DISSERTATION

submitted to the

Combined Faculty of Mathematics, Engineering

and Natural Sciences

of

Heidelberg University, Germany

for the degree of

Doctor of Engineering

Put forward by

Christian Ritter, M.Sc.

Born in: Leipzig, Germany

Data Fusion and Bayesian Smoothing for Tracking in Fluorescence Microscopy Images

Advisor: PD Dr. Karl Rohr

Oral examination: .....

Abstract

Obtaining spatial-temporal information of virus particles in fluorescence microscopy images is a prerequisite to gain insights into viral pathogens at a microscopic level about virus replication and assembly, but also at a macroscopic level to understand virus spread and infection in tissue-like structures. To obtain spatial-temporal information of virus particles, these structures need to be detected and tracked over time in time-lapse fluorescence microscopy image data. Since accurate manual determination of the position of many particles for all time points in microscopy image data is not feasible and introduces human bias, automatic computer vision methods for particle tracking and trajectory analysis are required.

In this thesis, new methods for probabilistic particle tracking are introduced based on data fusion and Bayesian smoothing. We propose data fusion approaches to incorporate image intensity and position as well as motion information and use a Bayesian framework to exploit uncertainties introduced by image noise and integrate a priori knowledge. To exploit image intensity and position information, we consider multiple measurements for each particle and fuse them by taking into account different uncertainties. Further, we developed a novel intensity-based probabilistic fusion approach which fuses results from multiple detectors and yields a consistent estimate of multiple fused detections to improve particle detection and localization. This approach integrates detections from classical and deep learning methods as well as exploits single-scale and multi-scale detections. To improve particle tracking by incorporating temporal information, we developed a novel Bayesian smoothing approach which integrates information from past and future time points. The covariance intersection algorithm is used to fuse position information and to obtain consistent trajectory estimates. In addition, motion information based on displacements from past and future time points is used to improve correspondence finding.

The novel methods were applied and evaluated on image data consisting of state-of-the-art benchmark data sets as well as live cell fluorescence microscopy image data showing immunodeficiency virus type-1 (HIV-1) and hepatitis C virus (HCV). It turns out, that the developed novel methods yield competitive or improved results compared to existing methods. We also applied the methods to quantify the motion and colocalization of HIV-1, HCV, and chromatin structures. Insights into viral and chromatin structures were obtained to better understand virus replication, assembly, spread and infectivity in tissue-like structures, as well as nuclear organization in mammalian cells.

Zusammenfassung

Die Gewinnung räumlich-zeitlicher Informationen über Viruspartikel in Fluoreszenzmikroskopiebildern ist eine Voraussetzung zur Erlanungung von Erkenntnissen über virale Krankheitserreger auf mikroskopischer Ebene in Bezug auf Virusreplikation und das Zusammenfügen, aber auch auf makroskopischer Ebene über das Verständnis der Virusausbreitung und Infektion in gewebeähnlichen Strukturen. Um räumlich-zeitliche Informationen von Viruspartikeln zu erhalten, müssen subzelluläre Strukturen in zeitlich aufgelösten fluoreszenzmikroskopischen Bilddaten detektiert und zeitlich verknüpft werden. Da eine genaue manuelle Bestimmung der Position einer großen Anzahl von Partikeln für alle Zeitpunkte in mikroskopischen Bilddaten nicht durchführbar ist und eine menschliche Verzerrung beinhaltet, sind automatische Methoden der Bildanalyse zur Partikelverfolgung und Trajektorienanalyse erforderlich.

In dieser Arbeit werden neue Methoden zur probabilistischen Partikelverfolgung vorgeschlagen, die neuartige Datenfusions- und Bayes'sche Glättungsansätze verwenden. Wir schlagen Datenfusionsansätze vor, um Bildintensitäts- und Positionssowie Bewegungsinformationen einzubeziehen, und verwenden einen Bayes'schen Ansatz, um Unsicherheiten zu nutzen, die durch Bildrauschen entstehen und integrieren a-priori-Wissen. Um Bildintensität und Positionsinformationen zu nutzen, betrachten wir mehrere Messungen für jedes Partikel und fusionieren diese unter Berücksichtigung verschiedener Unsicherheiten. Zur Verbesserung der Partikeldetektion und -lokalisation haben wir einen neuartigen intensitätsbasierten probabilistischen Fusionsansatz entwickelt, der die Ergebnisse mehrerer Detektoren vereinigt und eine konsistente Schätzung für die Partikelerkennung liefert. Darüber hinaus führt der neuartige Fusionsansatz eine Datenfusion von Detektionen von klassischen und Deep-Learning-Methoden durch und nutzt ein- und mehrskalige Detektionen. Um die Partikelverfolgung durch Einbeziehung zeitlicher Informationen zu verbessern, haben wir einen neuartigen Bayes'schen Glättungsansatz entwickelt, der Informationen aus vergangenen und zukünftigen Zeitpunkten fusioniert. Der Covariance-Intersection-Algorithmus wird verwendet, um Positionsinformationen zu fusionieren und konsistente Trajektorienschätzungen zu erhalten. Darüber hinaus werden Bewegungsinformationen, die auf Verschiebungen von vergangenen und zukünftigen Zeitpunkten basieren, zur Verbesserung der Korrespondenzfindung verwendet.

Die neuen Methoden wurden auf Bilddaten angewandt und ausgewertet, die aus Benchmark-Datensätzen sowie aus realen fluoreszenzmikroskopischen Bilddaten bestehen, die das Humane Immundefizienz-Virus Typ-1 (HIV-1) und das Hepatitis-C-Virus (HCV) zeigen. Es zeigt sich, dass die neu entwickelten Methoden im Allgemeinen konkurrenzfähige oder verbesserte Ergebnisse im Vergleich zu bestehenden Methoden liefern. Wir haben die Methoden auch auf Bilddaten von HIV-1, HCV und Chromatin Strukturen angewendet, um die Bewegung und Kolokalisierung zu analysieren. Dadurch konnten Einblicke in Virus- und Chromatinstrukturen

erhalten werden um die Virusreplikation, die Virusausbreitung und Infektiosität in gewebeähnlichen Strukturen, sowie die Zellkernorganisation in Säugetierzellen besser zu verstehen.

Acknowledgements

I am very grateful to many people who strongly supported me and my work, and without whom this thesis would not have been possible.

First of all, I want to thank Heidelberg University and Heidelberg University Hospital for providing a productive work environment and the possibility to meet people who are also dedicated to science. I want to thank *PD Dr. Karl Rohr* for supervising and supporting my work as well as his patience and guidance during preparation of publications and presentations of our work.

I am very grateful to many *people at BioQuant*, who made this place to a productive and stimulating place, but also to a home place to some extend. Especially, I would like to thank *Nina Beil* and *Jürgen Beneke* from the group of *Holger Erfle* for pleasant and humorous but also serious discussions about life and science during all lunches we had. A very special thanks goes to *Jürgen Beneke* for providing me (and almost complete BioQuant) with coffee without this thesis and work would not have been possible.

I am also very grateful to former and current colleagues at the Biomedical Computer Vision (BMCV) group for great discussions and shared ideas. A special thanks goes to *Thomas Wollmann* for countless fruitful discussions about science but also real life. It was a pleasure to work with you and having collaborative science projects with you. Further, a special thanks goes to my office roommate *Leonid Kostrykin.* I really loved the time sharing the office with you, all the endless discussions about my and your work and having some beers together. I really want to thank my particle tracking colleague *Roman Spilger* for sharing ideas, working together, and bringing a great spirit to BMCV. I am also very grateful to *Carola Krug* who showed unlimited support and highly motivated me during the last months of finishing this work. I am also very thankful for proof reading my thesis and giving comments. I am very grateful to all students who I came across during my work. It was a pleasure to work with you and all of you added a great value to this work. It was also a great pleasure to teach the students from molecular biotechnology in computer vision and bioinformatics. Last but not least, I want to thank our secretary *Sabrina Wetzel* and our former secretary *Manuela Schäfer* for beeing always available and your unlimited support in all kinds of administrative matters.

I also would like acknowledge the DFG-funded *Sonderforschungsbereich 1129* for financial support and I am very thankful for providing a productive environment and showing me the beauty and fascination of virology. I also would like to thank the group of *Ralf Bartenschlager* for various successful cooperation projects. Special thanks goes to the group of *Loan Dao Thi* and *Barbara Müller* for giving me the opportunity to have an internship and introducing me in cell culture and microscopy imaging.

I am very thankful to all guys from the *Skatrunde.* Especially, I thank *Julia* and *Stefan* for supporting me and beeing there whenever it was necessary.

I am very grateful to all my *Schwetzinger Freunde*. Without you and your unlimited support with strongly believe in me, this work would not have been possible. A special thanks goes to the *Forstner's*. You provided me with constant support, criticism at the right time, and providing me an environment where I could regain new strength. *Paula* and *Clara*, you both have been born during the time of this work. You both showed me, that there is more than science and with your smiles, all negative thoughts have been vanished. Thank you!

I also would like to thank *Meike* for being one of my deepest and oldest friends since high school. I am very grateful for all your constant support and openness.

My deepest gratitude belongs to my family, especially my Mom and my Dad. Without your unlimited and steadfast support you have given me throughout my life, I would not be at the point where I am today. In Love Thank You!

Contents

List of Figures

List of Tables

Nomenclature

Mathematical symbols frequently used in this thesis.

$*$	Convolution
$\mathrm{E}[\,]$	Expected value
$p(\,)$	Probability density function
$P(\,)$	Probability
$J(\,)$	Cost function
$\mathcal{L}$	Loss function
t	Time point
$\mathbf{x}$	State vector
$\mathbf{m}$	Mean state vector
$\mathbf{P}$	State covariance matrix
$\mathbf{m}_\mathrm{f}$	Fused mean state vector
$\mathbf{P}_\mathrm{f}$	Fused state covariance matrix
$\mathbf{K}_\mathrm{f}$, $\mathbf{L}_\mathrm{f}$, $\mathbf{M}_\mathrm{f}$	Data fusion gain matrices
$\mathbf{y}$	Measurement vector
β	Association probability
$\mathrm{MSD}(\Delta t)$	Mean squared displacement for time lag Δt
D	Diffusion coefficient
α	Anomalous diffusion exponent
Γ	Transport coefficient

Publications

Major parts of this thesis have been published in peer-reviewed journals and as peer-reviewed conference articles.

Peer-Reviewed Journal Articles

C. Ritter, J.-Y. Lee, M.-T. Pham, M. K. Pabba, M. C. Cardoso, R. Bartenschlager, and K. Rohr, "Multi-detector fusion and Bayesian smoothing for tracking viral and chromatin structures," *Under review*, 2022

C. Ritter, R. Thielemann, J.-Y. Lee, M.-T. Pham, R. Bartenschlager, and K. Rohr, "ColocQuant and ColocJ: Multi-channel colocalization analysis of viral proteins in fluorescence microscopy images," *In preparation*, 2022

M.-T. Pham, J.-Y. Lee, **C. Ritter**, R. Thielemann, U. Haselmann, C. Funaya, V. Laketa, K. Rohr, and R. Bartenschlager, "Intercellular transmission of viral RNA by Apolipoprotein E associated extracellular vesicles," *Under review*, 2022

C. Ritter, T. Wollmann, J.-Y. Lee, A. Imle, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "Data fusion and smoothing for probabilistic tracking of viral structures in fluorescence microscopy images," *Medical Image Analysis*, vol. 73, 102168, 2021

C. Ritter, T.Wollmann, P. Bernhard, M. Gunkel, D. M. Braun, J.-Y. Lee, J. Meiners, R. Simon, G. Sauter, H. Erfle, K. Rippe, R. Bartenschlager, and K. Rohr, "Hyperparameter optimization for image analysis: Application to prostate tissue images and live cell data of virus-infected cells," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 1847–1857, 2019

A. Imle, P. Kumberger, N. D. Schnellbächer, J. Fehr, P. Carrillo-Bustamante, J. Ales, P. Schmidt, **C. Ritter**, W. J. Godinez, B. Müller, K. Rohr, F. A. Hamprecht, U. S. Schwarz, F. Graw, and O. T. Fackler, "Experimental and computational analyses reveal that environmental restrictions shape HIV-1 spread in 3D cultures," *Nature Communications*, vol. 10, no. 2144, pp. 1–18, 2019

J.-Y. Lee, M. Cortese, U. Haselmann, K. Tabata, I. Romero-Brey, C. Funaya, N. L. Schieber, Y. Qiang, M. Bartenschlager, S. Kallis, **C. Ritter**, K. Rohr, Y. Schwab, A. Ruggieri, and R. Bartenschlager, "Spatiotemporal coupling of the hepatitis C virus replication cycle by creating a lipid droplet-proximal membranous replication compartment," *Cell Reports*, vol. 27, no. 12, pp. 3602–3617.e5, 2019

Peer-Reviewed Conference Articles

C. Ritter, R. Spilger, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "Deep learning for particle detection and tracking in fluorescence microscopy images," in *Proc. International Symposium on Biomedical Imaging (ISBI 2021)*, Nice, France, pp. 873–876, IEEE, 2021

D. Schacherer, **C. Ritter**, and K. Rohr, "Multiple hypothesis tracking with integrated cell division detection," in *Proc. International Symposium on Biomedical Imaging (ISBI 2021)*, Nice, France, pp. 165–168, IEEE, 2021

C. Ritter, A. Newrly, S. Schifferdecker, I. Roggenbach, B. Müller, and K. Rohr, "Optimizing particle detection by colocalization analysis in multi-channel fluorescence microscopy images," in *Proc. International Symposium on Biomedical Imaging (ISBI 2020)*, Iowa City, Iowa, USA, pp. 882–885, IEEE, 2020

C. Ritter, T. Wollmann, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "Deep learning particle detection for probabilistic tracking in fluorescence microscopy images," in *Proc. International Symposium on Biomedical Imaging (ISBI 2020)*, Iowa City, Iowa, USA, pp. 977–980, IEEE, 2020

T. Wollmann, **C. Ritter**, J.-N. Dohrke, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "DetNet: Deep neural network for particle detection in fluorescence microscopy images," in *Proc. International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, pp. 517–520, IEEE, 2019

C. Ritter, A. Imle, J.-Y. Lee, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "Two-filter probabilistic data association for tracking of virus particles in fluorescence microscopy images," in *Proc. International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, D.C., USA, pp. 957–960, IEEE, 2018

1 Introduction

1.1 Motivation

Viruses have the potential to affect all forms of life starting from simple life forms such as bacteria to complex organisms such as humans where a virus infection might represent a threat to life. Furthermore, viruses have the potential to spread fast and therefore infect a large population of people which can lead to pandemic situations such as the current global COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Other pandemic situations are caused by the outbreak of other viruses such as the human immunodeficiency virus type-1 (HIV-1) and the hepatitis C virus (HCV).

An HIV-1 infection causes the acquired immune deficiency syndrome (AIDS) and is therefore a major global public health issue. As reported by the WHO, 36 million people have lost their lives and approximately 37.7 million people live with HIV at the end of 2020 [1]. An HIV-1 infection can be caused by the exchange of a variety of body fluids from infected people such as blood, breast milk, semen, and vaginal secretions. Further, HIV-1 can also transmitted from the mother to her child during pregnancy and delivery. HIV-1 is an enveloped virus with a lipid membrane which embeds spikes of glycoprotein complexes and two copies of genomic single stranded RNA are encapsulated by the cone-shaped capsid [2]. During virus replication, the virus genomic information is integrated into the host-cell DNA [3]. As virus-host interactions that govern the infection of individual cells with HIV-1, the host receptor (CD4), the chemokine receptor CC5, and the CXC-chemokine receptor 4 (CXCR4) are identified and the main target cells for infection are $CD4^+$ T cells and macrophages [4, 5].

The infection with HCV can cause both acute and chronic hepatitis which is an inflammation of the liver ranging from mild to serious progressions with lifelong illness including liver cirrhosis and cancer. The WHO estimates that globally 58 million people have a chronic hepatitis C virus infection and about 1.5 million infections occur per year [6]. An HCV infection can be caused by inadequate sterilization of medical equipment (e.g., syringes, needles), transfusion of HCV infected blood and blood products, but also due to drug injection with shared injection equipment which is HCV infected. The existence of HCV was demonstrated in mid-1970s since patients with hepatitis after blood transfusion have been reported which was not caused by hepatitis A or B viruses [7]. The hepatitis C virus is an enveloped plus-stranded RNA virus with a lipid membrane which embeds two glycoproteins, E1 and E2,

and consists of non-structural proteins (e.g., NS5A) which are important in virus replication and assembly [8]. In contrast to HIV-1, HCV genomic information is released into the cytoplasm and exploited for replication and assembly which takes place at the endoplasmic reticulum aided by lipid droplets [9].

However, the replication and assembly of HCV is not fully understood. Especially, spatial-temporal knowledge in conjunction with information on virus-host interaction is still missing to fully understand HCV assembly. Concerning HIV-1, knowledge about the HIV-1 post-entry phase consisting of nucleus entry and capsid uncoating could be achieved in recent works [10, 2]. However, a functional understanding in conjunction with a spatial-temporal view is still incomplete. As for the microscopic level, virus spread and infection on a macroscopic level with complex biological systems taking into account tissue characteristics is not fully understood.

To gain insights into virus replication or virus spread, time-lapse live cell fluorescence microscopy in conjunction with virus particle detection and tracking allows quantification of viral structures at a high spatial-temporal resolution. Insights into virus assembly can be obtained by colocalization analysis of tracked viral structures. To understand virus-host interactions, cell tracking is required as an additional step, and cell dependent genome organization can be studied by tracking and motion analysis of chromatin structures. In order to draw sound statistical conclusions, a large number of correctly detected and tracked virus particles are required. However, manual annotation of image sequences consisting of a large number of frames is not feasible and additionally prone to errors since human bias is introduced. Therefore, automatic computer vision methods and complex image analysis pipelines with optimized hyperparameters are required. To deal with uncertainties in the image data caused by noise and to exploit available a priori knowledge, this thesis presents methods based on a Bayesian framework.

1.2 Contributions and Organization of the Thesis

This thesis introduces methods based on data fusion and Bayesian smoothing for integrating spatial and temporal information to determine trajectories of virus and chromatin structures in live cell fluorescence microscopy images. The proposed methods improve main tasks of trajectory determination, namely object detection and tracking. The methods were applied to real fluorescence microscopy image data for infectious disease and genome organization research projects involving motion and colocalization analysis of HIV-1 and HCV as well as chromatin structures. The main contributions of this thesis are:

- **Two-Filter Probabilistic Data Association for Particle Tracking:** A novel particle tracking method for microscopy images is proposed which is based on a two-filter smoothing approach and probabilistic data association. The approach exploits spatial and temporal information by fusing information from past and future time points, integrates multiple measurements, and combines Kalman filtering and particle filtering. The novel tracking method was applied to data of the Particle Tracking Challenge and fluorescence microscopy data of HCV proteins and HIV-1. The tracking method improved the results compared to previous methods. The work was published in [11].

- **Bayesian Smoothing using Covariance Intersection for Particle Tracking:** A novel method for particle tracking in microscopy images is proposed based on Bayesian smoothing using covariance intersection to incorporate temporal information by fusing information from past and future time points. Predictions are obtained by two filters running in opposite temporal directions and fused in the state space by the covariance intersection algorithm which is a general multi-sensor data fusion method for unknown cross-covariances. In addition, motion information is exploited by fused displacements from past and future time points and integrated in the cost function for correspondence finding. Further, a multi-sensor data fusion approach with probabilistic data association is used which fuses multiple measurements from separate measurement processes by integrating detection-based and prediction-based measurements. The approach allows taking into account different uncertainties to improve update estimation. Both detection-based and prediction-based measurements are obtained from a Kalman filter using probabilistic data association with elliptical sampling. The novel tracking method was applied to data of the Particle Tracking Challenge and fluorescence microscopy data of HIV-1 and HCV proteins acquired with different types of microscopes and spatial-temporal resolutions. It turned out that the tracking method yielded state-of-the-art results or outperformed previous methods. The work was published in [12].

- **Multi-Detector Fusion and Bayesian Smoothing for Particle Tracking:** A novel probabilistic tracking approach for multiple particle tracking based on multi-detector data fusion and Bayesian smoothing is proposed. The approach fuses results from multiple detectors using a novel intensity-based covariance intersection method which takes into account information about image intensities, positions, and uncertainties. The method ensures a consistent estimate of multiple fused particle detections and does not require an optimization step. The tracking approach performs data fusion of detections from classical and deep learning methods as well as exploits single-scale and multi-scale particle detections. Further, for each particle a time-varying estimate of the measurement noise covariance is computed to improve update estimation for particle tracking. In addition, Bayesian smoothing is used to fuse predictions obtained from both past and future time points. The novel tracking method was evaluated using data of the Particle Tracking Challenge and achieved state-of-the-art results or outperformed previous methods. Further, the method was applied to challenging time-lapse fluorescence microscopy image data of HCV associated proteins and chromatin structures and outperformed existing methods. The work has been submitted for publication [13].

- **Deep Learning for Particle Detection and Tracking:** For particle tracking, a novel approach is proposed which combines deep learning and Bayesian sequential estimation by using a domain adapted network for particle detection in conjunction with probabilistic data association for tracking. Data association parameters that depend on the detection result are identified and automatically determined by hyperparameter optimization. Further, an extension is introduced which uses deep learning for both particle detection and particle association by combining a domain adapted network for detection with an LSTM-based recurrent neural network for tracking. The work was published in [14], [15], and [16].

- **ColocQuant and ColocJ - Colocalization Analysis for Multi-Channel Microscopy Images:** A novel colocalization analysis approach for multi-channel microscopy images is proposed which consists of object-based detection, quantification, and visualization of colocalizations of virus structures in multi-channel fluorescence microscopy images. The approach uses a multi-dimensional graph-based method to efficiently identify multi-channel colocalized particles and visualizes the color composition of colocalizations by a Maxwell color triangle. The approach has been implemented in a software suite with integrated graphical user interface and was applied to three-channel fluorescence microscopy images of HCV to characterize HCV assembly sites [17].

- **Motion Analysis of Subcellular Structures:** The developed probabilistic tracking methods were applied to time-lapse fluorescence microscopy image sequences of HIV-1 particles. Diffusion parameters of HIV-1 particles were quantified and used for mathematical modeling of virus spread kinetics in suspension and 3D collagen structures. Further, probabilistic tracking was applied to time-lapse confocal microscopy image sequences of chromatin within mammalian cells. Diffusion parameters of chromatin were quantified and correlated to sites in proximity of active DNA replication based on colocalization analysis. The work was published in [18], [19], and has been submitted for publication [20].

Organization of the Thesis Chapter 2 describes fundamentals of fluorescence microscopy imaging, data fusion, and Bayesian smoothing. Further, previous work on particle detection and tracking within the field of computer vision for biological images is reviewed. In Chapter 3, novel Bayesian smoothing approaches for particle tracking in microscopy images are proposed which incorporate temporal information by fusing information from past and future time points. Chapter 4 introduces a novel multi-detector fusion approach with intensity-based covariance intersection for particle detection and proposes a multi-detector tracking approach based on Bayesian smoothing. Chapter 5 describes novel methods for particle tracking which integrate deep learning methods for particle detection and tracking. Chapter 6 presents the experimental results of the proposed methods described in Chapters 3 to 5. The methods were applied to data of the Particle Tracking Challenge as well as for live cell real fluorescence microscopy image data displaying different types of viral and chromatin structures. In Chapter 7, a framework for hyperparameter optimization is described and used to optimize hyperparameters for particle detection in multi-channel microscopy images. Further, a novel probabilistic cell tracking approach with integrated cell division detection is introduced. In Chapter 8, a novel colocalization analysis approach for multi-channel microscopy images is proposed. The developed probabilistic particle tracking methods were applied to real fluorescence microscopy image data and motion as well as colocalization analysis of virus and chromatin structures was performed. A summary and outlook with suggestions for future work are given in Chapter 9.

2 Fundamentals and Previous Work

This chapter introduces fundamental concepts essential for this thesis. First, fundamentals of microscopy imaging such as fluorescence microscopy and fluorescence labeling are reviewed. Then, data fusion concepts such as multi-sensor data fusion algorithms, Kalman filtering, and Bayesian smoothing are described. Afterwards, a brief introduction of deep learning for biomedical computer vision is given. Further, approaches for particle tracking in microscopy images are presented and discussed, followed by introducing methods for colocalization and motion analysis.

2.1 Microscopy Imaging

Virus and chromatin structures have a small size and require a magnified image provided by microscopy techniques for observation. To gain insights into cellular processes, the subcellular structures need to be observed *in situ* to conduct measurements directly and dynamically in live cells. Time-lapse *fluorescence microscopy* capture such structures and is able to uncover mechanisms of virus infection, assembly, release, or cell proliferation in high spatial and temporal resolution [21, 22, 23].

Fluorescence Microscopy

Fluorescence microscopy belongs to optical microscopy techniques and utilizes the principle of fluorescence which is the emission of light occurring after the absorption of light with a shorter wavelength [21, 22]. The subcellular structure of interest is labeled with a fluorescent protein (e.g., green fluorescent protein (GFP)) which is a fluorophore of small size and consisting of electrons which can reach an excited state with higher energy due to absorption of light (photons) with a specific wavelength. Due to vibrations, the electrons lose energy and when falling back to ground state the released energy during this process is emitted as light. The wavelength of the emitted light is larger than the wavelength of the absorbed light, which means that the energy of the emitted photons is smaller than that of the absorbed photons. This shift of wavelengths (energies) is denoted as Stokes shift. The main task is to separate the spectrum of emission and stimulation light efficiently and accurately obtain high quality signals from the subcellular structure of interest. The stimulation light is usually generated by high pressure metal halide lamps (with mercury and/or xenon) or by light-emitting diodes (LED). However, if high intensity of light in

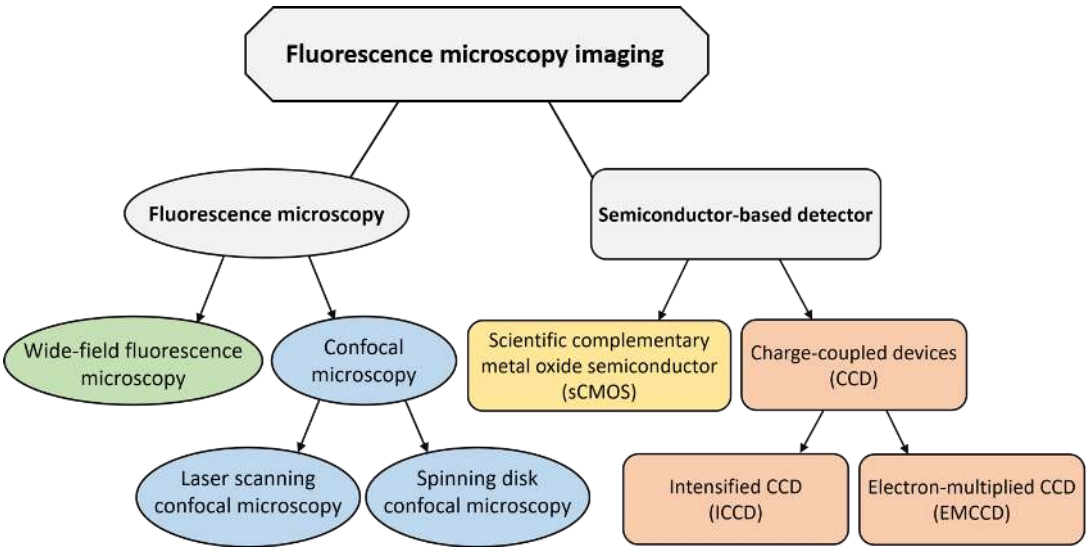


Figure 2.1: Overview of fluorescence microscopes and detectors.

a small area of the specimen is required, lasers are used as light sources [22]. An overview of fluorescence microscopes and detectors is given in Fig. 2.1.

Wide-field fluorescence microscopy (WF) illuminates a large field of the specimen, generates a large field of view, and uses a highly sensitive camera [23]. WF has been used to investigate virus-cell interaction for the HIV-1 entry process [24] and to study HIV-1 assembly and release [25]. However, in WF the fluorophores above and below the image plane in focus are excited. This is suboptimal for thick and highly scattering specimen, as the focused image is overlaid by blurred images, resulting in low image contrast and resolution. [22].

Confocal microscopy (CM) reduces these perturbations, and thick specimen like living samples can be observed. CM exploits the principle of pinholes to eliminate out-of-focus signals [22, 23] produced by fluorophores outside the image plane in focus. A pinhole in front of the detector is used to eliminate emission light generated outside the image plane in focus and a second pinhole placed after the light source prevents the stimulation of fluorophores outside the image plane in focus. For CM, two types of systems exist (see Fig. 2.1), *laser scanning confocal microscopy* (LSCM) and *spinning disk confocal microscopy* (SDCM) [26]. LSCM uses lasers as light sources and scans the specimen point-by-point due to a scanning unit either by moving scanning mirrors or moving specimen slide. The data is collected one pixel at a time by a photomultiplier tube as detector. Disadvantages are that the microscopy system is limited in scanning speed and by the number of photons available per pixel [23]. In contrast, SDCM scans multiple positions simultaneously by using a mechanically spinning Nipkow disk which consists of multiple pinholes arranged in a specific pattern. Both LSCM and SDCM use dichroic mirrors to separate emission and stimulation light. In contrast to LSCM, SDCM uses an electron multiplying charge-coupled device (EMCCD) camera as detector to collect the emission light which leads to higher quantum efficiency, higher image quality, and faster read

out [23]. Combing SDCM with a scan device for the z-coordinate, the system allows fast recording of three-dimensional (3D) images which has been intensively used for 3D live cell imaging to study viral replication (e.g., [10, 2, 18, 27]).

The most common sensitive detectors for modern fluorescence microscopes are camera systems based on modern semiconductor technologies [28, 29, 23] (see Fig. 2.1). The most commonly used cameras are based on *charge-coupled devices* (CCD) which exploit the inner photoelectric effect and measure the amount of electrical charges that has built up during the time that the CCD was illuminated [30]. *Intensified charge-coupled device* (ICCD) cameras use an image intensifier in front of the CCD to improve quantum efficiency which leads to a higher sensitivity [28, 29]. The image is transferred with fiber optics to a microchannel plate which converts the emission photons to electrons, multiplies them by impact ionization, and converts the enhanced electron signal back to photons. The multiplied emission photons are measured by CCD behind the image intensifier. The *electron-multiplied charge-coupled device* (EMCCD) camera exploits a shift register and high voltage electron multiplying register to perform electron amplification on chip before the detected signal is read-out [29, 23]. EMCCD cameras simultaneously achieve a fast read-out speed and high sensitivity due to high quantum efficiency. Compared to CCD-based camera systems, *scientific complementary metal oxide semiconductor* (sCMOS) use semiconductor-based electronic circuits directly integrated into each pixel and thus only parts of the pixel are photon sensitive [29]. sCMOS cameras have a very fast read-out speed with several hundred frames per second compared to CCD-based cameras with frame rates of around one hundred frames per second [31, 29, 23].

Fluorescence Labeling

Cell structures can be fluorescently imaged after the specimen is stained or viral components are labeled by fluorophores. An overview and classification of fluorescence labeling is given in Fig. 2.2. Cell nuclei can be imaged with 4',6-diamidino-2-phenylindole (*DAPI*) staining. DAPI binds to deoxyribonucleic acid (DNA) regions rich in adenine-thymine and has the advantage that it can be combined with other dyes (e.g., GFP, Cy3, Cy5, Alexa Fluor 488) [22]. Example images of DAPI staining are given in Fig. 2.3 acquired at the Infectious Diseases Imaging Platform at the Center for Integrative Infectious Diseases Research (CIID), Heidelberg University Hospital. Directly labeling viral components is a challenging task, since viral structures are densely packed and therefore are limited in size and location for the fluorophore. The fluorophores are required to not inhibit the functionality of the viral structure as well as not hamper the infectivity of the virus [23]. The fluorophores are desired to be bright with a strong ability to capture photons by a large molecular absorption coefficient and high fluorescence quantum yield in order to obtain a high spatial resolution and fast image acquisition. Further, a high photostability is required for many excitation-deexcitation cycles.

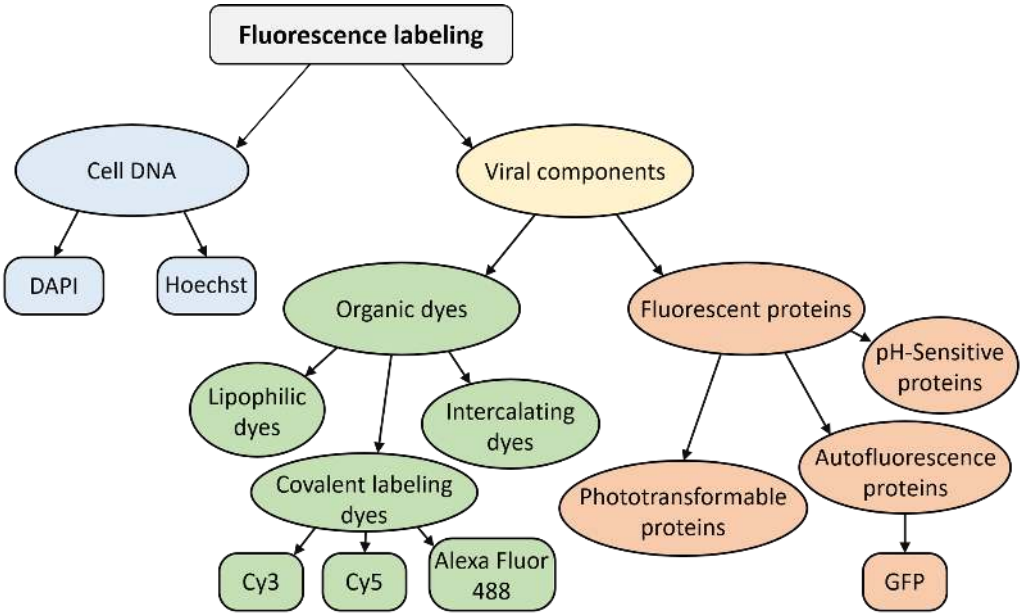


Figure 2.2: Overview and classification of fluorescence labeling techniques and used fluorophores.

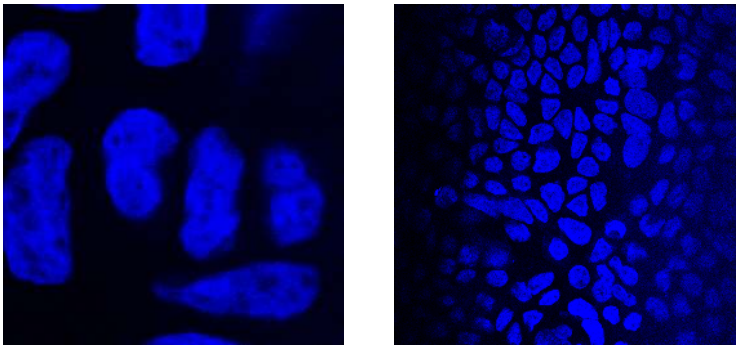


Figure 2.3: Live cell microscopy images of DAPI stained cell nuclei imaged with a spinning disk confocal microscope.

There exist several types of fluorophores including organic dyes and fluorescent proteins. *Organic dyes* are small fluorescent labels with good photophysical properties and many reactive groups are available for various labeling strategies with a wide spectral range. Depending on their virus labeling strategy, organic dyes can be subdivided in covalent labeling dyes, lipophilic dyes, and intercalating dyes. Covalent labeling dyes bind to the viral component by a covalent bond and often used labeling dyes are, e.g., Cy3 (orange), Cy5 (far-red), or Alexa Fluor 488 (green). For example, Alexa Fluor 488 has been applied for labeling the capsid and envelope of human papillomavirus (HPV) to study the kinetics of the viral entry process [32]. Lipophilic dyes are able to bind to the lipid membrane of enveloped viruses (e.g., HIV [33, 34], HCV [35]) due to the incorporation into the envelope by hydrophobic-lipophilic interactions [23]. Intercalating dyes are able to penetrate the outer virus components in order to label the encapsulated virus genome to a certain extent. However, the

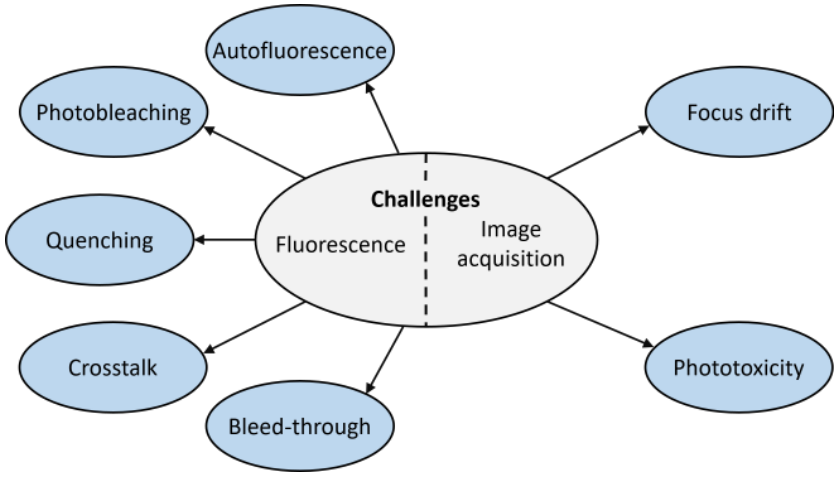


Figure 2.4: Overview of main challenges of fluorescence techniques and image acquisition for image analysis.

dyes have the disadvantage that they can interact with the viral RNA and thus inactivate the viral RNA during illumination [23].

Fluorescent proteins are a type of fluorophores which are genetically encodable such that the fluorescent protein DNA can be fused with the protein DNA of interest (target protein to be labeled). The host cell can be transfected with the fusion product and the expression level of the fluorophore is related to transcription and transfection efficiency. Types of fluorescent proteins can be classified into autofluorescence proteins, pH-sensitive fluorescent proteins, and phototransformable fluorescent proteins [23]. The green fluorescent protein (GFP) and GFP-like proteins are autofluorescence proteins, and the property of fluorescence is directly encoded in the sequence and preserved when fused with the DNA of proteins of interest. The pH-sensitive fluorescent proteins change their optical property when the pH (potential of hydrogen) of the enviroment changes. These proteins can be used to detect pH changes in the cell and have been applied to HIV to study the fission of newly assembled virus particles [36]. Phototransformable fluorescent proteins can be activated or their spectrum can be shifted using light.

Challenges of Fluorescence Microscopy Imaging for Image Analysis

In live cell time-lapse fluorescence microscopy imaging, various challenges of fluorescence techniques and image acquisition for image analysis exist, e.g., *autofluorescence*, *photobleaching*, *quenching*, *crosstalk*, *bleed-through*, *phototoxicity*, or *focus drift* [21, 22]. An overview and classification of the challenges is given in Fig. 2.4. The image contrast is reduced due to background noise generated by *autofluorescence* which is the natural emission of photons by biological structures within the specimen to be observed (e.g., mitochondria, lysosomes) [37]. *Photobleaching* is caused by destruction of the fluorophores due to the light source and bleaches the fluorescence signal. It can be prevented by using lower stimulation light intensity and/or faster

acquisition which reduces the degradation of the fluorescent proteins. After stimulation by light, the energy of fluorophores in excited state (donator) can be given to a second fluorophore (acceptor) due to Förster resonance energy transfer (FRET). This reduces the fluorescence of the molecules and is denoted as *quenching*. Challenges arise if multiple fluorophores are used. Fluorophores with overlapping stimulation light spectra lead to simultaneous illumination (*crosstalk*) and fluorophores with overlapping emission spectra can not be separated by filters (*bleed-through*). During live cell imaging, *phototoxicity* is the damaging effect of living cells due to illumination by stimulation light (e.g., laser). The fluorophores in excited state react with molecular oxygen and produce cell-damaging free radicals. Low stimulation light intensity and short image acquisition time can prevent phototoxicity. *Focus drift* in time-lapse fluorescence microscopy is a major challenge induced by the inability of a microscopy system to maintain the selected focal plane over the time period of data acquisition. It leads to artifacts in the image sequence and wrong focus position caused by several faults (e.g., specimen movement, vibrations and thermal drift, lateral observation stage movement, mechanically instability of components in the microscope).

2.2 Data Fusion

Data fusion is a framework comprising techniques and methods that combine data from multiple sensors in order to achieve improved accuracy and more specific inferences compared to using a single sensor. The quality of the requested information can be improved by using multiple sensors or using data from various sources of different nature [38, 39, 40, 41]. A sensor is typically a physical device that is sensitive to a physical property in order to provide a measurement as primary output [42, 40]. Also, the term sensor can be more general and refers to mechanisms that derive measurements which can be signals or images [38, 40]. However, environmental factors lead to systematic errors in the measurement which are considered by uncertainty of the sensor measurements. Measurements from multiple sensors can be interpreted as estimates of the state of a common system and can be fused by multi-sensor data fusion approaches. However, determining the best procedure (estimator) to combine estimates (measurements) obtained by multiple sensors is challenging. Using a probabilistic framework for data fusion has the advantage that explicit probabilistic models are employed which consider the uncertainty of the estimates.

The aim of multi-sensor data fusion is to fuse multiple estimates to obtain a *consistent estimate* with mean $\mathbf{m}_f \in \mathbb{R}^{n\times 1}$ and covariance $\mathbf{P}_f \in \mathbb{R}^{n\times n}$ for which the actual error covariance $\mathrm{E}[\tilde{\mathbf{m}}\,\tilde{\mathbf{m}}^T]$ is bounded by the estimated covariance $\mathbf{P}_f$, i.e., $\mathbf{P}_f \geq \mathrm{E}[\tilde{\mathbf{m}}\,\tilde{\mathbf{m}}^T]$, where $\tilde{\mathbf{m}} = \mathbf{m}_f - \mathbf{m}$ is the error with respect to the (true) mean $\mathbf{m} \in \mathbb{R}^{n\times 1}$ [43]. Further, the inequality of the estimated and actual error covariance $\mathbf{P}_f \geq \mathrm{E}[\tilde{\mathbf{m}}\,\tilde{\mathbf{m}}^T]$ means that $\mathbf{P}_f - \mathrm{E}[\tilde{\mathbf{m}}\,\tilde{\mathbf{m}}^T]$ is positive semidefinite. Consistency implies that $\mathbf{P}_f$ is an upper bound of the actual error covariance matrix.

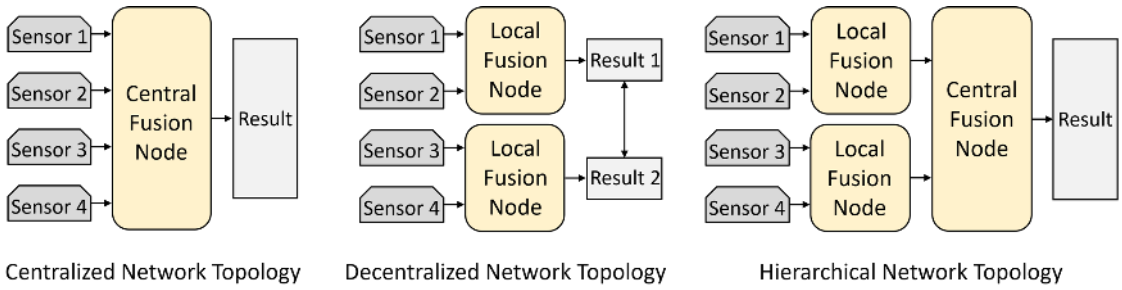


Figure 2.5: Comparison of different multi-sensor network typologies.

Multi-Sensor Network Topology

Fusing the information obtained by multiple sensors depends on the topology of the fusion nodes. A fusion node obtains data provided by multiple sensors and applies data fusion based on the used data fusion approach. Generally, the fusion nodes and sensors can be arranged in three different multi-sensor network topologies, *centralized*, *decentralized*, and *hierarchical* network topology [40, 41]. The different network topologies are schematically depicted in Fig. 2.5.

A *centralized* network topology exploits a central fusion node as central processor that collects all information provided from different sensors. Multi-sensor data fusion is done in the central fusion node and provides the fusion result. In principle, the central fusion node as single fusion node yields the best performance under the assumption that all sensors are accurately aligned. Sensor alignment requires a single common representational format of the sensory data and considers spatial, temporal, or semantic alignment or data normalization. Therefore, a disadvantage of centralized network topologies is that small errors in sensor alignment can cause strong reduction in the fusion performance. Further, all sensors are required to communicate with the central fusion node which may cause a communication bottleneck. The centralized network topology is also hard to scale in structure since new sensors with new sensor data can not directly be added to the centralized system.

A *decentralized* network topology fuses multiple sensors with multiple local fusion nodes compared to a single central fusion node. The main advantage of the decentralized network topology is the absence of the requirement of sensor alignment. Further, the communication and data processing is distributed over the network. The decentralized network is also scalable since the structure is not constrained by a central processor and the communication bandwidth limit, and a modular design of the network structure can be accomplished. As shown in Fig. 2.5, the local fusion nodes are only able to compute partial results which need to be combined in order to obtain a result representation similar to the centralized network topology. Often the network structure needs to be known in order to combine the partial results. However, a main disadvantage of decentralized networks is the effect of redundant information wrongly decreasing the uncertainty in the fused estimate which occurs if information is fused under the (wrong) assumption of independence.

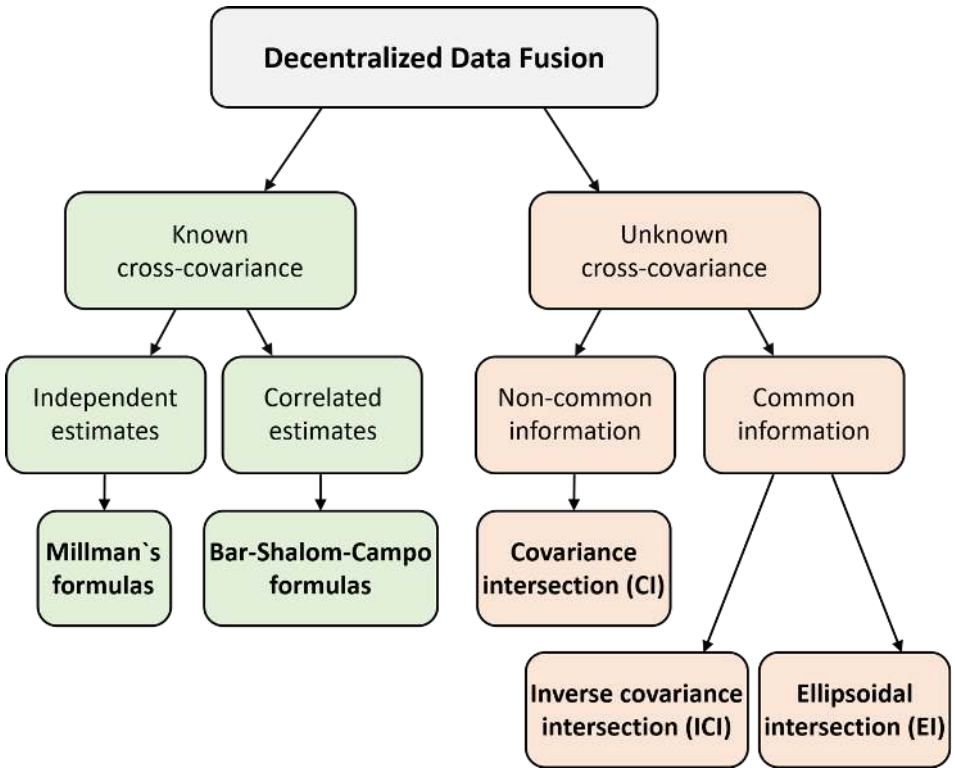


Figure 2.6: Overview and classification of multi-sensor data fusion algorithms.

A hybrid network topology combining centralized and decentralized network topology is the *hierarchical* network topology. This topology exploits local fusion nodes and uses a central fusion node in order to fuse the information to obtain the result. As with the decentralized network topology, the hierarchical network has the disadvantage that it may suffer from the negative effect of redundant information (e.g., data incest) which wrongly decreases the uncertainty of the fused estimate.

Multi-Sensor Data Fusion Algorithms

An optimal fusion method requires a centralized network topology with a central fusion node which has the disadvantage of a high communication demand and needs sensor alignment with a common representational format of the data [40, 44, 43]. Therefore, in multi object tracking a decentralized network topology is preferred [45, 44] and has also been applied to particle tracking in microscopy image data (e.g., [13, 12, 11, 46, 47, 48, 49]). The treatment of cross-correlations between the sensor estimates to be fused is challenging and has received strong attention in the field of multi-sensor data fusion.

In this section, multi-sensor data fusion algorithms for decentralized network topologies are reviewed and a classification of the fusion algorithms is given as an overview chart in Fig. 2.6. The data fusion methods are considered to solve as a two-sensor data fusion problem of two unbiased estimates provided by sensor A and B with mean $\mathbf{m_A} \in \mathbb{R}^{n\times 1}$ and $\mathbf{m_B} \in \mathbb{R}^{n\times 1}$ and with covariance matrix $\mathbf{P_A} \in \mathbb{R}^{n\times n}$

and $\mathbf{P}_\mathrm{B} \in \mathbb{R}^{n \times n}$. The mean of the fused estimate $\mathbf{m}_f$ can be generally obtained by a linear combination [50, 40]

$$\mathbf{m}_\mathrm{f} = \mathbf{K}_\mathrm{f}\,\mathbf{m}_\mathrm{A} + \mathbf{L}_\mathrm{f}\,\mathbf{m}_\mathrm{B} \tag{2.1}$$

with the fusion gain matrices $\mathbf{K}_\mathrm{f} \in \mathbb{R}^{n \times n}$ and $\mathbf{L}_\mathrm{f} \in \mathbb{R}^{n \times n}$. The corresponding covariance $\mathbf{P}_\mathrm{f}$ of the fused estimate is

$$\mathbf{P}_\mathrm{f} = \mathbf{K}_\mathrm{f}\,\mathbf{P}_\mathrm{A}\,\mathbf{K}_\mathrm{f}^T + \mathbf{K}_\mathrm{f}\,\mathbf{P}_\mathrm{AB}\,\mathbf{L}_\mathrm{f}^T + \mathbf{L}_\mathrm{f}\,\mathbf{P}_\mathrm{BA}\,\mathbf{K}_\mathrm{f}^T + \mathbf{L}_\mathrm{f}\,\mathbf{P}_\mathrm{B}\,\mathbf{L}_\mathrm{f}^T \tag{2.2}$$

where $\mathbf{P}_\mathrm{AB} \in \mathbb{R}^{n \times n}$ and $\mathbf{P}_\mathrm{BA} \in \mathbb{R}^{n \times n}$ are the cross-covariances. The fusion gain matrices $\mathbf{K}_\mathrm{f}$ and $\mathbf{L}_\mathrm{f}$ can be computed by minimizing a cost function $J(\mathbf{P}_\mathrm{f})$ of the fused covariance $\mathbf{P}_\mathrm{f}$ with respect to $\mathbf{K}_\mathrm{f}$ and $\mathbf{L}_\mathrm{f}$. $J(\mathbf{P}_\mathrm{f})$ can be any arbitrary strictly monotonically increasing cost function such as the trace or determinant of $\mathbf{P}_\mathrm{f}$ [51, 52] which corresponds to minimizing the mean squared error $\tilde{\mathbf{m}} = \mathbf{m}_f - \mathbf{m}$ [45, 50]. The equations (2.1) and (2.2) are denoted as *Generalized Millman's formulas.*

Millman's formulas

For *known cross-covariances* and assuming *independent estimates* for which the cross-covariances are $\mathbf{P}_\mathrm{AB} = \mathbf{P}_\mathrm{BA} = \mathbf{0}$, the *Millman's formula* is a special case of the generalized Millman's formulas. The fused estimate mean $\mathbf{m}_\mathrm{f}$ can be computed by $\mathbf{m}_\mathrm{A}$ and $\mathbf{m}_\mathrm{B}$ using [50, 40]

$$\mathbf{m}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{B}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B}}\,\mathbf{m}_\mathrm{A} + \frac{\mathbf{P}_\mathrm{A}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B}}\,\mathbf{m}_\mathrm{B} \tag{2.3}$$

where the gain matrices are $\mathbf{K}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{B}}{\mathbf{P}_\mathrm{A}+\mathbf{P}_\mathrm{B}}$ and $\mathbf{L}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{A}}{\mathbf{P}_\mathrm{A}+\mathbf{P}_\mathrm{B}}$. The fused covariance $\mathbf{P}_\mathrm{f}$ is

$$\mathbf{P}_\mathrm{f} = \left((\mathbf{P}_\mathrm{A})^{-1} + (\mathbf{P}_\mathrm{B})^{-1}\right)^{-1} \tag{2.4}$$

Bar-Shalom-Campo formulas

If the *cross-covariances* are *known* and the *estimates* from sensor A and B are *correlated* ($\mathbf{P}_\mathrm{AB} = \mathbf{P}_\mathrm{BA} \neq \mathbf{0}$), the fused estimate can be computed by the *Bar-Shalom-Campo formulas* [53]. Compared to Millman's formula, the Bar-Shalom-Campo formulas consider the cross-covariances $\mathbf{P}_\mathrm{AB}$ and $\mathbf{P}_\mathrm{BA}$, and the estimated fused mean $\mathbf{m}_\mathrm{f}$ is computed by

$$\mathbf{m}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{BA}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{AB} - \mathbf{P}_\mathrm{BA}}\,\mathbf{m}_\mathrm{A} + \frac{\mathbf{P}_\mathrm{A} - \mathbf{P}_\mathrm{AB}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{AB} - \mathbf{P}_\mathrm{BA}}\,\mathbf{m}_\mathrm{B} \tag{2.5}$$

where the optimal gain matrices are $\mathbf{K}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{BA}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{AB} - \mathbf{P}_\mathrm{BA}}$ and $\mathbf{L}_\mathrm{f} = \frac{\mathbf{P}_\mathrm{A} - \mathbf{P}_\mathrm{AB}}{\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{AB} - \mathbf{P}_\mathrm{BA}}$. The covariance $\mathbf{P}_\mathrm{f}$ of the fused estimate is

$$\mathbf{P}_\mathrm{f} = \mathbf{P}_\mathrm{A} - (\mathbf{P}_\mathrm{A} - \mathbf{P}_\mathrm{AB}) \left(\mathbf{P}_\mathrm{A} + \mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{AB} - \mathbf{P}_\mathrm{BA}\right)^{-1} (\mathbf{P}_\mathrm{A} - \mathbf{P}_\mathrm{BA}) \tag{2.6}$$

A geometric interpretation of the Bar-Shalom-Campo formulas is that the optimal fused covariance $\mathbf{P}_\mathrm{f}$ is located in the intersection of the covariances $\mathbf{P}_\mathrm{A}$ and $\mathbf{P}_\mathrm{B}$ [51].

Covariance Intersection

Covariance intersection is an optimal fusion method if the *estimates* are *correlated* ($\mathbf{P}_\mathrm{AB} = \mathbf{P}_\mathrm{BA} \neq \mathbf{0}$) and the *cross-covariances* are *not known* [54]. A consistent estimate by fusing the two estimates with mean $\mathbf{m}_\mathrm{A}$ and $\mathbf{m}_\mathrm{B}$ is only feasible if a covariance matrix $\mathbf{P}_\mathrm{f}^* \in \mathbb{R}^{n \times n}$ bounds $\mathbf{P}_\mathrm{f}$ in the intersection of $\mathbf{P}_\mathrm{A}$ and $\mathbf{P}_\mathrm{B}$ [55, 51]. The best optimal fused covariance $\mathbf{P}_\mathrm{f}^*$ (which holds for all possible $\mathbf{P}_\mathrm{AB}$ and $\mathbf{P}_\mathrm{BA}$) is a representative from the possible set of optimal ellipsoids that tightly circumscribe the intersection of the two centered ellipsoids [56]. Therefore, $\mathbf{P}_\mathrm{f}^*$ provides the smallest upper bound on the actual error covariance $\mathrm{E}[\tilde{\mathbf{m}}\,\tilde{\mathbf{m}}^T]$. Using set theory, the intersection can be characterized by convex combination of the covariances, and the covariance intersection algorithm [55, 52, 51] is defined by

$$\begin{aligned}
\mathbf{m}_\mathrm{f} &= \mathbf{K}_\mathrm{f}\,\mathbf{m}_\mathrm{A} + \mathbf{L}_\mathrm{f}\,\mathbf{m}_\mathrm{B} && (2.7)\\
\mathbf{P}_\mathrm{f}^* &= \left(\omega\,(\mathbf{P}_\mathrm{A})^{-1} + (1-\omega)\,(\mathbf{P}_\mathrm{B})^{-1}\right)^{-1} && (2.8)\\
\mathbf{K}_\mathrm{f} &= \omega\,\mathbf{P}_\mathrm{f}^*\,(\mathbf{P}_\mathrm{A})^{-1} && (2.9)\\
\mathbf{L}_\mathrm{f} &= (1-\omega)\,\mathbf{P}_\mathrm{f}^*\,(\mathbf{P}_\mathrm{B})^{-1} && (2.10)
\end{aligned}$$

with $\omega \in [0, 1]$. The optimal parameter $\omega^* \in [0, 1]$ can be obtained by numerical optimization [54]

$$\omega^* = \arg\min_{\omega} J(\mathbf{P}_\mathrm{f}^*) \tag{2.11}$$

with an arbitrary strictly monotonically increasing cost function $J(\mathbf{P}_\mathrm{f}^*)$ (e.g., trace or determinant of $\mathbf{P}_\mathrm{f}^*$). There exist suboptimal but fast solutions [57, 58] for ω^*, novel optimization criteria [59], or even closed-form solutions for low dimensional matrices [60]. It has been proofed in [52, 51] that covariance intersection yields the optimal fusion approach for unknwon and not exploitable cross-covariances, however, it yields conservative fusion results [54].

Common Information

Even though the covariance intersection algorithm yields consistent estimates and is proofed to be the optimal fusion method for estimates with unknown cross-covariances, it yields conservative fusion results. In recent works, novel data fusion methods are

introduced which demonstrated to provide less conservative fusion results compared to covariance intersection [61, 43]. Both, *ellipsoidal intersection* [61] and *inverse covariance intersection* [43] exploit a structure of correlation denoted as *common information* [62] which is shared between the sensors and is incorporated in the cross-covariances $\mathbf{P}_{\mathrm{AB}}$ and $\mathbf{P}_{\mathrm{BA}}$. In [61], a decomposition of the estimates to be fused is proposed:

$$\mathbf{m}_{\mathrm{A}} = \mathbf{P}_{\mathrm{A}} \left(\left(\mathbf{P}_{\mathrm{A}}^{\mathrm{I}}\right)^{-1} \mathbf{m}_{\mathrm{A}}^{\mathrm{I}} + \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} \right) \tag{2.12}$$

$$\mathbf{m}_{\mathrm{B}} = \mathbf{P}_{\mathrm{B}} \left(\left(\mathbf{P}_{\mathrm{B}}^{\mathrm{I}}\right)^{-1} \mathbf{m}_{\mathrm{B}}^{\mathrm{I}} + \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} \right) \tag{2.13}$$

with

$$\mathbf{P}_{\mathrm{A}} = \left(\left(\mathbf{P}_{\mathrm{A}}^{\mathrm{I}}\right)^{-1} + \boldsymbol{\Gamma}^{-1} \right)^{-1} \tag{2.14}$$

$$\mathbf{P}_{\mathrm{B}} = \left(\left(\mathbf{P}_{\mathrm{B}}^{\mathrm{I}}\right)^{-1} + \boldsymbol{\Gamma}^{-1} \right)^{-1} \tag{2.15}$$

where the common estimate with mean $\boldsymbol{\gamma} \in \mathbb{R}^{n \times 1}$ and covariance $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$ is shared by the sensors and each sensor A and B contains exclusive information represented by the means $\mathbf{m}_{\mathrm{A}}^{\mathrm{I}} \in \mathbb{R}^{n \times 1}$ and $\mathbf{m}_{\mathrm{B}}^{\mathrm{I}} \in \mathbb{R}^{n \times 1}$ and covariances $\mathbf{P}_{\mathrm{A}}^{\mathrm{I}} \in \mathbb{R}^{n \times n}$ and $\mathbf{P}_{\mathrm{B}}^{\mathrm{I}} \in \mathbb{R}^{n \times n}$, respectively. Each of the errors $\tilde{\mathbf{m}}_{\mathrm{A}}^{\mathrm{I}} \in \mathbb{R}^{n \times 1}$, $\tilde{\mathbf{m}}_{\mathrm{B}}^{\mathrm{I}} \in \mathbb{R}^{n \times 1}$, and $\tilde{\boldsymbol{\gamma}} \in \mathbb{R}^{n \times 1}$ have zero mean and are assumed to be mutually correlated

$$\mathrm{E}[\tilde{\mathbf{m}}_{\mathrm{A}}^{\mathrm{I}} \left(\tilde{\mathbf{m}}_{\mathrm{B}}^{\mathrm{I}}\right)^{T}] = \mathrm{E}[\tilde{\mathbf{m}}_{\mathrm{A}}^{\mathrm{I}} \tilde{\boldsymbol{\gamma}}^{T}] = \mathrm{E}[\tilde{\mathbf{m}}_{\mathrm{B}}^{\mathrm{I}} \tilde{\boldsymbol{\gamma}}^{T}] = \mathbf{0} \tag{2.16}$$

Based on the decomposition of estimates to be fused into common and exclusive information, consistent fusion of two estimates is obtained by a linear combination (similar to (2.1))

$$\mathbf{m}_{\mathrm{f}} = \mathbf{K}_{\mathrm{f}} \mathbf{m}_{\mathrm{A}}^{\mathrm{I}} + \mathbf{L}_{\mathrm{f}} \mathbf{m}_{\mathrm{B}}^{\mathrm{I}} + \mathbf{M}_{\mathrm{f}} \boldsymbol{\gamma} \tag{2.17}$$

which correspondence to fusing three independent estimate means $\mathbf{m}_{\mathrm{A}}^{\mathrm{I}}$, $\mathbf{m}_{\mathrm{B}}^{\mathrm{I}}$, and $\boldsymbol{\gamma}$. The fused estimate mean can be computed similar to Millman's formulas

$$\mathbf{m}_{\mathrm{f}} = \mathbf{P}_{\mathrm{f}} \left(\left(\mathbf{P}_{\mathrm{A}}^{\mathrm{I}}\right)^{-1} \mathbf{m}_{\mathrm{A}}^{\mathrm{I}} + \left(\mathbf{P}_{\mathrm{B}}^{\mathrm{I}}\right)^{-1} \mathbf{m}_{\mathrm{B}}^{\mathrm{I}} + \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} \right) \tag{2.18}$$

$$= \mathbf{P}_{\mathrm{f}} \left(\left(\mathbf{P}_{\mathrm{A}}\right)^{-1} \mathbf{m}_{\mathrm{A}} + \left(\mathbf{P}_{\mathrm{B}}\right)^{-1} \mathbf{m}_{\mathrm{B}} - \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma} \right) \tag{2.19}$$

where the fused covariance is

$$\mathbf{P}_{\mathrm{f}} = \left(\left(\mathbf{P}_{\mathrm{A}}^{\mathrm{I}}\right)^{-1} + \left(\mathbf{P}_{\mathrm{B}}^{\mathrm{I}}\right)^{-1} + \boldsymbol{\Gamma}^{-1} \right)^{-1} = \left(\left(\mathbf{P}_{\mathrm{A}}\right)^{-1} + \left(\mathbf{P}_{\mathrm{B}}\right)^{-1} - \boldsymbol{\Gamma}^{-1} \right)^{-1} \tag{2.20}$$

Ellipsoidal Intersection

The *ellipsoidal intersection* method exploits the representation of the estimates to be fused in (2.12) to (2.15) which implies that for the covariance matrix of the common information $\mathbf{\Gamma}$ the inequalities $\mathbf{P}_\mathrm{A} \leq \mathbf{\Gamma}$ and $\mathbf{P}_\mathrm{B} \leq \mathbf{\Gamma}$ hold, since the covariances of the exclusive information of sensor A and B must satisfy $\left(\mathbf{P}_\mathrm{A}^\mathrm{I}\right)^{-1} = (\mathbf{P}_\mathrm{A})^{-1} - \mathbf{\Gamma}^{-1} \geq \mathbf{0}$ and $\left(\mathbf{P}_\mathrm{B}^\mathrm{I}\right)^{-1} = (\mathbf{P}_\mathrm{B})^{-1} - \mathbf{\Gamma}^{-1} \geq \mathbf{0}$ [61]. Ellipsoidal intersection obtains the best fused result by computing the common information that has the maximum possible $\mathbf{\Gamma}^{-1}$ (highest certainty) which corresponds to the smallest ellipsoid that encloses the covariance ellipsoids belonging to the covariances $\mathbf{P}_\mathrm{A}$ and $\mathbf{P}_\mathrm{B}$ of sensor A and B [61, 63]. It has been proven that the smallest ellipsoid is the Löwner-John ellipsoid [64, 61] and therefore the covariance of the common information can be computed by

$$\mathbf{\Gamma} = \mathbf{T}^{-1}\,\mathbf{D}\,\left(\mathbf{T}^{-1}\right)^T \tag{2.21}$$

where $\mathbf{T} \in \mathbb{R}^{n \times n}$ is a transformation matrix obtained by an eigenvalue decomposition of $\mathbf{P}_\mathrm{A}$ and $\mathbf{P}_\mathrm{B}$, and a diagonal matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ where each diagonal component is obtained by a component-wise maximum $\mathbf{D}_{ii} = \max\{(\mathbf{D}_\mathrm{A})_{ii}, (\mathbf{D}_\mathrm{B})_{ii}\}$ of the jointly transformed ellipsoids $\mathbf{D}_\mathrm{A} = \mathbf{T}\,\mathbf{P}_\mathrm{A}\,\mathbf{T}^T$ and $\mathbf{D}_\mathrm{B} = \mathbf{T}\,\mathbf{P}_\mathrm{B}\,\mathbf{T}^T$ corresponding to sensor A and B. The mean of the common information is computed by

$$\begin{aligned} \boldsymbol{\gamma} \;=\; & \left((\mathbf{P}_\mathrm{A})^{-1} + (\mathbf{P}_\mathrm{B})^{-1} - 2\,\mathbf{\Gamma}^{-1} + 2\,\eta\,\mathbf{I}\right)^{-1} \\ & \left[\left((\mathbf{P}_\mathrm{A})^{-1} - \mathbf{\Gamma}^{-1} + \eta\,\mathbf{I}\right)\,\mathbf{m}_\mathrm{A} + \left((\mathbf{P}_\mathrm{B})^{-1} - \mathbf{\Gamma}^{-1} + \eta\,\mathbf{I}\right)\,\mathbf{m}_\mathrm{B}\right] \end{aligned} \tag{2.22}$$

Even though ellipsoidal intersection obtains less conservative fusion results compared to the covariance intersection method [61], it has been shown that the fused estimate is general not a consistent estimate [43].

Inverse Covariance Intersection

The *inverse covariance intersection* method also exploits the decomposition of the estimates to be fused in (2.12) to (2.15) and aims to find the maximum possible $\mathbf{\Gamma}^{-1}$ (highest certainty of the common information) [43, 54]. In [43] it has been shown that all ellipsoides related to $\mathbf{\Gamma}^{-1}$ which provide a decomposition in (2.12) to (2.15) are bounded by the intersection of the inverse covariances $(\mathbf{P}_\mathrm{A})^{-1}$ and $(\mathbf{P}_\mathrm{B})^{-1}$ of sensor A and B. Hence, the maximum possible ellipsoid related to $\mathbf{\Gamma}^{-1}$ is obtained by covariance intersection of the inverse covariances $(\mathbf{P}_\mathrm{A})^{-1}$ and $(\mathbf{P}_\mathrm{B})^{-1}$, and is computed by

$$\mathbf{\Gamma}^{-1} = \left(\omega\,\left((\mathbf{P}_\mathrm{A})^{-1}\right)^{-1} + (1-\omega)\,\left((\mathbf{P}_\mathrm{B})^{-1}\right)^{-1}\right)^{-1} \tag{2.23}$$

$$= \left(\omega\,\mathbf{P}_\mathrm{A} + (1-\omega)\,\mathbf{P}_\mathrm{B}\right)^{-1} \tag{2.24}$$

where $\omega \in [0,1]$. As in the covariance intersection algorithm, the optimal parameter ω can be obtained by numerical optimization with an arbitrary strictly monotonically increasing cost function of the fused covariance

$$\mathbf{P}_\mathrm{f} = \left((\mathbf{P}_\mathrm{A})^{-1} + (\mathbf{P}_\mathrm{B})^{-1} - (\omega\,\mathbf{P}_\mathrm{A} + (1-\omega)\,\mathbf{P}_\mathrm{B})^{-1} \right)^{-1} \tag{2.25}$$

(e.g., trace or determinant of $\mathbf{P}_\mathrm{f}$). Based on $\mathbf{P}_\mathrm{f}$, the fused estimate mean $\mathbf{m}_\mathrm{f}$ is computed by [43, 54]

$$\begin{aligned}
\mathbf{m}_\mathrm{f} &= \mathbf{K}_\mathrm{f}\,\mathbf{m}_\mathrm{A} + \mathbf{L}_\mathrm{f}\,\mathbf{m}_\mathrm{B} && (2.26)\\
\mathbf{K}_\mathrm{f} &= \mathbf{P}_\mathrm{f}\left((\mathbf{P}_\mathrm{A})^{-1} - \omega\,(\omega\,\mathbf{P}_\mathrm{A} + (1-\omega)\,\mathbf{P}_\mathrm{B})^{-1} \right) && (2.27)\\
\mathbf{L}_\mathrm{f} &= \mathbf{P}_\mathrm{f}\left((\mathbf{P}_\mathrm{B})^{-1} - (1-\omega)\,(\omega\,\mathbf{P}_\mathrm{A} + (1-\omega)\,\mathbf{P}_\mathrm{B})^{-1} \right) && (2.28)
\end{aligned}$$

where $\omega \in [0,1]$. Compared to ellipsoidal intersection, it has been proven that the inverse covariance intersection algorithm yields a consistent estimate and a less conservative fusion result compared to covariance intersection [43]. However, a disadvantage compared to covariance intersection is that the inverse covariance intersection algorithm is tailored to a specific correlation structure between the sensor estimates (cf. (2.12) to (2.15)) which is the common information [54] shared between sensor A and B.

Kalman Filter and Bayesian Smoothing

In many applications, sensors are the source of input data in a multi-sensor data fusion system [42, 40]. The actual sensor element is a physical element which directly interacts with the environment and can be any device sensitive to a physical property or environmental attribute in order to measure, e.g., light intensity, temperature, pressure, or motion. However, a sensor can only measure the physical property with a sensor specific certainty which means that the measurements include uncertainty. Also, the sensor is directly affected by environmental factors leading to systematic errors in the measurement and thus, a measurement is only an estimate of the system state (e.g., physical property). However, application specific knowledge of the environment or physical property is available and can be integrated in a multi-sensor system.

Bayesian inference is a powerful method for data fusion where information about the estimate (measurement) and the environment with their physical property are fused in order to obtain an estimate of the unknown quantity of interest (system state). Such a quantity can be, e.g., the position and velocity of an object in object tracking or the electrical activity in the brain. In many applications the unknown quantity might not be fully observable and therefore is not directly measurable by any sensor. In general, the unknwon quantity is described by an hidden state $\mathbf{x}_t \in \mathbb{R}^{n\times 1}$ and the measurement by $\mathbf{y}_t \in \mathbb{R}^{m\times 1}$ at time point t. Bayesian inference has the aim to compute the joint posterior distribution of all the (hidden) states $\mathbf{x}_{0:t}$ given all

the measurements $\mathbf{y}_{1:t}$ up to time point t by using Bayes' theorem [65]

$$p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right) = \frac{p\left(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}\right) p\left(\mathbf{x}_{0:t}\right)}{p\left(\mathbf{y}_{1:t}\right)} \tag{2.29}$$

where $p\left(\mathbf{x}_{0:t}\right)$ is the prior distribution of the state vector at time point $t = 0$, $p\left(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}\right)$ is the likelihood model for the measurements, and $p\left(\mathbf{y}_{1:t}\right)$ is a normalization constant defined by

$$p\left(\mathbf{y}_{1:t}\right) = \int p\left(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}\right) p\left(\mathbf{x}_{0:t}\right) \, \mathrm{d}\mathbf{x}_{0:t} \tag{2.30}$$

In this case, Bayesian inference is used for *single-sensor multi-temporal data fusion* (fusion of data from a single sensor over multiple time points) to compute the posterior distribution. However, computing the full posterior distribution $p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right)$ in (2.29) is not tractable since when new measurements are obtained a recomputation is required. Instead of computing the full posterior distribution, *Bayesian filtering* relaxes the estimation of hidden states by computing only the marginal distribution $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ of the current state $\mathbf{x}_t$ given the current and all previous measurements $\mathbf{y}_{1:t}$ [65]. Two properties are assumed, namely the *Markov property of states* and the *conditional independence of measurements*. The Markov property of states assumes that a state $\mathbf{x}_t$ at time point t given $\mathbf{x}_{t-1}$ at time point $t-1$ is independent of all states and measurements until time point $t-1$ [65, 66]

$$p\left(\mathbf{x}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}\right) = p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) \tag{2.31}$$

Conversely, the past state is also independent of the future state when the present state is given. The conditional independence of measurements assumes that the current measurement $\mathbf{y}_t$ given the current state $\mathbf{x}_t$ is conditionally independent of all states and measurements until time point t [65]

$$p\left(\mathbf{y}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t-1}\right) = p\left(\mathbf{y}_t|\mathbf{x}_t\right) \tag{2.32}$$

Due to the Markov property, the joint distribution of the current state $\mathbf{x}_t$ and the previous state $\mathbf{x}_{t-1}$ given $\mathbf{y}_{1:t-1}$ is

$$\begin{aligned} p\left(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) &= p\left(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) && (2.33) \\ &= p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) && (2.34) \end{aligned}$$

Integrating the joint distribution $p\left(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right)$ over $\mathbf{x}_{t-1}$ gives the marginal distribution $p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right)$ which is denoted as the Chapman–Kolmogorov equation

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right) = \int p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) \mathrm{d}\mathbf{x}_{t-1} \tag{2.35}$$

The probability distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is denoted as dynamic model describing the temporal evolution of the state. The Chapman–Kolmogorov equation is a recursive equation for computing the predicted distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and denotes the *state prediction* step of Bayesian filtering. The step of *update estimation* is defined by computing the posterior distribution of the current state $\mathbf{x}_t$ given all previous measurements $\mathbf{y}_{1:t}$ and can be obtained by Bayes' theorem in conjunction with the property of conditional independence of measurements [65]

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1})\ p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{Z_k} \tag{2.36}$$

$$= \frac{p(\mathbf{y}_t|\mathbf{x}_t)\ p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{Z_k} \tag{2.37}$$

with

$$Z_k = \int p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{y}_{1:t-1})\, \mathrm{d}\mathbf{x}_t \tag{2.38}$$

$p(\mathbf{y}_t|\mathbf{x}_t)$ is the measurement model and describes the distribution of the measurement $\mathbf{y}_t$ given the state $\mathbf{x}_t$. For time point $t = 0$, no measurement can be obtained which is denoted as *initialization step* and the prior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = p(\mathbf{x}_0) \tag{2.39}$$

and contains prior information about the hidden state. The *Kalman filter* is the closed form solution to the Bayesian filtering equations and computes the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \sim \mathcal{N}(\mathbf{x}_t; \mathbf{m}_t, \mathbf{P}_t)$ with mean $\mathbf{m}_t \in \mathbb{R}^{n\times 1}$ and covariance $\mathbf{P}_t \in \mathbb{R}^{n\times n}$ [67, 65, 66, 68], and assumes linear and Gaussian models for the dynamic and measurement model:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim \mathcal{N}(\mathbf{x}_t; \mathbf{F}\,\mathbf{m}_{t-1}, \mathbf{Q}) \tag{2.40}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) \sim \mathcal{N}(\mathbf{y}_t; \mathbf{H}\,\mathbf{m}_t, \mathbf{R}) \tag{2.41}$$

The dynamic model includes the transition matrix $\mathbf{F} \in \mathbb{R}^{n\times n}$ and the covariance matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ which reflects the uncertainty of the dynamic model. The measurement model $\mathbf{y}_t = \mathbf{H}\,\mathbf{m}_t$ contains the measurement matrix $\mathbf{H} \in \mathbb{R}^{m\times m}$ and the covariance matrix $\mathbf{R} \in \mathbb{R}^{m\times m}$ which reflects the uncertainty of the measurement model. The mean $\hat{\mathbf{m}}_t \in \mathbb{R}^{n\times 1}$ and covariance $\hat{\mathbf{P}}_t \in \mathbb{R}^{n\times n}$ of the predicted state $\hat{\mathbf{x}}_t \in \mathbb{R}^{n\times 1}$ is computed using the state estimate $\mathbf{x}_{t-1}$ at time point $t-1$:

$$\hat{\mathbf{m}}_t = \mathbf{F}\,\mathbf{m}_{t-1} \tag{2.42}$$

$$\hat{\mathbf{P}}_t = \mathbf{F}\,\mathbf{P}_{t-1}\,\mathbf{F}^T + \mathbf{Q} \tag{2.43}$$

The update estimate of the mean $\mathbf{m}_t \in \mathbb{R}^{n\times 1}$ and covariance $\mathbf{P}_t \in \mathbb{R}^{n\times n}$ of the state

$\mathbf{x}_t$ is computed by

$$\mathbf{m}_t = \hat{\mathbf{m}}_t + \mathbf{K}\left(\mathbf{y}_t - \mathbf{H}\,\hat{\mathbf{m}}_t\right) \tag{2.44}$$
$$\mathbf{P}_t = \left(\mathbf{I} - \mathbf{K}\,\mathbf{H}\right)\hat{\mathbf{P}}_t \tag{2.45}$$

with the Kalman gain $\mathbf{K} \in \mathbb{R}^{n \times m}$

$$\mathbf{K} = \hat{\mathbf{P}}_t\,\mathbf{H}^T\left(\mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{2.46}$$

and the identity matrix $\mathbf{I} \in \mathbb{R}^{n \times n}$.

Bayesian smoothing has the aim to compute the posterior probability distribution $p\left(\mathbf{x}_t|\mathbf{y}_{1:T}\right)$ conditionally on all measurements $\mathbf{y}_{1:T} \in \mathbb{R}^{m \times T}$ up to time point T, where $T > t$ is a future time point [69, 65]:

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:T}\right) = p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) \int \frac{p\left(\mathbf{x}_{t+1}|\mathbf{x}_t\right)\, p\left(\mathbf{x}_{t+1}|\mathbf{y}_{1:T}\right)}{p\left(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}\right)}\,\mathrm{d}\mathbf{x}_{t+1} \tag{2.47}$$

Note that the distribution $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ is the filtering distribution of time point t. A closed form solution to Bayesian smoothing is the Rauch–Tung–Striebel smoother [70] or a two-filter smoothing approach introduced by Fraser and Potter [71] with two independent Kalman filters running forward and backward in time.

2.3 Deep Neural Networks for Biomedical Computer Vision

In this section, fundamentals of deep learning are briefly introduced and basic network architectures used for biomedical computer vision (e.g., particle detection and tracking, cell segmentation and classification) are reviewed.

Artificial neural networks (ANN) are inspired by biological neural networks of brains (e.g., perceptron [72]) and consist of a collection of connected artificial neurons which are able to receive a signal, process it, and transmit signals to other connected neurons. In an ANN, q neurons form a layer which is parameterized by weights $\mathbf{W} \in \mathbb{R}^{p \times q}$ and bias $\mathbf{b} \in \mathbb{R}^q$, where p is the number of neurons of the previous layer. The activation $y_i \in \mathbb{R}$ of the i-th neuron in a layer is obtained by a weighted sum of the neuron activations of the previous layer $\mathbf{x} \in \mathbb{R}^{p \times 1}$ using the weights $\mathbf{w}_i \in \mathbb{R}^{p \times 1}$ and the bias $b_i \in \mathbb{R}$, given as input to an activation function $\sigma(\cdot)$ [73, 74, 75]:

$$y_i = \sigma\left(\left(\mathbf{w}_i\right)^T\mathbf{x} + b_i\right) \tag{2.48}$$

For $\sigma(\cdot)$, linear functions (e.g., identity function) or non-linear functions (e.g., sigmoid, hyperbolic tangent function) are used [74]. Loss functions are used during training phase, where ANNs learn to perform a defined task by changing the network

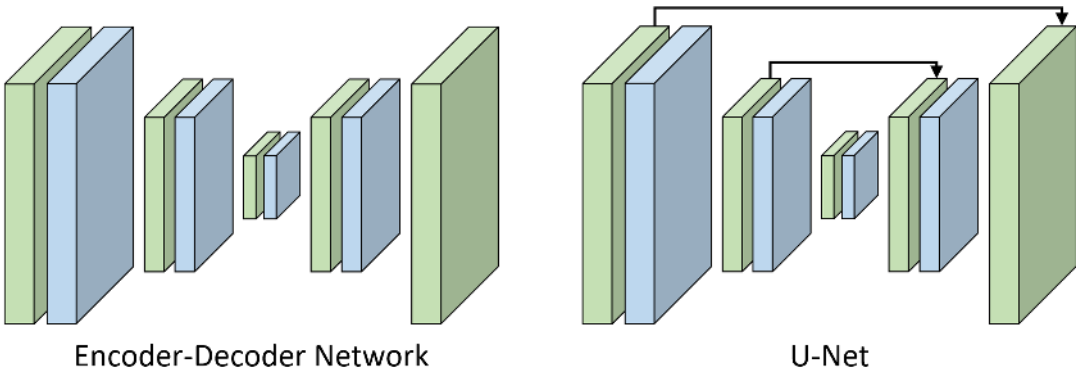


Figure 2.7: Simplified schematic of the Encoder-Decoder Network and U-Net architecture. Green boxes represent convolutional layers and blue layers are pooling layers. The black arrows in the U-Net show the skipping connections between opposed convolutional layers.

parameters (e.g., weights) using optimization algorithms [73]. If the ANN consists of multiple consecutive hidden layers, the network is denoted as *Deep Neural Network* (DNN). With an increased number of neurons, feed-forward neural networks are able to approximate any continuous function in a compact set $\Omega \in \mathbb{R}^n$ with an decreasing error bound [75].

Convolutional Neural Networks

For microscopy applications such as image segmentation [76, 77, 78, 79], object detection [80, 81, 82, 83, 84, 85], and object classification [86, 87], *convolutional neural networks* (CNNs) have been employed. CNNs are based on convolutional layers where each layer is a series of 2D learnable filters which perform convolution operations on images and output a series of feature maps $\mathbf{y}$ [74, 88]. Therefore, a convolutional layer is represented by a convolution ($*$) with filter $\mathbf{W} \in \mathbb{R}^{k \times k \times p \times q}$ subjected to an element-wise non-linear activation function $\sigma(\cdot)$:

$$\mathbf{y} = \sigma\left(\mathbf{W} * \mathbf{x} + \mathbf{b}\right) \tag{2.49}$$

where $k \times k$ is the window size of the convolutional kernel, p the number of input feature maps, and q the number of output feature maps. Note that in many CNN implementations, flipping of the convolutional kernel is omitted which results in computing the cross-correlation instead of convolution [89]. The feature maps can be reduced in dimensionality by applying pooling operators (e.g., max pooling, average pooling) and make the representation approximately invariant to small shifts and distortions which can be helpful for classification when only the presence of features is important and not their position [73, 89].

Encoder-Decoder architectures are composed of a contracting (encoder) and an expanding (decoder) path [88]. A schematic of the Encoder-Decoder architecture is shown in Fig. 2.7. The encoder reduces the dimensionality of the input through a

sequence of convolutional and pooling layers. Thus, the information of the input image is encoded into features of low dimensionality which is a compression of the input image. The decoder exploits the encoded input information in order to reconstruct the original image or determines a representation of an image (e.g., binary image for segmentation). Upsampling within the decoder path can either be performed by techniques which are predefined and not learnable from the data such as nearest neighbor interpolation, bi-linear interpolation, bed of nails, and max unpooling, or can be done by transposed convolutions with learnable parameters [90]. However, transposed convolution can produce uneven overlap at some parts of the feature maps causing checkerboard artifacts in the resulting feature maps [91]. Transposed convolution has been used in the Fully Convolutional Network (FCN) proposed for semantic segmentation [90]. In comparison, the *U-Net* displayed in Fig. 2.7 exploits additionally long-range skip connections between directly opposed convolutional layers of the contracting and expanding path [76]. This has the advantage that finer image details are maintained during upsampling within the expanding path.

A CNN can be trained (supervised) by performing a specific task (e.g., detection, classification) where the network is shown an image and produces an output (e.g, an image, binary image, vector of scores) which is compared to the desired output. The error between the predicted network output and the desired output is measured by an application specific objective function (loss function). The derivative of the loss function is exploited by propagating it backwards through the network (*Backpropagation*) by applying the chain rule for derivatives in order to compute a (negative) gradient for each weight indicating its adjustment [73]. In conjunction with stochastic gradient descent (SGD), the weights of the CNN can be updated in order to obtain the smallest prediction error of the network measured by the loss function. Often used loss functions for CNNs are *pixel-wise cross-entropy* and *soft Dice* which are utilized for classification or binary image segmentation (binary classification between foreground and background pixels). For classification, the *pixel-wise cross-entropy* (CE) examines each pixel individually by comparing the class prediction vector $\mathbf{x}_i \in \mathbb{R}^n$ to a target vector $\mathbf{y}_i \in \mathbb{R}^n$ which encodes the expected class as on-hot vector. Averaging the pixel-wise CE over all M image pixels by computing

$$\mathcal{L}_{\text{CE}} = \frac{1}{M} \sum_{i=1}^{M} -\mathbf{y}_i \log\left(\mathbf{x}_i\right) \tag{2.50}$$

performs equal learning for all pixels in an image. However, in the case of class imbalance, $\mathcal{L}_{\text{CE}}$ has the disadvantage that the training is dominated by the most prevalent class. A solution is provided in [90] where the loss is weighted for each output channel to avoid class inbalance. For binary image segmentation, the *soft Dice* loss is a smooth approximation of the Dice coefficient and can be computed

by [92]

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{M} y_i\, x_i}{\sum_{i=1}^{M} y_i^2 + \sum_{i=1}^{M} x_i^2} \tag{2.51}$$

using the pixel prediction probability $x_i \in [0, 1]$ and the pixel target probability $y_i \in [0, 1]$. An advantage of the soft Dice loss is that the denominator incorporates the quantity of activations which avoids the challenges of class imbalance.

During training of DNNs, the distribution of the activations of each layer is changing (covariate shift [93]) so that the layers should be adjusted in each training step. This change can be large since in deep networks many layers are updated simultaneously and small changes in one layer can have large effects on subsequent layers. Thus, a small learning rate is needed and slows down training. DNN training can be improved by using *Batch Normalization* [94], which normalizes the distribution of activations of each layer to zero mean and unit variance calculated by the mean and variance of the training mini-batch. To maintain the layer representation, the normalized activation distributions of each layer are shifted and scaled by learnable parameters [94, 89]. Due to Batch Normalization, the DNN can be trained with a much higher learning rate leading to accelerated training and the normalization also acts as a regularizer which supersedes Dropout in some cases [94]. In comparison, *Instance Normalization* normalizes the distribution of each activation using the mean and variance of each feature map (channel) of each instance of a batch individually [95].

Additional, challenges of training a DNN are the vanishing gradient problem [96] and the problem of degradation, where the network suffers from a higher training error when adding more layers [97]. A solution is provided in [97] by introducing *residual learning* which is able to skip layers by connecting the output of one layer with the input of an earlier layer. This reduces the effect of gradient vanishing since there are fewer layers to propagate through. Another advantage is that during the initial training phase, the network is simplified due to the skipping connections leading to an accelerated training.

Recurrent Neural Networks

To process sequential data (e.g., image series) the *Recurrent Neural Network* (RNN) was developed. The RNN passes information between consecutive time points by giving a hidden state $\mathbf{h}_t \in \mathbb{R}^p$ which is the output of a non-linear mapping from $\mathbf{x}_t \in \mathbb{R}^n$ and the previous hidden state $\mathbf{h}_{t-1}$ as input to the RNN block in order to produce an output $\mathbf{o}_t \in \mathbb{R}^p$ [74]. Thus, the RNN forms a cycled computation graph whose gradients are calculated via unfolding which is illustrated in Fig. 2.8. It has been stated in [98] that an RNN suffers from similar challenges in network training compared to regular DNN since the gradient from the output is required to be backpropagated over time which makes the RNN inherently deep.

Implementations of RNN units, which are specialized memory units, are the *Long*

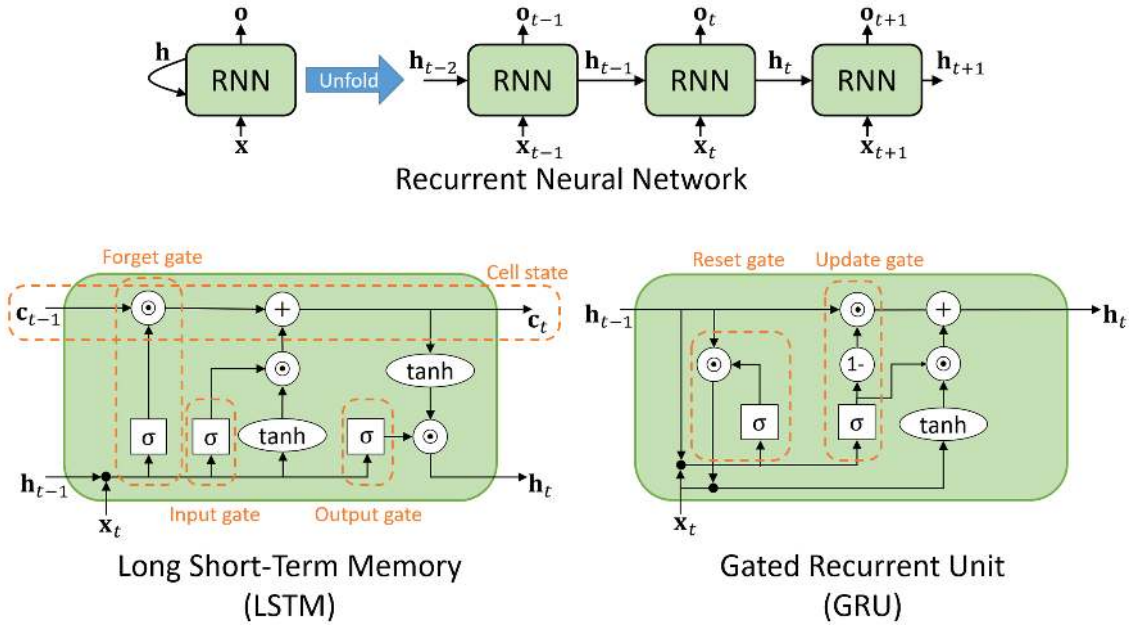


Figure 2.8: Illustration of unfolding the RNN with observations $\mathbf{x}$ as well as the LSTM and GRU architecture as popular implementations of RNN units. For the LSTM and GRU, σ indicates the sigmoid function, tanh is the hyperbolic tangent, $\odot$ denotes the Hadamard product (element-wise multiplication), $+$ signifies the element-wise addition, and $\bullet$ is vector concatenation.

Short-Term Memory (LSTM) [99] and the *Gated Recurrent Unit* (GRU) [100]. The LSTM and GRU are illustrated in Fig. 2.8. To avoid gradient vanishing and exploding in RNNs, the LSTM has been proposed which exploits core components to regulate the flow of information. In principle, the gates are able to learn which information in a sequence is important to be kept or being discarded. The previous hidden state $\mathbf{h}_{t-1}$, the current input $\mathbf{x}_t$, and the previous cell state $\mathbf{c}_{t-1} \in \mathbb{R}^m$ are given as input to the LSTM unit. The forget gate discards previous and current information by using $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$ as input to a sigmoid function and therefore regulates the reset of the cell state (memory). The input gate decides which input information is relevant and added to the cell state $\mathbf{c}_t$. The output gate regulates the output activation which is exploited for calculating the new hidden state $\mathbf{h}_t$. Compared to LSTM, GRU does not exploit a cell state and uses instead the hidden state to regulate the memory of the unit (see Fig. 2.8). The GRU reset gate is used in order to decide which previous information can be discarded, whereas the update gate is similar to the forget and input gate. The output of the GRU is only the hidden state $\mathbf{h}_t$.

2.4 Methods for Particle Tracking in Microscopy Images

In this section, fundamentals of particle tracking in microscopy images and main approaches are described. We distinguish between *deterministic*, *probabilistic*, and *deep learning* approaches.

2.4.1 Deterministic Approaches

Deterministic approaches follow a two step-paradigm consisting of particle detection and correspondence finding (e.g., [101, 102, 103, 47, 104]). Some approaches determine correspondences by minimal cost paths in a spatial-temporal volume (e.g., [105, 106, 107]). In this section, first, particle detection approaches are discussed and second, motion correspondence approaches are introduced.

Particle Detection

Particle detection methods can be subdivided into single-scale and multi-scale approaches. Performance evaluations of different methods were carried out in [108, 109, 110]. *Single-scale* approaches exploit information from only one image scale (e.g., [111, 112, 113, 106, 114, 115, 116]). The spot-enhancing filter (SEF) [106] enhances particles while reducing noise and applies a Laplacian-of-Gaussian (LoG) filter. The standard deviation of the LoG needs to be adjusted according to the size of the particles. Model fitting approaches that fit Gaussian models to the image intensities [113, 47, 115] and detection approaches based on grayscale morphology using h-dome transformation were also proposed [117, 114, 116]. These approaches assume a relatively simple appearance model of particles (Gaussian functions). Top-hat filters for particle detection [111, 112] exploit the mean intensities in local neighborhoods. The size of the local neighborhoods needs to be adjusted according to the size of the particles. *Multi-scale* approaches use information from multiple image scales to detect particles with varying size (e.g., [118, 119, 120, 84, 83, 14]). [118] proposed a detection method based on Wavelet transformation which employs wavelet multi-scale products. [120] proposed a multi-scale SEF approach (MSSEF) which iteratively exploits multiple scales. [119] introduced ATLAS which is an adaptive thresholding scheme with autoselected scale selection. All these multi-scale detection approaches use only a single detection method.

Correspondence Finding

Once particles are detected, a set of N measurements $\mathcal{Y}_t = \{\mathbf{y}_t^1, \mathbf{y}_t^2, \ldots, \mathbf{y}_t^N\}$ is obtained for an image at time point t. A measurement $\mathbf{y}_t^k \in \mathbb{R}^{m \times 1}$ represents a particle k in frame t typically defined by its position $\mathbf{p}_t^k = (x_t^k, y_t^k, z_t^k)$. However, errors in particle detection lead to missed objects or false positive detected objects in $\mathcal{Y}_t$.

Motion correspondence algorithms determine which two measurements $\mathbf{y}_t^k$ and $\mathbf{y}_{t+1}^l$ originate from the same object and therefore define a track $\mathcal{I} = \{(t, k, \mathbf{p}^k), (t+1, l, \mathbf{p}^l)\}$ consisting of triples with time point, object index, and position [101, 121]. In order to ensure obtaining feasible correspondences, two assumptions are typically made: (i) A particle can generate at most one detection, and (ii) a detection can be associated with at most one particle [121]. Using a distance function $d(\cdot)$ to quantify the degree of correspondence, most frequently a *nearest neighbor* approach is used, which determines correspondences by considering one measurement at time point t and all measurements at time point $t+1$ by optimizing [121]:

$$\min_{\mathbf{y}_{t+1}^l} d(\mathbf{y}_t^k, \mathbf{y}_{t+1}^l) \quad \text{with} \quad d(\mathbf{y}_t^k, \mathbf{y}_{t+1}^l) = ||\mathbf{p}_t^k - \mathbf{p}_{t+1}^l||_2 \tag{2.52}$$

Gating can improve motion correspondence by considering only the most likely associations which have a distance less or equal to a distance threshold $d_{\max}$ [47, 121]. The nearest neighbor approach can yield conflicting correspondences when two measurements $\mathbf{y}_t^k$ and $\mathbf{y}_t^m$ (with $k \neq m$) are associated with $\mathbf{y}_{t+1}^l$, since only a single measurement of time point t is considered in (2.52) without taking into account the remaining measurements at time point t.

To consider all measurements of time point t, the motion correspondence problem can be extended by using a weighted bipartite graph $G = (V, E)$ with vertices $V = \{\mathcal{Y}_t, \mathcal{Y}_{t+1}\}$. An edge $e_{k,l} = (\mathbf{y}_t^k, \mathbf{y}_{t+1}^l) \in E$ represents a possible correspondence between measurements $\mathbf{y}_t^k$ and $\mathbf{y}_{t+1}^l$, and is weighted by $w_{k,l} = d(\mathbf{y}_t^k, \mathbf{y}_{t+1}^l)$. A set of valid correspondences ($e_{k,l} \in \mathcal{M}$) is denoted as matching set $\mathcal{M}$, if the set consists of edges with no shared vertices (see the assumptions for feasible correspondences mentioned above). A valid edge $e_{k,l} \in \mathcal{M}$ is indicated by a binary assignment variable $a_{k,l} = 1$ compared to non-valid edges with $a_{k,l} = 0$. Therefore, motion correspondence finding between two sets of measurements obtained from two consecutive frames can be formulated as *assignment problem* [121]:

$$\begin{aligned} &\min_{a_{k,l}} \sum_{k=0}^{K} \sum_{l=0}^{L} a_{k,l}\, w_{k,l} \\ &\text{subject to} \\ &\sum_{k=0}^{K} w_{k,l} = 1, \quad l = 1, \ldots, L \quad \text{and} \quad \sum_{l=0}^{L} w_{k,l} = 1, \quad k = 1, \ldots, K \end{aligned} \tag{2.53}$$

where $w_{k,l}$ represents the cost of assigning measurement $\mathbf{y}_t^k$ to $\mathbf{y}_{t+1}^l$. Since the assignment problem considers all correspondences of all measurements it is denoted as *global nearest neighbor* approach [47]. For $a_{k,0} = 1$ and $a_{0,l} = 1$ in (2.53) the measurements are assigned to dummy particle detections. Assigning a measurement $\mathbf{y}_t^k$ to a dummy detection represents track termination, whereas assigning $\mathbf{y}_{t+1}^l$ to a dummy detection corresponds to track initialization. Since there are no constraints on the dummy detections, multiple measurements can be assigned to a dummy detection.

The optimal assignments can be obtained by using optimization algorithms for (2.53), such as the Hungarian algorithm [122], the Munkres algorithm which is an extended version of the Hungarian algorithm [123], the Jonker-Volgenant shortest augmenting path algorithm [124], or algorithms based on a graph-theoretical approach for the transportation problem [102, 47].

Since missing detections can lead to too early track termination, or false positive detections can cause false track initialization or false track continuation, the assignment problem in (2.53) for two time points can be extended to the *generalized multi-frame assignment problem* for multiple time points [121]:

$$\begin{aligned}
&\min_{a_{k,l,\ldots,m,n}} \sum_{k=0}^{K} \sum_{l=0}^{L} \cdots \sum_{m=0}^{M} \sum_{n=0}^{N} a_{k,l,\ldots,m,n}\, w_{k,l,\ldots,m,n} \qquad (2.54)\\
&\text{subject to}\\
&\sum_{k=0}^{K} \cdots \sum_{m=0}^{M} w_{k,l,\ldots,m,n} = 1, \quad m = 1, \ldots, M\\
&\qquad\qquad \vdots\\
&\sum_{l=0}^{L} \cdots \sum_{n=0}^{N} w_{k,l,\ldots,m,n} = 1, \quad k = 1, \ldots, K
\end{aligned}$$

For particle tracking, the generalized multi-frame assignment problem was used in [104] and the optimal assignments are determined by an integer programming framework where tracks are extended by gated detections within a sliding window of multiple frames.

Compared to the generalized multi-frame assignment problem, the *two-step linear assignment procedure* (LAP) introduced in [103] determines the optimal assignments by combinatorial optimization in two steps and also exploits measurements obtained from multiple time points. In the first step, LAP assigns measurements in consecutive frames by solving the assignment problem in (2.53) using a distance-based cost function. In the second step, the obtained tracklets are linked in order to close gaps and deal with events of track splitting and merging by using application specific closing, splitting, and merging cost function.

The approaches in [105, 106, 107] formulate tracking as finding trajectories in a 3D volume, where a temporal sequence of 2D images is represented by a 3D spatial-temporal volume. Motion correspondence is solved by minimizing energy functions which consider the image data as well as prior constraints on the particle displacements on a frame-to-frame level.

2.4.2 Probabilistic Approaches

Deterministic approaches are computationally efficient but do not take into account uncertainties. Often, they suffer from low signal-to-noise ratio (SNR) for detection or challenging object constellations for correspondence finding (e.g., high object density, spurious objects). In comparison, probabilistic tracking approaches consider spatial-temporal uncertainties and are formulated within a Bayesian framework (e.g., [125, 46, 48, 120, 49, 126]). Based on Bayesian sequential estimation, the approaches robustly determine the position of a particle from the posterior distribution using noisy particle detections. Probabilistic approaches for correspondence finding determine a probability of possible assignments between particle predictions and measurements. Below, we review such approaches including *probabilistic data association*, *joint probabilistic data association*, and *multiple hypothesis tracking*.

Probabilistic Data Association

The basic idea of the *probabilistic data association* (PDA) filter [45, 127] is to exploit a decomposition of state estimation based on the origin of each measurement in $\mathcal{Y}_t$. Each measurement $\mathbf{y}_t^i$ in $\mathcal{Y}_t$ is examined if the measurement may have originated from a particle k by using an elliptical region denoted as validation region (gating) [45, 127]

$$\mathcal{V}(t,\gamma) = \left\{ \mathbf{y}_t^i \;\middle|\; \left(\mathbf{y}_t^i - \hat{\mathbf{y}}\right)^T \left(\mathbf{S}_t\right)^{-1} \left(\mathbf{y}_t^i - \hat{\mathbf{y}}\right) \leq \gamma^2 \right\} \tag{2.55}$$

which is centered around the predicted measurement $\hat{\mathbf{y}} \in \mathbb{R}^{m \times 1}$ of particle k, and $\mathbf{S}_t = \mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H} + \mathbf{R}_t$ is the innovation covariance obtained from the predicted covariance $\hat{\mathbf{P}}_t \in \mathbb{R}^{n \times n}$ (cf. Section 2.2). The parameter $\gamma^2 \in \mathbb{R}$ is the gate threshold. All $\mathbf{y}_t^i$ within the validation region form a set $\mathcal{Y}_t^k = \{\mathbf{y}_t^i\}_{i=1}^K$ of validated measurements associated with particle k. A sequence of sets containing validated measurements up to time point t is defined as $\mathcal{Y}_{1:t} = \{\mathcal{Y}_t^k\}_{t=1}^t$. The PDA filter computes the updated estimate of the mean by [45]:

$$\mathbf{m}_t \;=\; \sum_{i=0}^{K} \mathrm{E}[\mathbf{x}_t | A_i, \mathcal{Y}_{1:t}]\, P(A_i | \mathcal{Y}_{1:t}) = \sum_{i=0}^{K} \mathbf{m}_{i,t}\, \beta_{i,t} \tag{2.56}$$

where $\mathbf{m}_{i,t} \in \mathbb{R}^{n \times 1}$ is the mean update estimate conditioned that the measurement $\mathbf{y}_t^i$ associated to particle k has an association probability $\beta_{i,t} \in [0,1]$. A_i is an association event indicating whether the measurement $\mathbf{y}_t^i$ has originated from the particle. Using a Kalman filter approach, the closed form solution for the update estimation of the PDA filter is [127]

$$\mathbf{m}_t \;=\; \hat{\mathbf{m}}_t + \mathbf{K}\,\mathbf{v}_t \tag{2.57}$$

$$\mathbf{P}_t \;=\; \beta_{0,t}\,\hat{\mathbf{P}}_t + (1 - \beta_{0,t})\left(\hat{\mathbf{P}}_t - \mathbf{K}\,\mathbf{S}_t\,\mathbf{K}^T\right) + \tilde{\mathbf{P}}_t \tag{2.58}$$

where $\mathbf{v}_t \in \mathbb{R}^{m \times 1}$ is the combined innovation

$$\mathbf{v}_t = \sum_{i=1}^{K} \beta_{i,t}\, \mathbf{v}_{i,t} \tag{2.59}$$

$$\mathbf{v}_{i,t} = \mathbf{y}_t^i - \hat{\mathbf{y}}_t \tag{2.60}$$

and the association probability $\beta_{0,t}$ represents that none of the measurements is correct and therefore $\mathbf{m}_t = \hat{\mathbf{m}}_t$. $\tilde{\mathbf{P}}_t \in \mathbb{R}^{n \times n}$ is the spread of the innovation term [45], which represents the effect of the measurement origin uncertainty (similar to the spread of the means term in a mixture of probability density functions [128]).

The association probabilities $\beta_{i,t}$ for $i = 0, 1, \ldots, K$ are determined based on K measurements (number of measurements in $\mathcal{V}(t, \gamma)$) and their location by [45]

$$\beta_{i,t} = P(A_i | \mathcal{Y}_{1:t}) = P(A_i | \mathcal{Y}_t^k, K, \mathcal{Y}_{1:t-1}) \tag{2.61}$$

$$= \frac{1}{c} P(A_i | K, \mathcal{Y}_{1:t-1})\, p(\mathcal{Y}_t^k | A_i, K, \mathcal{Y}_{1:t-1}) \tag{2.62}$$

$p(\mathcal{Y}_t^k | A_i, K, \mathcal{Y}_{1:t-1})$ is the joined density of $\mathcal{Y}_t^k$ conditioned on A_i and is computed as product of the Gaussian probability density function of the target originated measurements and the independent identically distributed measurements originated from clutter (false measurements) with uniform spatial distribution

$$p(\mathcal{Y}_t^k | A_i, K, \mathcal{Y}_{1:t-1}) = \begin{cases} V_t^{-K+1}\, P_{\mathrm{G}}^{-1}\, \mathcal{N}(\mathbf{y}_t^i; \hat{\mathbf{y}}_t, \mathbf{S}_t) & i = 1, \ldots, K \\ V_t^{-K} & i = 0 \end{cases} \tag{2.63}$$

$P_{\mathrm{G}}^{-1} \in \mathbb{R}$ is a factor which restricts $\mathcal{N}(\mathbf{y}_t^i; \hat{\mathbf{y}}_t, \mathbf{S}_t)$ to the validation region $\mathcal{V}(t, \gamma)$ and $V_t \in \mathbb{R}^m$ is the volume of $\mathcal{V}(t, \gamma)$. The probabilities of the assignment A_i are only conditioned on the number of measurements within $\mathcal{V}(t, \gamma)$ and computed by [45]

$$P(A_i | K, \mathcal{Y}_{1:t-1}) = P(A_i | K) \tag{2.64}$$

$$= \begin{cases} \frac{1}{K} P_{\mathrm{D}} P_{\mathrm{G}} \left(P_{\mathrm{D}} P_{\mathrm{G}} + (1 - P_{\mathrm{D}} P_{\mathrm{G}}) \frac{\mu(K))}{\mu(K-1)} \right)^{-1} & i = 1, ..., K \\ (1 - P_{\mathrm{D}} P_{\mathrm{G}}) \frac{\mu(K))}{\mu(K-1)} \left(P_{\mathrm{D}} P_{\mathrm{G}} + (1 - P_{\mathrm{D}} P_{\mathrm{G}}) \frac{\mu(K))}{\mu(K-1)} \right)^{-1} & i = 0 \end{cases} \tag{2.65}$$

where $\mu(\cdot)$ is the probability mass function of the number of false measurements in $\mathcal{V}(t, \gamma)$. Note, $\mu(K)$ means that a particle is not detected and all measurements in $\mathcal{V}(t, \gamma)$ are false measurements, whereas $\mu(K-1)$ implies that a particle is detected and at most one of the measurements has originated from the particle [45]. $\mu(\cdot)$ can either be modeled by a Poisson model or a diffuse prior model. For the Poisson model, assuming N discretized (independent) positions within $\mathcal{V}(t, \gamma)$, the occurrence of a false measurement at a position in $\mathcal{V}(t, \gamma)$ can be obtained by a Bernoulli process with probability $P_{\mathrm{F}} \in [0, 1]$ and thus, the number of false measurements m_{F} follows

a binomial distribution [128]

$$P(K = m_{\mathrm{F}}) \;=\; \binom{N}{m_{\mathrm{F}}} P_{\mathrm{F}}^{m_t} (1 - P_{\mathrm{F}})^{N - m_{\mathrm{F}}} \tag{2.66}$$

In the continuous case ($N \to \infty$), the binomial distribution becomes a Poisson distribution

$$\mu(K = m_{\mathrm{F}}) \;=\; \exp^{-\lambda V} \frac{(\lambda V)^{m_{\mathrm{F}}}}{m_{\mathrm{F}}!} \tag{2.67}$$

with spatial density λ and leads to the parametric PDA filter. The diffuse prior model yields the non-parametric PDA with $\lambda = \frac{K}{V_t}$ [45].

The particle tracking approach in [48] extends the PDA filter by using probabilistic data association with elliptical sampling (PDAE). The PDAE exploits multiple measurements as in a particle filter generated by elliptical sampling around the positions of particle detection and prediction, and integrates them via combined innovations. Thus, the PDAE performs probabilistic tracking based on a combination of Kalman filter and particle filter.

Joint Probabilistic Data Association

The *joint probabilistic data association* (JPDA) filter [45] is an extension of the PDA filter, which determines the association probabilities $\beta_{i,j,t}$ jointly across the particles detected at time point t [127]. The joint association event is defined by $A = \bigcap_{i=1}^{K} A_{i,j}$ [129, 45], where $A_{i,j}$ is the association event that measurement $\mathbf{y}_t^i$ originated from a particle is associated with particle j with $j = 1, \ldots, N_{\mathrm{P}}$ and N_{P} is the (known) number of particles at time point t. The association probabilities $\beta_{i,j,t}$ are jointly computed over A by

$$\beta_{i,j,t} \;=\; P(A|\mathcal{Y}_{1:t}) = \frac{1}{c} P(A|K)\, p(\mathcal{Y}_t^k|A, K, \mathcal{Y}_{1:t-1}) \tag{2.68}$$

where the likelihood of the measurements is

$$p(\mathcal{Y}_t^k|A, K, \mathcal{Y}_{1:t-1}) \;=\; \prod_{i=1}^{K} p(\mathbf{y}_i^k|A_{i,j}, \mathcal{Y}_{1:t-1}) \tag{2.69}$$

with K being the number of measurements $\mathbf{y}_t^i$ in the *union of validation regions*. The decomposition of the likelihood function is based on the assumption that the states of the particles conditioned on the past observations are mutually independent. The probability distribution $p(\mathbf{y}_i^k|A_{i,j}, \mathcal{Y}_{1:t-1})$ of a measurement is

$$p(\mathbf{y}_i^k|A_{i,j}, \mathcal{Y}_{1:t-1}) \;=\; \begin{cases} \mathcal{N}(\mathbf{y}_t^i; \hat{\mathbf{y}}_t^j, \mathbf{S}_t^j) & \mathbf{y}_t^i \text{ associated with particle j} \\ V_t^{-1} & \mathbf{y}_t^i \text{ not associated with a particle} \end{cases} \tag{2.70}$$

where $\hat{\mathbf{y}}_t^j$ is the predicted measurement with innovation covariance $\mathbf{S}_t^j$ for particle j. The probability of joint association events is computed by [45]

$$P(A|K) = \frac{m_\mathrm{F}!}{K!}\,\mu(m_\mathrm{F}) \prod_j^{N_\mathrm{P}} (P_\mathrm{D}^j)^{\delta_j}\,(1-P_\mathrm{D}^j)^{1-\delta_j} \tag{2.71}$$

and exploits all combinations of measurement-to-particle assignments (equally likely) such that (i) each measurement is assigned to at most one particle, and (ii) each particle is uniquely associated to a measurmeent [130]. P_D^j is the detection probability for particle j and $\mu(m_\mathrm{F})$ is the probability mass function of the number of false measurements m_F (cf. PDA filter).

Multiple Hypothesis Tracking

While most Bayesian filtering approaches consider only two frames for correspondence finding, *multiple-hypothesis tracking* (MHT) uses multiple frames (e.g., for particle tracking [131, 132, 133] and cell tracking [134, 135]). MHT approaches seek globally optimal solutions, but may not determine locally best associations. MHT exploits multiple association hypothesis in order to solve ambiguities of motion correspondences by using more information from subsequent image frames [136]. The posterior distribution $p(\mathcal{X}_t|\mathcal{Y}_{1:t})$ for multiple objects can be obtained by conditioning over all possible association hypotheses $\mathcal{A}$ [131]

$$p(\mathcal{X}_t|\mathcal{Y}_{1:t}) = \sum_{\mathcal{A}} p(\mathbf{X}_t|\mathcal{Y}_{1:t},\mathcal{A})\,p(\mathcal{A}|\mathcal{Y}_{1:t}) \tag{2.72}$$

where $\mathcal{X}_t = \{\mathbf{x}_t^i\}_{i=1}^K$ is a set of states of K particles at time point t and $\mathcal{Y}_{1:t}$ is a set of measurements until time point t. MHT obtains the (best) particle state estimates $\mathcal{X}_t^*$ based on the maximum a posteriori (MAP) *globally best hypothesis* $\mathcal{A}^*$ [131]

$$\mathcal{A}^* = \arg\max_{\mathcal{A}} p(\mathcal{A}|\mathcal{Y}_{1:t}) \tag{2.73}$$

$$\mathcal{X}_t^* = \arg\max_{\mathcal{X}_t} p(\mathcal{X}_t|\mathcal{Y}_{1:t},\mathcal{A}^*) \tag{2.74}$$

MHT algorithms can be categorized into *hypothesis-oriented* MHT (HOMHT) and *track-oriented* MHT (TOMHT). The HOMHT approach was proposed in [136] and exploits a single tree of association hypotheses in order to compute the globally optimal assignment solution. The TOMHT approach was introduced in [137, 138] and maintains a set of track trees where a node in each tree represents a measurement and branches corresponds to possible tracks. For each time point t, the track trees are expanded by using all feasible measurement-to-particle associations and each track hypothesis is assigned a track score [139]. TOMHT obtains the globally optimal assignments by representing correspondence finding as maximum weighted independent set (MWIS) problem [140] where the highest scoring set of compatible

track hypothesis is determined. A set of track hypotheses is compatible if the track hypotheses do not share any measurement at any time point t. TOMHT was proved to be mathematically equivalent to HOMHT [138].

The particle tracking approach in [131] uses a multi-stage TOMHT approach where the first stage forms tracklets based on a low number of measurements and the second stage performs track fusion with the retained measurements. The approach in [133] solves MHT as integer programming problem and exploits tracking in forward and backward temporal direction to determine particle merging and splitting. However, the correspondence space is explored over multiple time points and therefore the computational complexity is generally high for high object density. The particle tracking approach in [132] and the cell tracking approach in [135] focus on non- or weakly interacting objects and do not address object division events. The cell tracking MHT approach in [134] was used to investigate cell behavior during angiogenesis, however, cell division events are detected by a deterministic (distance-based) approach which does not take into account uncertainties.

2.4.3 Deep Learning Approaches

In recent years, deep learning approaches have been introduced for particle detection in microscopy images [84, 83, 14]. Approaches based on convolutional neural networks (CNNs) have been proposed [84, 83]. These methods require a relatively large number of parameters or are based on a sliding window scheme. [14] proposed an hourglass-shaped Deconvolution Network denoted as DetNet which has a significantly reduced number of parameters and does not require a sliding window scheme.

For particle tracking, deep learning methods were presented (e.g., [141, 142, 143, 144, 145, 146]). These supervised methods can yield good results but have the disadvantage that they require ground truth data for training, and interpretability is lacking. For tracking objects in a sequence of images, RNNs are typically employed to compute assignment scores for correspondence finding (e.g., [147]) by exploiting appearance features (e.g., [148]). However, a main challenge in particle tracking is the lack of prominent shape and appearance characteristics. Therefore, RNNs have been used to learn particle dynamics or compute assignment probabilities based on particle motion prediction. The approaches in [141, 144] exploit handcrafted and learned features. In [143], the motion model within a classical MHT framework is learned by exploiting a denoising autoencoder and score matching. In [142], an LSTM-based RNN is used to compute assignment probabilities jointly across multiple detections by exploiting only past temporal information. This method has been extended in [145] by exploiting a bidirectional RNN (stacked bidirectional LSTMs) which integrates past and future information as well as multiple track hypotheses for correspondence finding. The assignment probabilities are directly determined and computed jointly across multiple detections, and also probabilities of missing detections are computed. Recently, a probabilistic deep learning approach for particle

tracking has been introduced which mimics Bayesian filtering by using Bayesian layers and considering aleatoric and epistemic uncertainty [146]. The aleatoric uncertainty reflects the noise in the data (e.g., particle detection noise, particle motion noise) and the epistemic uncertainty the model uncertainty.

2.5 Colocalization and Motion Analysis in Microscopy Images

In this section we describe methods for colocalization and motion analysis of particles in microscopy image sequences. First, approaches for *colocalization analysis* are reviewed which provide information of spatial relationships of particles in multi-channel fluorescence microscopy images. Second, an approach for *motion analysis* is described which uses the mean squared displacement for computing diffusion parameters to characterize particle kinetics.

Colocalization Analysis

To gain insights into virus-host interactions at high spatial resolution, intensive research is carried out using multi-channel fluorescence microscopy images in combination with automatic object detection and quantification. For the hepatitis C virus, subcellular locations of RNA replication and virus particle components are known, but the exact location of virus assembly and the spatio-temporal coupling of the relevant processes remains unclear[149, 18, 150]. The location of viral proteins in the cell and their spatial relationship with subcellular structures can be accessed by colocalization analysis of fluorescently labeled proteins.

Previous work on colocalization analysis can be subdivided into *intensity-based* (pixel-based), *object-based*, and *track-based* approaches. *Intensity-based* approaches perform correlation analysis of the image intensities [151, 152, 153, 154, 155, 156, 157, 158, 159]. A correlation between the intensities of two-channel images can be calculated by the Pearson correlation coefficient [154] ranging from -1 to 1. However, the negative values are hard to interpret [155]. Therefore, the Manders overlap coefficient was introduced which ranges from 0 to 1 [155]. This coefficient does not consider the average intensity values of the two channels compared to the Pearson correlation coefficient. Another extension of the Pearson correlation coefficient determines a cross-correlation function by shifting one channel relative to the other channel [156]. However, this approach requires cross-correlation analysis for each image dimension. The approach by Costes uses a statistical significance algorithm based on the Pearson correlation coefficient and image randomization to exclude image points with random colocalization [157]. The approach by Li is based on an intensity correlation analysis and assesses the difference of pixel intensities to the mean intensity for each channel to determine segregated or dependent staining in

image channels [158]. To exclude colocalizations by chance, a threshold overlap score was proposed which classifies the intensity distributions between colocalization, anti-colocalization, and non-colocalization [159]. However, the intensity-based approaches do not consider spatial information (e.g., positions) of the colocalization and do not perform colocalization for a particular object of interest. In comparison, *object-based* colocalization approaches localize objects of interest and perform a colocalization on the object level (e.g., nearest-neighbour distance, intensity model). Since these approaches determine colocalization for individual objects (e.g., proteins, cells), they are more appropriate to determine molecular interaction in order to gain insights into virus assembly. Different object-based colocalization approaches have been introduced [160, 161, 162, 163, 164, 165, 166, 167]. Often, colocalization is determined by a nearest-neighbour approach which compares the positions of object centroids between two channels [160, 162, 164]. The approach of Jaskolski [161] combines binary images obtained by object segmentation using a Boolean operation (disjunction) and determines colocalization within region-of-interests by the difference of image intensities to the mean intensity. Statistical colocalization approaches have also been introduced which use distance-based multiple hypothesis tests [163] or exploit statistical tests to compare detected objects with objects within a region-of-interest [166]. For 3D multi-channel microscopy images, the approach by Wörz et al. [165] uses different 3D parametric intensity models and determines colocalization based on the estimated geometry of the subcellular structures. Another colocalization approach is based on two channel 3D image cross-correlation, which is determined at each object position obtained by single-particle tracking [167]. A temporal extension of object-based colocalization are *track-based* approaches which combine spatial and temporal information [168, 169, 170]. Single-particle tracking and colocalization by trajectory correlation was used in [168]. Trajectory correlation within a window and adaptive thresholding based on the Pearson correlation coefficient was also employed [169]. In recent work, a probabilistic tracking approach with integrated colocalization analysis was introduced [170]. This approach is based on a combination of the Kalman filter and particle filter for tracking, and jointly performs tracking and colocalization analysis. However, the track-based colocalization approaches described above determine colocalization only based on two-channel fluorescence microscopy images. To gain further insights into virus assembly, more than two fluorophores could be used for labelling structures of interest and therefore colocalization needs to be determined in multi-channel fluorescence microscopy images with more than two channels.

A broad diversity of publicly available plugins and software suites for colocalization analysis are available to support biologists in image analysis, such as JACoP v2.0 [151], Squassh [171], DiAna [172], EzColocalization [173], and Colocalization Colormap [174]. Further, the ImageJ [175] platform provides plugins such as Coloc 2, Colocalization Finder, and ComDet. Intensity-based colocalization analysis can be performed with JACoP v2.0, Squassh, EzColocalization, Coloc 2, and Colocalization Finder.

Object-based colocalization can be performed by JACop v2.0, Squassh, DiAna, Colocalization Colormap, and ComDet. The ImageJ plugins JACoP v2.0, Squassh, DiAna, Colocalization Colormap, Coloc 2, and Colocalization Finder determine colocalization only for two-channel images. Only EzColocalization and ComDet are able to analyse multi-channel images with more than two channels. However, all publicly available plugins and software suites for colocalization analysis contain a visualization of colocalization statistics (e.g., scatterplot, histogram, tables), but do not provide a quantification and visualization of the color distribution and spatial distribution for each colocalized object, respectively.

Motion Analysis

To understand viral replication and infection in a spatial-temporal context, studying the kinetics of labeled viral compartments in order to extract motion properties of the particles to infer diffusion parameters is important [176, 25, 177, 19, 178]. Diffusion of particles is a stochastic process and can be mathematically modeled by random walk (e.g., Brownian motion) [179]. It has been derived by Einstein in 1905 that the fundamental quantity of random walk is the mean of the quadratic spatial shifts over time [180] which corresponds to the variance of the Gaussian distributed increments of a Wiener process [181]. Hence, a particle can be assumed to perform random walk and the explored space over time can be measured by the mean square displacement (MSD). The MSD can be obtained based on the particle trajectory by computing [182, 183]

$$\mathrm{MSD}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}(t_i + \Delta t) - \mathbf{x}(t_i)||^2 \tag{2.75}$$

where $\mathbf{x} \in \mathbb{R}^n$ are the particle position coordinates measured at N discrete time lags Δt. In order to quantify diffusion parameters for a whole system of $j = \{1, 2, ..., L\}$ moving particles, the MSD can be computed as mean over a set of L trajectories with [179]

$$\mathrm{MSD}(\Delta t) = \frac{1}{L}\sum_{j=1}^{L}\frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}_j(t_i + \Delta t) - \mathbf{x}_j(t_i)||^2 \tag{2.76}$$

The motion type and diffusion coefficient can be characterized and estimated from the measured MSD values by fitting a diffusion model to the MSD values. For Brownian motion, the MSD values grow linearly in time ($\mathrm{MSD} \sim t$) and the diffusion model

$$\mathrm{MSD}(\Delta t) = 2\,n\,D\,t \tag{2.77}$$

is fitted to the MSD values from the particle coordinates $\mathbf{x} \in \mathbb{R}^n$ in order to obtain the diffusion coefficient $D \in \mathbb{R}$. For anomalous diffusion the MSD values show a power scaling (MSD $\sim t^\alpha$) and are fitted to an anomalous diffusion model

$$\mathrm{MSD}(\Delta t) = 2\,n\,\Gamma\,t^\alpha \tag{2.78}$$

with the anomalous diffusion exponent $\alpha \in \mathbb{R}$ and the transport coefficient $\Gamma \in \mathbb{R}$ [182]. Based on the anomalous diffusion exponent α, the motion type can be classified into confined diffusion with $\alpha \leq 0.1$, obstructed diffusion with $0.1 < \alpha < 0.9$, and normal diffusion with $\alpha \geq 0.9$ [184]. For $\alpha = 1$, the anomalous diffusion model is equal to the diffusion model for Brownian motion

$$2\,n\,D\,t = 2\,n\,\Gamma\,t^\alpha \tag{2.79}$$

$$D = \Gamma\,t^{\alpha-1} = \Gamma \tag{2.80}$$

and the diffusion coefficient D is equal to the transport coefficient Γ.

3 Bayesian Smoothing for Particle Tracking

In this chapter, methods for integrating temporal information for particle tracking are proposed. We formulate particle tracking as a Bayesian smoothing problem. First, we introduce a novel two-filter probabilistic data association approach for particle tracking. Second, a novel Bayesian smoothing approach which uses the covariance intersection algorithm is described. The proposed methods have been published in Ritter *et al.* [11, 12].

3.1 Particle Tracking as Bayesian Smoothing Problem

Particle tracking can be formulated as Bayesian sequential estimation problem. In our approach, we assume that a fluorescently labeled particle is represented by a state vector $\mathbf{x}_t = (p_x, \dot{p}_x, p_y, \dot{p}_y, I_{\max}, \sigma_{xy})^T$ at time point t. The state $\mathbf{x}_t$ contains the position $\mathbf{p}_t = (p_x, p_y)^T$, velocity $\mathbf{v}_t = (\dot{p}_x, \dot{p}_y)^T$, and intensity information $\mathbf{I}_t = (I_{\max}, \sigma_{xy})^T$ consisting of the maximum intensity $I_{\max}$ and the width σ_{xy} of a Gaussian appearance model $g_{\text{Gauss}}(x, y; \mathbf{x}_t) = I_{\text{b}} + (I_{\max} - I_{\text{b}}) \exp\left(-\frac{(x-p_x)^2+(y-p_y)^2}{2\sigma_{xy}^2}\right)$ for a spot-like structure (I_{b} denotes the background intensity). The temporal evolution of the state $\mathbf{x}_t$ can be formulated by a dynamic model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. Further, it is assumed that the noisy measurement $\mathbf{y}_t = (\overline{p}_x, \overline{p}_y, \overline{I}_{\max}, \overline{\sigma}_{xy})^T$ reflects the state $\mathbf{x}_t$. The measurements can be incorporated via a measurement model $p(\mathbf{y}_t|\mathbf{x}_t)$. Bayesian estimation determines the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ conditionally on all measurements $\mathbf{y}_{1:t} \in \mathbb{R}^{4\times t}$ up to time point t using Bayes' theorem [65]

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{1}{Z_k}\, p(\mathbf{y}_t|\mathbf{x}_t)\; p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \tag{3.1}$$

where

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\; p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, \mathrm{d}\mathbf{x}_{t-1} \tag{3.2}$$

$$Z_k = \int p(\mathbf{y}_t|\mathbf{x}_t)\; p(\mathbf{x}_t|\mathbf{y}_{1:t-1})\, \mathrm{d}\mathbf{x}_t \tag{3.3}$$

By assuming linear and Gaussian models, the posterior distribution

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) \sim \mathcal{N}\left(\mathbf{x}_t; \mathbf{m}_t, \mathbf{P}_t\right) \tag{3.4}$$

with mean $\mathbf{m}_t \in \mathbb{R}^{6\times1}$ and covariance matrix $\mathbf{P}_t \in \mathbb{R}^{6\times6}$ can be resolved in a closed form via the Kalman filter. For the dynamic model

$$p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) \sim \mathcal{N}\left(\mathbf{x}_t; \mathbf{F}\,\mathbf{m}_{t-1}, \mathbf{Q}\right) \tag{3.5}$$

which includes the transition matrix $\mathbf{F} \in \mathbb{R}^{6\times6}$ we assume a random walk model with $\mathbf{F} = \mathrm{diag}(1,1,1,1,1,1)$. The uncertainty of the dynamic model is reflected by the covariance matrix $\mathbf{Q} = \mathrm{diag}(q_{p_x}, q_{\dot{p}_x}, q_{p_y}, q_{\dot{p}_y}, q_{I_{\max}}, q_{\sigma_{xy}})$. The measurement model

$$p\left(\mathbf{y}_t|\mathbf{x}_t\right) \sim \mathcal{N}\left(\mathbf{y}_t; \mathbf{H}\,\mathbf{m}_t, \mathbf{R}\right) \tag{3.6}$$

with

$$\mathbf{y}_t = \mathbf{H}\,\mathbf{m}_t \tag{3.7}$$

is given by the measurement matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.8}$$

and the covariance matrix $\mathbf{R} = \mathrm{diag}(r_{p_x}, r_{p_y}, r_{I_{\max}}, r_{\sigma_{xy}})$. The Kalman filter consists of two main steps, state prediction and update estimation [65]. The mean $\hat{\mathbf{m}}_t \in \mathbb{R}^{6\times1}$ and covariance $\hat{\mathbf{P}}_t \in \mathbb{R}^{6\times6}$ of the predicted state $\hat{\mathbf{x}}_t \in \mathbb{R}^{6\times1}$ is computed using the state estimate $\mathbf{x}_{t-1}$ at time point $t-1$:

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

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

The update estimate of the mean $\mathbf{m}_t \in \mathbb{R}^{6\times1}$ and covariance $\mathbf{P}_t \in \mathbb{R}^{6\times6}$ of the state $\mathbf{x}_t$ is computed by

$$\mathbf{m}_t = \hat{\mathbf{m}}_t + \mathbf{K}\left(\mathbf{y}_t - \mathbf{H}\,\hat{\mathbf{m}}_t\right) \tag{3.11}$$

$$\mathbf{P}_t = \left(\mathbf{I} - \mathbf{K}\,\mathbf{H}\right)\hat{\mathbf{P}}_t \tag{3.12}$$

with the identity matrix $\mathbf{I} \in \mathbb{R}^{6\times6}$ and the Kalman gain $\mathbf{K} \in \mathbb{R}^{6\times4}$

$$\mathbf{K} = \hat{\mathbf{P}}_t\,\mathbf{H}^T\left(\mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{3.13}$$

However, when calculating the posterior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, only information from past time points (and the current time point) is used. To improve the estimation of $\mathbf{x}_t$, a smoothing approach can be used which also incorporates information from future time points. Thus, both past and future time points are exploited. The smoothing approach determines the posterior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ conditionally on all measurements up to time point T, where $T > t$ is a future time point [65]:

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = p(\mathbf{x}_t|\mathbf{y}_{1:t}) \int \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t)\, p(\mathbf{x}_{t+1}|\mathbf{y}_{1:T})}{p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})} \mathrm{d}\mathbf{x}_{t+1} \tag{3.14}$$

3.2 Two-Filter Probabilistic Data Association for Particle Tracking

Tracking subcellular structures displayed as small spots in fluorescence microscopy images is important to determine quantitative information of biological processes. In previous work, different types of particle tracking approaches have been introduced (e.g., [106, 102, 185, 186]). Probabilistic tracking approaches based on Bayesian filtering have the advantage that spatial and temporal uncertainties are taken into account (e.g., [132, 48, 49]). Godinez et al. [48] introduced a probabilistic particle tracking approach based on probabilistic data association. There, information from past time points is used to find correspondences between detected particles in consecutive image frames. However, incorporation of information from future time points can improve the tracking results. Roudot et al. [49] described an approach based on piecewise-stationary motion smoothing to recover heterogeneous motion of particles. Past and future information are exploited to select the motion model, which is represented by an additional random variable.

We have developed a new approach for tracking multiple particles in time-lapse microscopy images which is based on smoothing and probabilistic data association. We propose a two-filter smoothing method, where two filters run in opposite temporal directions, and which integrates a localization scheme based on probabilistic data association. Compared to [49], our approach incorporates past and future information directly in the state space by exploiting multiple measurements as in a particle filter. The measurements are integrated into a Kalman filter via combined innovation.

In our approach, fluorescently labeled particles are represented by a state vector $\mathbf{x}_t$ which is reflected by the noisy measurement $\mathbf{y}_t$. For time point t, the aim is to estimate the true state $\mathbf{x}_t$ given a series of measurements $\mathbf{y}_{1:t}$. However, when calculating the posterior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, only information from past time points (and the current time point) is used. To improve the estimation of $\mathbf{x}_t$, a smoothing approach can be used which also incorporates information from future time points. Thus, both past and future time points are exploited. The smoothing approach determines the posterior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ conditionally

on all measurements up to time point T, where $T > t$ is a future time point [65]. Fraser and Potter [71] introduced a two-filter smoothing approach by fusing two independent Kalman filters running forward and backward in time. In our new approach, we suggest combining this idea with the PDAE filter [48]. The PDAE filter is based on probabilistic data association [187] with elliptical sampling and exploits multiple measurements similar to a particle filter. However, only information from past time points is exploited. In our approach, we extend the PDAE filter by two-filter smoothing and taking into account information from past and future time points. We denote our approach by S-PDAE. Assuming linear and Gaussian models, the state space equations for the forward and backward PDAE filters can be defined as:

$$\mathbf{x}_{t+1} = \mathbf{F}\,\mathbf{x}_t \tag{3.15}$$

$$\mathbf{y}_t = \mathbf{H}\,\mathbf{x}_t \tag{3.16}$$

For the motion model represented by $\mathbf{F} \in \mathrm{R}^{6\times 6}$ we assume random walk and for the measurement model we use the identity matrix $\mathbf{H} \in \mathrm{R}^{4\times 6}$.

At time point t, the forward PDAE filter provides a measurement $\mathbf{y}_t$ based on the spot-enhancing filter (SEF) [106] and an uncertainty represented by the covariance matrix $\mathbf{R}$. These measurements are referred to as bottom-up measurements. In addition, using the predicted state $\hat{\mathbf{x}}_t$ and the measurement model in (3.16), the measurement $\hat{\mathbf{y}}_t$ with the covariance matrix $\hat{\mathbf{S}}$ is determined. The backward PDAE filter starts from a future time point T and yields the measurement $\tilde{\mathbf{y}}_t$ using the predicted state $\tilde{\mathbf{x}}_t$ with covariance matrix $\tilde{\mathbf{S}}$. The measurements $\hat{\mathbf{y}}_t$ and $\tilde{\mathbf{y}}_t$ are denoted as top-down measurements.

For the forward and backward predicted states $\hat{\mathbf{x}}_t$ and $\tilde{\mathbf{x}}_t$, validation regions are defined based on the covariance matrices $\hat{\mathbf{S}}$ and $\tilde{\mathbf{S}}$. With the submatrices $\hat{\mathbf{S}}_p$ and $\tilde{\mathbf{S}}_p$, including only the position variables, and the predicted positions $\hat{\mathbf{p}}$ and $\tilde{\mathbf{p}}$, elliptical validation regions are determined by

$$V_{\hat{\mathbf{p}},\hat{\mathbf{S}}_p}(\gamma) \equiv \{\mathbf{p} \mid (\mathbf{p}-\hat{\mathbf{p}})^T\,\hat{\mathbf{S}}_p^{-1}\,(\mathbf{p}-\hat{\mathbf{p}}) \le \gamma^2\} \tag{3.17}$$

$$V_{\tilde{\mathbf{p}},\tilde{\mathbf{S}}_p}(\gamma) \equiv \{\mathbf{p} \mid (\mathbf{p}-\tilde{\mathbf{p}})^T\,\tilde{\mathbf{S}}_p^{-1}\,(\mathbf{p}-\tilde{\mathbf{p}}) \le \gamma^2\}\,, \tag{3.18}$$

where the sampling positions $\mathbf{p}$ have a Mahalanobis distance to $\hat{\mathbf{p}}$ and $\tilde{\mathbf{p}}$ which is less than or equal to γ^2. Through diagonalizing $\hat{\mathbf{S}}_p$ and $\tilde{\mathbf{S}}_p$, the semi-axes of the validation regions $V_{\hat{\mathbf{p}},\hat{\mathbf{S}}_p}$ and $V_{\tilde{\mathbf{p}},\tilde{\mathbf{S}}_p}$ are obtained by

$$\hat{\mathbf{r}}_i = \gamma\sqrt{\hat{\lambda}_i}\hat{\mathbf{e}}_i \tag{3.19}$$

$$\tilde{\mathbf{r}}_i = \gamma\sqrt{\tilde{\lambda}_i}\tilde{\mathbf{e}}_i\,, \tag{3.20}$$

where $\hat{\lambda}_i$ and $\hat{\mathbf{e}}_i$ are the eigenvalues and eigenvectors of $\hat{\mathbf{S}}_p$, and $\tilde{\lambda}_i$ and $\tilde{\mathbf{e}}_i$ are defined

analogously. N_j measurements based on $\hat{\mathbf{p}}$ and N_l measurements based on $\tilde{\mathbf{p}}$ are generated in the validation regions respectively by sampling positions $\mathbf{p}$ along N_c concentric ellipsoidal contours with:

$$\mathbf{p}_{j,c} = \hat{\mathbf{p}} + \frac{c}{N_c}\hat{\mathbf{A}}\mathbf{u}_j \tag{3.21}$$

$$\mathbf{p}_{l,c} = \tilde{\mathbf{p}} + \frac{c}{N_c}\tilde{\mathbf{A}}\mathbf{u}_l \tag{3.22}$$

The elliptical contours are centered around the positions $\hat{\mathbf{p}}$ and $\tilde{\mathbf{p}}$ of the forward and backward PDAE filter, respectively. $\hat{\mathbf{A}}$ and $\tilde{\mathbf{A}}$ are rotation matrices containing the eigenvectors $\hat{\mathbf{e}}_i$ or $\tilde{\mathbf{e}}_i$ as column vectors. The factor $c = 1, 2, \ldots, N_c$ is the concentric index and regulates the density of measurements within both validation regions. The assignment of the validation regions in (3.17) and (3.18) is performed by a global nearest neighbor method based on an graph-theoretical approach for the transportation problem [102]. At the current time point t, N_k bottom-up measurements are incorporated analogously. In total, our approach incorporates $N_m = N_c\,(N_j + N_k + N_l) + 3$ measurements. All measurements are taken into account via combined innovation $\nu = \sum_{k=1}^{N_m} \beta_k\,\nu_k$ with $\nu_k = \mathbf{y}_k - \hat{\mathbf{y}}$ and $\sum_{k=1}^{N_m} \beta_k = 1$ as in [48]. The measurements $\mathbf{y}_k$ are determined at the elliptical samples. The association probabilities β_k are computed based on the image likelihood $p\,(\mathbf{z}|\mathbf{x})$ using the Euclidean distance between the image intensities $\mathbf{z}$ within a region-of-interest around the position of the state $\mathbf{x}$ and a Gaussian appearance model for particles.

3.3 Bayesian Smoothing and Multi-Sensor Data Fusion for Particle Tracking

In previous work, different methods for particle tracking have been proposed, which can be subdivided into deterministic, probabilistic, and supervised deep learning approaches. *Deterministic* approaches follow a two step-paradigm consisting of particle detection and correspondence finding [101, 102, 103, 47]. Some approaches determine correspondences by the minimal cost paths in a spatial-temporal volume [105, 107]. However, being computationally efficient, deterministic approaches do not take into account uncertainties. Often, they suffer from low signal-to-noise ratio (SNR) for detection or challenging object constellations for correspondence finding (e.g., high object density, spurious objects). In comparison, *probabilistic* tracking approaches consider spatial-temporal uncertainties and are formulated within a Bayesian framework [125, 46, 48, 120, 49, 11, 126]. Based on Bayesian sequential estimation, the approaches robustly determine the position of a particle from the posterior distribution using noisy particle detections. While most Bayesian filtering approaches consider only two frames for correspondence finding, multiple-hypothesis tracking uses multiple frames [131, 132, 133]. These approaches seek globally optimal solutions, but may not determine locally best associations. Further, the correspondence space

is explored over multiple time points and therefore the computational complexity is generally high for high object density.

Bayesian smoothing approaches for particle tracking that take into account past and future information have recently been introduced in [49, 11]. In [49] an iterative approach was presented which is based on piecewise-stationary motion smoothing to cope with heterogeneous movements. The approach uses a Kalman filter with single measurements. In our previous work [11] described in Section 3.2, we introduced a two-filter smoothing approach that exploits multiple measurements using probabilistic data association with elliptical sampling (PDAE) [48]. However, detection-based and prediction-based measurements are not treated separately, and uncertainty information is not used to combine the measurements. In recent work, *deep learning* methods for particle tracking were presented [141, 142, 143, 145]. However, these supervised methods have the disadvantage that they require ground truth data for training, and interpretability is lacking.

We have developed a novel probabilistic approach for particle tracking in fluorescence microscopy images based on *multi-sensor data fusion*, which integrates *multiple* measurements from separate measurement processes. Our approach integrates detection-based and prediction-based measurements by separate sensor models which allows taking into account different uncertainties to improve update estimation. The approach combines the principle of a particle filter exploiting multiple measurements [188] with Kalman filter predictions [67]. For the measurement process, probabilistic data association with elliptical sampling (PDAE) [48] is used. In addition, predictions from both past and future time points are exploited by a *Bayesian smoothing* method. Predictions are obtained by two filters running in opposite temporal directions and fused in the state space by the *covariance intersection* algorithm. This algorithm is a general multi-sensor data fusion method for unknown cross-covariance to ensure a consistent estimate of the fused prediction. Smoothing and covariance intersection yield improved state prediction and track initialization. Also, our approach exploits motion information based on displacements from past and future time points instead of using positions, and integrates them in the cost function for correspondence finding.

The proposed particle tracking approach is the first that combines multi-sensor data fusion and Bayesian smoothing methods as well as integrates multiple measurements with different uncertainties. Both main steps of Bayesian sequential estimation, namely update estimation and prediction are addressed.

3.3.1 Multi-Sensor Data Fusion with Probabilistic Data Association

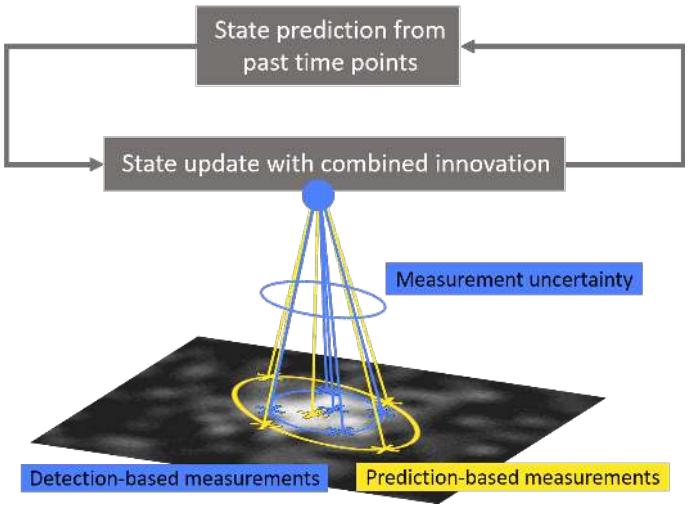
Typical Kalman filter-based approaches for biological particle tracking (e.g., [125, 132, 49]) use *single* measurements $\mathbf{y}_t$ to determine the predictions. Instead, probabilistic data association with elliptical sampling (PDAE) [48] exploits *multiple* measurements

as in a particle filter. With this approach, detection-based and prediction-based measurements are combined using combined innovation. However, detection-based and prediction-based measurements are not treated separately, and uncertainty information is not exploited for combining the measurements. We propose a tracking approach based on a sequential *multi-sensor data fusion* method, which integrates detection-based and prediction-based measurements by exploiting *separate* uncertainties to improve update estimation. Since the measurements are generated using PDAE we denote this approach as Multi-Sensor PDAE (MS-PDAE). For the measurements, detections from a spot detector (e.g., the spot-enhancing filter, SEF, [106]) and Kalman filter predictions are employed. Detection-based measurements and prediction-based measurements of the image intensities are determined within local elliptical regions around the detections and predictions, respectively. We represent the measurement processes by two different sensor models, one for the detection-based measurements and one for the prediction-based measurements, and integrate the measurements by a multi-sensor data fusion method [45]. Such methods integrate measurements from multiple sensors to reduce the overall uncertainty and increase the accuracy [189, 40]. In previous work, multi-sensor data fusion with multiple measurements has been used for radar applications(e.g., [190]), vehicle navigation (e.g., [191, 192]), and remote sensing applications (e.g., [193]), but not yet for object tracking in microscopy images. Note that the term *sensor* is typically used for physical devices, but also refers to mechanisms that derive data from sensory data [40]. We here use the term in the latter more general sense. The sensors are the algorithms that determine the detection-based and prediction-based multiple measurements from the image data.

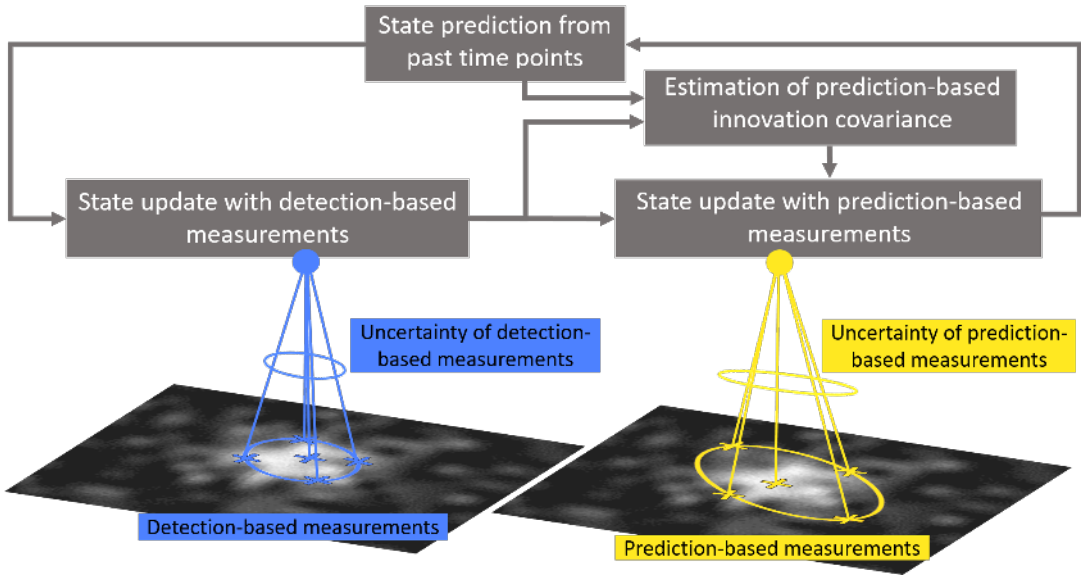
Multi-Sensor Data Fusion

In our application, the detection-based and prediction-based measurements have different uncertainties, and can be assumed to be independent from each other. The detection-based measurements $\mathbf{y}_{i,\mathrm{Det},t}$ at time point t given the current state $\mathbf{x}_t$ are independent of the detection-based measurement history, state history, and state prediction [65]. Thus, $\mathbf{y}_{i,\mathrm{Det},t}$ are independent of the prediction $\hat{\mathbf{x}}_t$ and independent of the prediction-based measurements $\mathbf{y}_{j,\mathrm{Pred},t}$. Therefore, the update estimation of our MS-PDAE can be computed sequentially [45, 189]. The MS-PDAE first determines the update estimate with the measurements $\mathbf{y}_{i,\mathrm{Det},t}$ and second with $\mathbf{y}_{j,\mathrm{Pred},t}$. This is a main difference to the previous PDAE approach [48, 11], where multiple measurements are combined using combined innovation. There, detection-based and prediction-based measurements are not treated separately, and uncertainty information is not exploited for combining the measurements. The difference between PDAE and MS-PDAE is illustrated in Fig. 3.1.

In the first step, the MS-PDAE updates the state $\mathbf{x}_{\mathrm{Det},t} \in \mathbb{R}^{6\times 1}$ (with mean $\mathbf{m}_{\mathrm{Det},t} \in \mathbb{R}^{6\times 1}$ and covariance $\mathbf{P}_{\mathrm{Det,t}} \in \mathbb{R}^{6\times 6}$) based on the predicted state $\hat{\mathbf{x}}_t$ (with



(a) PDAE



(b) MS-PDAE

Figure 3.1: Flowchart illustrating the difference between PDAE and MS-PDAE.

mean $\hat{\mathbf{m}}_t$ and covariance $\hat{\mathbf{P}}_t$) and the detection-based measurements $\mathbf{y}_{i,\mathrm{Det},t}$ using

$$\mathbf{m}_{\mathrm{Det},t} = \hat{\mathbf{m}}_t + \mathbf{K}_{\mathrm{Det},t}\,\mathbf{v}_{\mathrm{Det},t} \tag{3.23}$$

$$\mathbf{P}_{\mathrm{Det,t}} = (\mathbf{I} - \mathbf{K}_{\mathrm{Det},t}\,\mathbf{H})\,\hat{\mathbf{P}}_t \tag{3.24}$$

where $\mathbf{v}_{\mathrm{Det},t} \in \mathbb{R}^{4\times 1}$ is the combined innovation

$$\mathbf{v}_{\mathrm{Det},t} = \sum_{i=1}^{N_{\mathrm{Det}}} \beta_{i,\mathrm{Det},t}\,\mathbf{v}_{i,\mathrm{Det},t} \tag{3.25}$$

consisting of the association probabilities $\beta_{i,\mathrm{Det},t} \in [0,1]$ (based on a Gaussian

appearance model for spot-like particles, see Section 3.1) and the innovations $\mathbf{v}_{i,\mathrm{Det},t} \in \mathbb{R}^{4\times 1}$

$$\mathbf{v}_{i,\mathrm{Det},t} = \mathbf{y}_{i,\mathrm{Det},t} - \mathbf{H}\,\hat{\mathbf{m}}_t \tag{3.26}$$

The matrix $\mathbf{K}_{\mathrm{Det},t} \in \mathbb{R}^{6\times 4}$ is the Kalman gain for $\mathbf{y}_{i,\mathrm{Det},t}$:

$$\mathbf{K}_{\mathrm{Det},t} = \hat{\mathbf{P}}_t\,\mathbf{H}^T\,(\mathbf{S}_{\mathrm{Det},t})^{-1} \tag{3.27}$$
$$\mathbf{S}_{\mathrm{Det},t} = \mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H}^T + \mathbf{R}_{\mathrm{Det}} \tag{3.28}$$

and consists of the innovation covariance $\mathbf{S}_{\mathrm{Det},t} \in \mathbb{R}^{4\times 4}$, which represents the uncertainty of the innovation $\mathbf{v}_{\mathrm{Det},t}$. The uncertainty of $\mathbf{y}_{i,\mathrm{Det},t}$ is reflected by the covariance matrix $\mathbf{R}_{\mathrm{Det}} \in \mathbb{R}^{4\times 4}$.

In the second step, the MS-PDAE updates the state $\mathbf{x}_t$ (with mean $\mathbf{m}_t$ and covariance $\mathbf{P}_t$) based on the estimate $\mathbf{x}_{\mathrm{Det},t}$ (with mean $\mathbf{m}_{\mathrm{Det},t}$ and covariance $\mathbf{P}_{\mathrm{Det,t}}$) and the prediction-based measurements $\mathbf{y}_{j,\mathrm{Pred},t}$ by

$$\mathbf{m}_t = \mathbf{m}_{\mathrm{Det},t} + \mathbf{K}_{\mathrm{Pred},t}\,\mathbf{v}_{\mathrm{Pred},t} \tag{3.29}$$
$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_{\mathrm{Pred},t}\,\mathbf{H})\,\mathbf{P}_{\mathrm{Det},t} \tag{3.30}$$

where $\mathbf{v}_{\mathrm{Pred},t} \in \mathbb{R}^{4\times 1}$ is the combined innovation

$$\mathbf{v}_{\mathrm{Pred},t} = \sum_{j=1}^{N_{\mathrm{Pred}}} \beta_{j,\mathrm{Pred},t}\,\mathbf{v}_{j,\mathrm{Pred},t} \tag{3.31}$$

with the association probabilities $\beta_{j,\mathrm{Pred},t} \in [0,1]$ (based on a Gaussian appearance model) and the innovations $\mathbf{v}_{j,\mathrm{Pred},t} \in \mathbb{R}^{4\times 1}$

$$\mathbf{v}_{j,\mathrm{Pred},t} = \mathbf{y}_{j,\mathrm{Pred},t} - \mathbf{H}\,\mathbf{m}_{\mathrm{Det},t} \tag{3.32}$$

$\mathbf{K}_{\mathrm{Pred},t} \in \mathbb{R}^{6\times 4}$ is the Kalman gain for $\mathbf{y}_{j,\mathrm{Pred},t}$:

$$\mathbf{K}_{\mathrm{Pred},t} = \mathbf{P}_{\mathrm{Det},t}\,\mathbf{H}^T\,(\mathbf{S}_{\mathrm{Pred},t})^{-1} \tag{3.33}$$
$$\mathbf{S}_{\mathrm{Pred},t} = \mathbf{H}\,\mathbf{P}_{\mathrm{Det},t}\,\mathbf{H}^T + \mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H}^T \tag{3.34}$$

The prediction-based innovation covariance $\mathbf{S}_{\mathrm{Pred},t}$ is determined using the covariance $\mathbf{R}_{\mathrm{Pred},t} = \mathbf{H}\,\hat{\mathbf{P}}_t\,\mathbf{H}^T$ of $\mathbf{y}_{j,\mathrm{Pred},t}$, which is changing over time.

The detection-based measurements $\mathbf{y}_{i,\mathrm{Det},t}$ are determined within an elliptical sampling region $V_{\mathrm{Det},t}(\gamma)$ at the position $\overline{\mathbf{p}}_{\mathrm{Det},t} = \left(\overline{p}_{\mathrm{Det,x}}, \overline{p}_{\mathrm{Det,y}}\right)^T$ defined by the covariance matrix $\mathbf{R}_{\overline{\mathbf{p}}}$ in the position space $\mathcal{P} \in \mathbb{R}^{2\times 2}$ by taking only the position information of $\mathbf{R}_{\mathrm{Det}}$. Therefore, multiple measurement at positions $\overline{\mathbf{p}}_{l,c,t}$ are obtained

within the validation region [48]

$$V_{\mathrm{Det},t}(\gamma) \equiv \left\{ \overline{\mathbf{p}}_{l,c,t} \,\middle|\, \left(\overline{\mathbf{p}}_{l,c,t} - \overline{\mathbf{p}}_{\mathrm{Det},t}\right)^T \left(\mathbf{R}_{\overline{p}}\right)^{-1} \left(\overline{\mathbf{p}}_{l,c,t} - \overline{\mathbf{p}}_{\mathrm{Det},t}\right) \leq \gamma^2 \right\} \tag{3.35}$$

where the sampling positions $\overline{\mathbf{p}}_{l,c,t}$ have a Mahalanobis distance to $\overline{\mathbf{p}}_{\mathrm{Det},t}$ less or equal to a constant γ^2. The $\overline{\mathbf{p}}_{l,c,t}$ measurements are determined by sampling N_l positions along N_c concentric elliptical contours centered around $\overline{\mathbf{p}}_{\mathrm{Det},t}$. In total, we use $N_{\mathrm{Det}} = N_l\, N_c + 1$ detection-based measurements and incorporate them in (3.26). Analogously, the prediction-based measurements $\mathbf{y}_{j,\mathrm{Pred},t}$ at positions $\overline{\mathbf{p}}_{m,c,t}$ are obtained within a validation region $V_{\mathrm{Pred},t}(\gamma)$, and we incorporate $N_{\mathrm{Pred}} = N_m\, N_c + 1$ measurements in (3.32). In total, our MS-PDAE uses $N = N_{\mathrm{Det}} + N_{\mathrm{Pred}} = N_c\,(N_l + N_m) + 2$ measurements.

For the combined innovation in (3.25) the association probabilities $\beta_{i,\mathrm{Det},t}$ are interpreted as weights that quantify the probability that the image intensities within a region-of-interest (ROI) around $\overline{\mathbf{p}}_{l,c,t}$ conform to the intensities synthesized with the used Gaussian appearance model (cf. Section 3.1). Therefore, we query the image likelihood defined by the ratio [48]:

$$p\left(I_{\mathrm{ROI}}\left(\overline{\mathbf{p}}_{l,c,t}\right) \middle| I_{g_{\mathrm{Gauss}}}\left(\overline{\mathbf{p}}_{l,c,t}\right)\right) \triangleq \frac{p_o\left(I_{\mathrm{ROI}}\left(\overline{\mathbf{p}}_{l,c,t}\right) \middle| I_{g_{\mathrm{Gauss}}}\left(\overline{\mathbf{p}}_{l,c,t}\right)\right)}{p_b\left(I_{\mathrm{ROI}}\left(\overline{\mathbf{p}}_{l,c,t}\right) \middle| I_{\mathrm{ROI}}\left(\mathbf{p}_{\hat{\mathbf{m}}_t}\right)\right)} \tag{3.36}$$

The image object likelihood $p_o(\cdot)$ is computed using the Euclidean distance between the ROI image intensities around $\overline{\mathbf{p}}_{l,c,t}$ and the intensities synthesized with the appearance model of a particle. The image background likelihood $p_b(\cdot)$ is computed analogously using the ROI background value I_b. The weights $\beta_{i,\mathrm{Det},t}$ represent a probability and need to fulfill the property:

$$\beta_{0,t} + \sum_{i=1}^{N_{\mathrm{Det}}} \beta_{i,\mathrm{Det},t} = 1 \tag{3.37}$$

where $\beta_{0,t} \in [0, 1]$ is the probability that none of the measurements correspond to the tracked particle. The weights for $\mathbf{y}_{j,\mathrm{Pred},t}$ must fulfill the same property.

3.3.2 Bayesian Smoothing Using Covariance Intersection

The MS-PDAE described above uses information from past time points. We extend this approach by introducing a Bayesian smoothing method which exploits information from both past and future time points. We denote this extension as Smoothing MS-PDAE (SMS-PDAE) (see Fig. 3.2 for a flowchart). The SMS-PDAE computes the posterior probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ conditionally on all measurements $\mathbf{y}_{1:T} \in \mathbb{R}^{4\times T}$ up to time point T, where $T > t$ is a future time point [65]. To compute $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, Fraser and Potter [71] introduced a two-filter smoothing approach with two independent Kalman filters running forward and backward in time. For

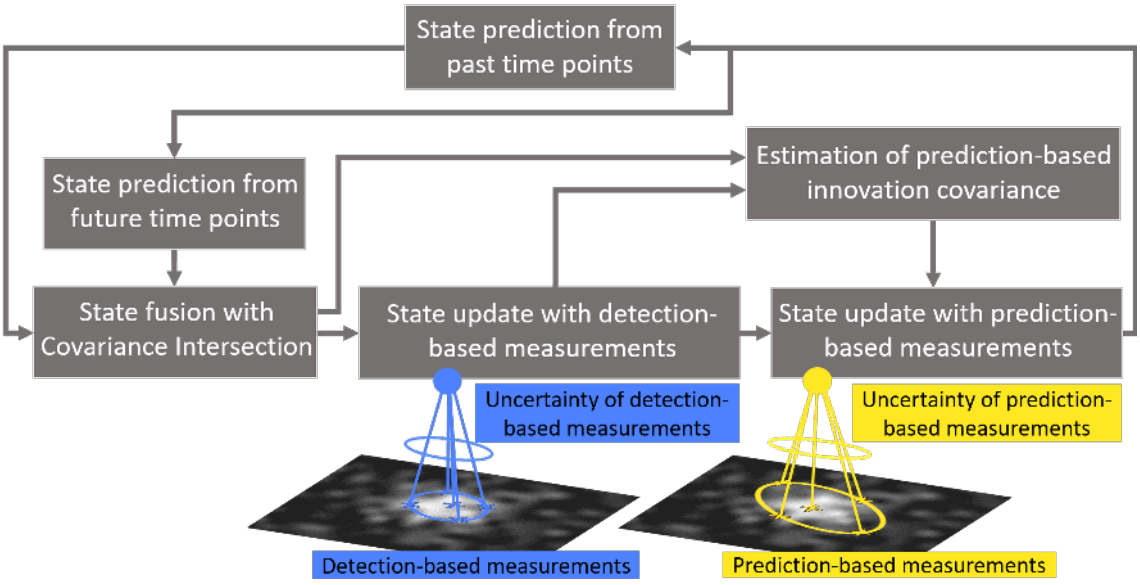


Figure 3.2: Flowchart of SMS-PDAE.

the SMS-PDAE, we exploit this idea, however, we use two MS-PDAE filters (cf. Section 3.3.1) running forward and backward in time. At time point t, the forward MS-PDAE provides a predicted state $\hat{\mathbf{x}}_{t|t-1} \in \mathbb{R}^{6\times 6}$ based on the state at time point $t-1$. The backward MS-PDAE provides a predicted state $\hat{\mathbf{x}}_{t|t+1} \in \mathbb{R}^{6\times 6}$ based on the state at time point $t+1$. Note that the predicted state of the forward MS-PDAE forms a Markov sequence and therefore $\hat{\mathbf{x}}_{t|t-1}$ only depends on $\mathbf{x}_{t-1}$ [65]

$$p\left(\hat{\mathbf{x}}_{t|t-1}\middle|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}\right) = p\left(\hat{\mathbf{x}}_{t|t-1}\middle|\mathbf{x}_{t-1}\right) \tag{3.38}$$

even though measurements $\mathbf{y}_{1:t-1} \in \mathbb{R}^{4\times(t-1)}$ and states $\mathbf{x}_{1:t-1} \in \mathbb{R}^{6\times(t-1)}$ up to time point $t-1$ are used. Analogously, the predicted state $\hat{\mathbf{x}}_{t|t+1}$ of the backward MS-PDAE only depends on $\mathbf{x}_{t+1}$

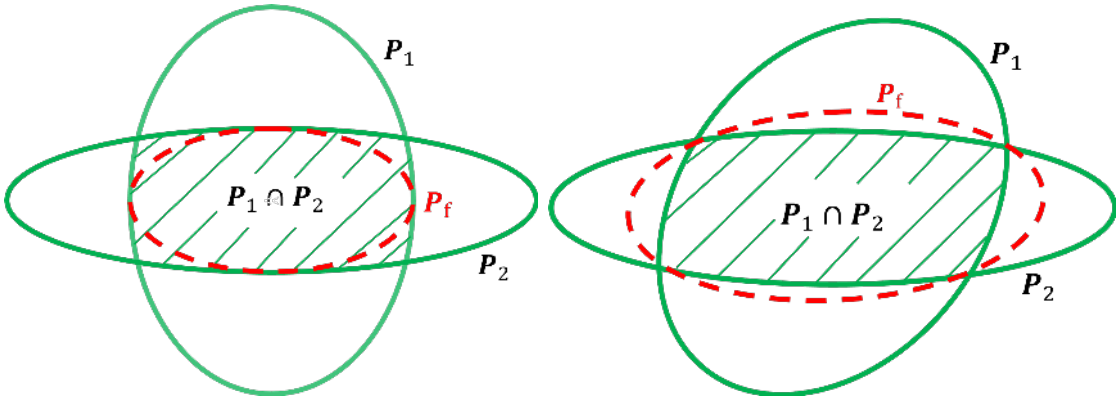
$$p\left(\hat{\mathbf{x}}_{t|t+1}\middle|\mathbf{x}_{t+1:T}, \mathbf{y}_{t+1:T}\right) = p\left(\hat{\mathbf{x}}_{t|t+1}\middle|\mathbf{x}_{t+1}\right) \tag{3.39}$$

even though measurements $\mathbf{y}_{t+1:T} \in \mathbb{R}^{4\times(T-t)}$ and states $\mathbf{x}_{t+1:T} \in \mathbb{R}^{6\times(T-t)}$ from future time points $t+1$ until T with $T > t$ are used. The fused predicted state $\hat{\mathbf{x}}_{\mathrm{f},t} \in \mathbb{R}^{6\times 1}$ at time point t

$$p\left(\hat{\mathbf{x}}_{\mathrm{f},t}\middle|\hat{\mathbf{x}}_{t|t-1}, \hat{\mathbf{x}}_{t|t+1}\right) \sim \mathcal{N}\left(\hat{\mathbf{x}}_{\mathrm{f},t}; \hat{\mathbf{m}}_{\mathrm{f},t}, \hat{\mathbf{P}}_{\mathrm{f},t}\right) \tag{3.40}$$

using the predicted states $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ can be computed based on the unbiased linear combination of the predicted means $\hat{\mathbf{m}}_{t|t-1} \in \mathbb{R}^{6\times 1}$ and $\hat{\mathbf{m}}_{t|t+1} \in \mathbb{R}^{6\times 1}$ [55, 51]

$$\hat{\mathbf{m}}_{\mathrm{f},t} = \mathbf{K}_1\,\hat{\mathbf{m}}_{t|t-1} + \mathbf{K}_2\,\hat{\mathbf{m}}_{t|t+1} \tag{3.41}$$



(a) Known cross-covariances (b) Unknown cross-covariances

Figure 3.3: Illustration of two different optimal fusion approaches for two states. The two green centered ellipses represent the covariances $\mathbf{P}_1$ and $\mathbf{P}_2$ of the two states to be fused. The intersection $\mathbf{P}_1 \cap \mathbf{P}_2$ is marked as green hatched area. a) Known cross-covariances $\mathbf{P}_{12} = \mathbf{P}_{21} = \mathbf{0}$: The optimal fused covariance $\mathbf{P}_\mathrm{f}$ lies within the intersection of $\mathbf{P}_1$ and $\mathbf{P}_2$ (red dashed ellipse). b) Unknown cross-covariances $\mathbf{P}_{12}$ and $\mathbf{P}_{21}$: The optimal fused covariance $\mathbf{P}_\mathrm{f}$ tightly circumscribes the intersection of $\mathbf{P}_1$ and $\mathbf{P}_2$ (red dashed ellipse).

with the gain matrices $\mathbf{K}_i \in \mathbb{R}^{6\times6}$. The gain matrices $\mathbf{K}_i$ are determined by optimizing a cost function $J(\hat{\mathbf{P}}_{\mathrm{f},t})$ of the covariance matrix $\hat{\mathbf{P}}_{\mathrm{f},t}$ of the fused state $\hat{\mathbf{x}}_{\mathrm{f},t}$ given as

$$\hat{\mathbf{P}}_{\mathrm{f},t} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{P}}_{t|t-1} & \hat{\mathbf{P}}_{t|t-1\,;\,t|t+1} \\ \hat{\mathbf{P}}_{t|t+1\,;\,t|t-1} & \hat{\mathbf{P}}_{t|t+1} \end{bmatrix} \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix} \tag{3.42}$$

with $\mathbf{K}_1 + \mathbf{K}_2 = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. $J(\hat{\mathbf{P}}_{\mathrm{f},t})$ is a strictly monotonically increasing function such as the trace or determinant of $\hat{\mathbf{P}}_{\mathrm{f},t}$ [51, 52]. If the cross-covariances $\hat{\mathbf{P}}_{t|t-1\,;\,t|t+1} \in \mathbb{R}^{6\times6}$ and $\hat{\mathbf{P}}_{t|t+1\,;\,t|t-1} \in \mathbb{R}^{6\times6}$ of $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ are *known*, the optimal gain matrices $\mathbf{K}_i$ are determined by the Bar-Shalom/Campo formulas [53]. If the cross-covariances are known and there is no cross-correlation, i.e. $\hat{\mathbf{P}}_{t|t-1\,;\,t|t+1} = \hat{\mathbf{P}}_{t|t+1\,;\,t|t-1} = \mathbf{0}$, the optimal ellipsoid corresponding to the fused covariance $\hat{\mathbf{P}}_{\mathrm{f},t}$ lies within the intersection of the ellipsoids corresponding to the covariances $\hat{\mathbf{P}}_{t|t-1} \in \mathbb{R}^{6\times6}$ and $\hat{\mathbf{P}}_{t|t+1} \in \mathbb{R}^{6\times6}$ of $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ (see Fig. 3.3 a).

Covariance Intersection Algorithm for Unknown Cross-Covariances

In our SMS-PDAE approach, the same motion and noise model is used for the forward and backward MS-PDAE. The measurement processes are not independent, i.e. the cross-covariances in (3.42) are not zero ($\hat{\mathbf{P}}_{t|t-1\,;\,t|t+1} = \hat{\mathbf{P}}_{t|t+1\,;\,t|t-1} \neq \mathbf{0}$) and *unknown*. In particular, the estimation errors of the forward predicted state $\hat{\mathbf{x}}_{t|t-1}$ and the backward predicted state $\hat{\mathbf{x}}_{t|t+1}$ are correlated due to the common motion model noise [53, 194]. When fusing the two predicted states $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$, a *consistent*

estimate (the probability of the estimate being arbitrarily close to the true parameters converges to one as the sample number increases) is only feasible if the ellipsoid corresponding to the fused covariance $\hat{\mathbf{P}}_{\mathrm{f},t} \in \mathbb{R}^{6\times 6}$ bounds the intersection of the ellipsoids of $\hat{\mathbf{P}}_{t|t-1}$ and $\hat{\mathbf{P}}_{t|t+1}$ [55, 51]. Using set theory, the intersection for unknown cross-covariances can be characterized by a convex combination of the covariances, and the *covariance intersection algorithm* [55] allows computing an optimal fused mean $\hat{\mathbf{m}}_{\mathrm{f},t} = \mathbf{K}_1\,\hat{\mathbf{m}}_{t|t-1} + \mathbf{K}_2\,\hat{\mathbf{m}}_{t|t+1}$ with

$$\left(\hat{\mathbf{P}}_{\mathrm{f},t}\right)^{-1} = \omega\,\left(\hat{\mathbf{P}}_{t|t-1}\right)^{-1} + (1-\omega)\,\left(\hat{\mathbf{P}}_{t|t+1}\right)^{-1} \tag{3.43}$$

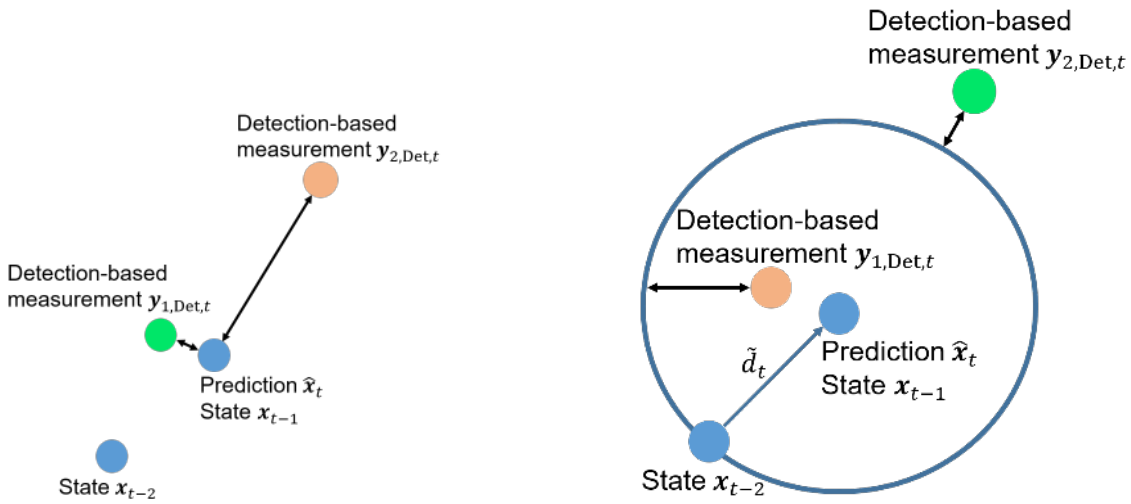
$$\mathbf{K}_1 = \omega\,\hat{\mathbf{P}}_{\mathrm{f},t}\,\left(\hat{\mathbf{P}}_{t|t-1}\right)^{-1} \tag{3.44}$$

$$\mathbf{K}_2 = (1-\omega)\,\hat{\mathbf{P}}_{\mathrm{f},t}\,\left(\hat{\mathbf{P}}_{t|t+1}\right)^{-1} \tag{3.45}$$

where $\omega \in [0,1]$ and $\mathbf{K}_i$ are the optimal gain matrices. The ellipsoid of the optimal fused covariance $\hat{\mathbf{P}}_{\mathrm{f},t}$ from the possible set of covariances tightly circumscribe the intersection of the two centered ellipsoids corresponding to $\hat{\mathbf{P}}_{t|t-1}$ and $\hat{\mathbf{P}}_{t|t+1}$ (see Fig. 3.3 b). The covariance intersection algorithm for fusing the two states $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ yields a consistent estimate $\hat{\mathbf{x}}_{\mathrm{f},t}$ (with mean $\hat{\mathbf{m}}_{\mathrm{f},t}$ and covariance $\hat{\mathbf{P}}_{\mathrm{f},t}$) which has been proofed in [55] as well as in [52, 51] for more general situations. Covariance intersection is a completely general data fusion method (for any cross-correlation), which is optimal when the cross-covariances are not exactly known. Using in this case instead the data fusion approach with known cross-covariances by the Bar-Shalom/Campo formulas [53] (cf. Fig. 3.3 a), then the uncertainties are underestimated, and an inconsistent and suboptimal estimate is generally obtained (e.g., [55, 39]). In our initial experiments, we had used the latter approach and empirically found that the uncertainties for the fused state were too low which deteriorated the tracking results.

Assignment Between Forward and Backward Predictions

The forward and backward MS-PDAE predicted states $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ in the SMS-PDAE are assigned by a global nearest neighbor method based on a graph-theoretical approach for the transportation problem [102]. If a forward predicted state $\hat{\mathbf{x}}_{t|t-1}$ is not assigned to a backward MS-PDAE predicted state $\hat{\mathbf{x}}_{t|t+1}$ ($\mathbf{K}_1 = \mathbf{I}$ and $\mathbf{K}_2 = \mathbf{0}$), the fused mean is $\hat{\mathbf{m}}_{\mathrm{f},t} = \hat{\mathbf{m}}_{t|t-1}$ with covariance $\hat{\mathbf{P}}_{\mathrm{f},t} = \hat{\mathbf{P}}_{t|t-1}$. If a backward predicted state $\hat{\mathbf{x}}_{t|t+1}$ is not assigned to a forward predicted state $\hat{\mathbf{x}}_{t|t-1}$ ($\mathbf{K}_1 = \mathbf{0}$ and $\mathbf{K}_2 = \mathbf{I}$), the fused mean is $\hat{\mathbf{m}}_{\mathrm{f},t} = \hat{\mathbf{m}}_{t|t+1}$ with covariance $\hat{\mathbf{P}}_{\mathrm{f},t} = \hat{\mathbf{P}}_{t|t+1}$, where $\mathbf{I} \in \mathbb{R}^{6\times 6}$ is the identity matrix.



(a) Position-based approach (nearest neighbor) (b) Displacement-based approach

Figure 3.4: Illustration of the two different correspondence finding approaches for a random walk motion model. The Euclidean distances are indicated by black arrows and represent the costs for correspondence finding. a) The position-based approach determines the closest point (green) which is an incorrect correspondence. b) The displacement-based approach uses the displacement $\tilde{d}_t$ between the previous and current position (blue arrow) to define a circle with the expected displacement (blue circle) and determines the closest point (green) to the circle which is the correct correspondence.

Displacement-based Correspondence Finding Between Predictions and Measurements

To determine correspondences between predictions and measurements in the MS-PDAE we use a displacement-based approach. In previous work, often a nearest-neighbor approach with a cost function based on the Euclidean distance between the predicted position and the measurement position is used (e.g., [106, 132, 48, 120, 121, 49]). In contrast to this position-based approach, in the MS-PDAE we exploit motion information in form of displacements in the cost function for correspondence finding. We use the expected displacement at a certain time point, which is the mean of all displacements obtained in the past. The difference between position-based (nearest neighbor) and displacement-based correspondence finding is illustrated in Fig. 3.4. We consider a random walk model, for which the predicted position is equal to the current position (e.g., [195]). The position-based approach determines the closest point (green) and yields an incorrect correspondence. In contrast, the displacement-based approach takes into account the expected displacement (displacement between previous and current position) with a uniformly distributed direction (indicated by the blue circle) and determines the closest point (green) to the circle which is the correct correspondence.

Data Fusion for Displacement-based Correspondence Finding

To determine correspondences between predictions and measurements in the SMS-PDAE we use a displacement-based approach. In the SMS-PDAE we exploit motion information in form of displacements in the cost function for correspondence finding. We use the expected displacement at a certain time point, which is determined based on displacements in both forward and backward direction. We exploit displacements for both the forward and backward MS-PDAE and fuse them for correspondence finding. We define the cost function J based on the displacement $d_t = ||\hat{\mathbf{p}}_{t|t-1} - \overline{\mathbf{p}}_t||_2$ between the predicted position $\hat{\mathbf{p}}_{t|t-1}$ from the forward MS-PDAE and a measurement position $\overline{\mathbf{p}}_t$:

$$J = |\tilde{d}_{\mathrm{f},t} - d_t| \tag{3.46}$$

with the expected fused displacement $\tilde{d}_{\mathrm{f},t}$, which is determined by a linear combination of the expected displacements for the forward and backward MS-PDAE, $\tilde{d}_{\mathrm{fwd},t}$ and $\tilde{d}_{\mathrm{bwd},t}$, respectively:

$$\tilde{d}_{\mathrm{f},t} = \frac{\sigma_{\mathrm{bwd},t}}{\sigma_{\mathrm{fwd},t} + \sigma_{\mathrm{bwd},t}} \tilde{d}_{\mathrm{fwd},t} + \frac{\sigma_{\mathrm{fwd},t}}{\sigma_{\mathrm{fwd},t} + \sigma_{\mathrm{bwd},t}} \tilde{d}_{\mathrm{bwd},t} \tag{3.47}$$

where $\tilde{d}_{\mathrm{fwd},t} \in \mathbb{R}$ denotes the mean displacement over the past displacements at time points $t - (T - t)$ to $t - 1$, $\tilde{d}_{\mathrm{bwd},t} \in \mathbb{R}$ is the mean displacement over the future displacements at time points $t + 1$ to T, and $\sigma_{\mathrm{fwd},t}$ and $\sigma_{\mathrm{bwd},t}$ are the corresponding standard deviations. In (3.47), the weight for the first term is large for small $\sigma_{\mathrm{fwd},t}$ (high certainty) relative to $\sigma_{\mathrm{bwd},t}$, and thus $\tilde{d}_{\mathrm{fwd},t}$ has a large influence. This can be seen by rewriting the weight as $\frac{1}{(\sigma_{\mathrm{fwd},t}/\sigma_{\mathrm{bwd},t})+1}$. The weight for the second term can be interpreted analogously. A graph-theoretical approach [102] is used to optimize the cost function in (3.46).

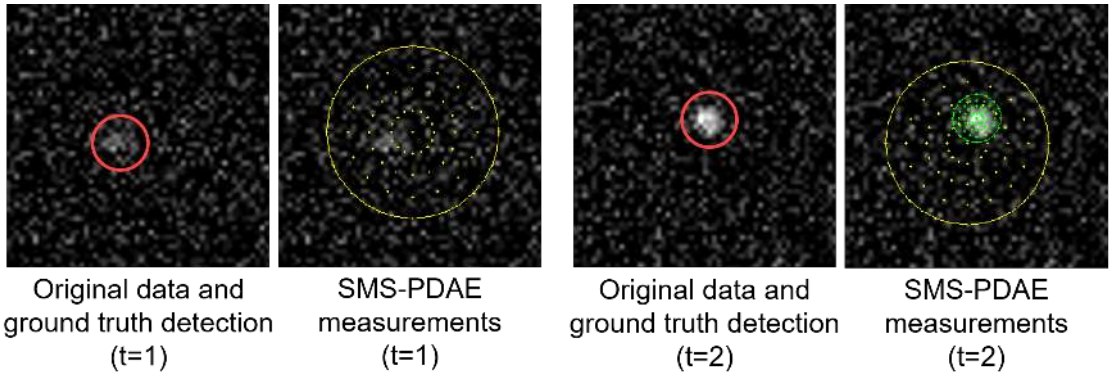


Figure 3.5: SMS-PDAE track initialization for synthetic data. Ground truth detections are indicated by a red circle. For $t = 1$, the particle was not detected and the SMS-PDAE uses only prediction-based measurements (yellow). For $t = 2$, the particle was detected and the SMS-PDAE uses detection-based (green) and prediction-based measurements (yellow).

Track Initialization using Information from Future Time Points

For track initialization, the SMS-PDAE exploits information from future time points by employing unassigned predicted states $\hat{\mathbf{x}}_{t|t+1}$ of the backward MS-PDAE. For an unassigned $\hat{\mathbf{x}}_{t|t+1}$, prediction-based measurements around $\overline{\mathbf{p}}_{\mathrm{Pred},t|t+1}$ are generated and the association probabilities $\beta_{j,\mathrm{Pred},t|t+1} \in [0, 1]$ are determined according to (3.36). If the sum over all $\beta_{j,\mathrm{Pred},t|t+1}$ is above a threshold, a track is initialized with $\mathbf{x}_t = \hat{\mathbf{x}}_{t|t+1}$. Exploiting future information enables starting tracking earlier when particles have not been detected due to low visibility. An example is shown in Fig. 3.5. At time point $t = 1$, the particle was not detected and thus only prediction-based measurements are used, while for $t = 2$, the particle was detected and thus both detection-based (green) and prediction-based (yellow) measurements are exploited.

4 Multi-Detector Fusion and Bayesian Smoothing for Particle Tracking

In this chapter, a novel method for particle tracking based on multi-detector fusion and Bayesian smoothing is described. Multi-detector fusion is achieved by a novel intensity-based covariance intersection algorithm which exploits image intensity information for data fusion. First, an overview of the particle tracking approach is given, followed by introducing the novel multi-detector fusion approach. Afterwards, the Bayesian smoothing approach for tracking is described. The work has been submitted for publication (Ritter *et al.* [13]).

4.1 Overview of the Approach

Due to the limited spatial resolution of optical microscopy, such structures have a spot-like appearance in the image data. The main steps of particle tracking are particle detection and association. Reliable detection and accurate localization of particles are important since errors are propagated to the association step and generally degrade the tracking performance. The main challenges for particle detection and tracking are low signal-to-noise ratio (SNR), small particle size, heterogeneity in particle size, high object density, lack of prominent particle shape, complex motion, and clutter.

We have developed a novel probabilistic approach for particle tracking in fluorescence microscopy images based on *multi-detector multi-scale data fusion* and *Bayesian smoothing*. Our approach integrates multiple measurements (detections) from multiple detectors using multiple image scales by a novel *intensity-based covariance intersection* method. Covariance intersection is a data fusion method for unknown cross-covariances which ensures a consistent estimate. Existing covariance intersection methods are position-based, and were applied to synthetic and remote sensing images [196, 193, 191] and microscopy images [12]. In comparison, the proposed intensity-based covariance intersection method exploits image intensities besides positions and uncertainties, and does not require an optimization step. In addition, we integrate detections from different methods, both classical and deep learning methods as well as exploit detections in multiple image scales. Further, for each particle a time-varying estimate of the measurement noise covariance is computed to improve update estimation. False positive detections are rejected using image likelihoods that represent the image intensities in the local neighborhood of detections. Information from future time points are integrated by Bayesian smooth-

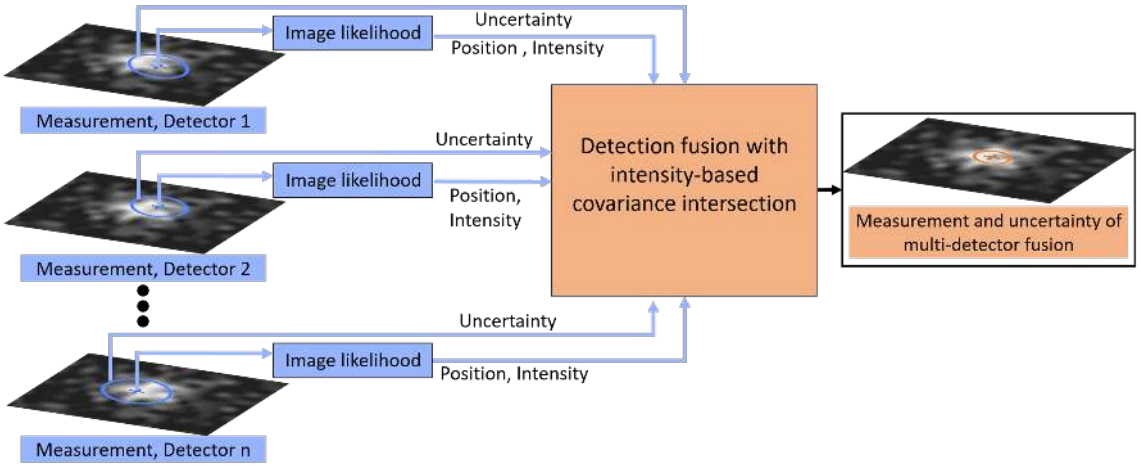


Figure 4.1: Overview of our multi-detector intensity-based covariance intersection method (MD-iCI).

ing. We fuse predictions as well as motion information from past and future time points obtained by two filters running in opposite temporal directions.

4.2 Multi-Detector Fusion for Particle Detection

Probabilistic particle tracking approaches based on Bayesian filtering often use the Kalman filter and exploit *single measurements* for prediction and update estimation (e.g., [125, 132, 49]). *Multiple measurements* (via elliptical sampling around the detection and prediction) were used in the probabilistic data association with elliptical sampling (PDAE) approach [48]. The Smoothing Multi-Sensor PDAE (SMS-PDAE) approach [12] described in Section 3.3 integrates multiple measurements and exploits separate uncertainties to improve update estimation. However, PDAE and SMS-PDAE employ a *single* detection method and use a *single* image scale. In contrast, the proposed tracking approach integrates detections from *multiple* methods (with *separate* uncertainties) and uses *multiple* image scales to improve measurement and update estimation. Particles are detected by different methods and the detections are fused to reduce the overall uncertainty and increase the accuracy. A consistent fused estimate can be obtained by the covariance intersection (CI) algorithm [55], which determines the optimal weighting coefficients of each detection on the fused result. However, an iterative optimization scheme is required and intensity information is not used. We suggest a different approach, where the calculation of weighting coefficients by optimization is replaced by computing image likelihoods directly from the image intensities. The proposed intensity-based covariance intersection (iCI) approach for multi-detector (MD) data fusion is denoted by MD-iCI and takes into account image intensities, positions, and uncertainties from multiple detectors. An overview is given in Fig. 4.1.

For each detector $i \in \{1, 2, ..., n\}$ at time point t, a measurement $\mathbf{y}_{i,t} \in \mathbb{R}^4$ with

uncertainty $\mathbf{R}_i \in \mathbb{R}^{4\times4}$ is obtained. The fused measurement $\mathbf{y}_{\mathrm{f},t} \in \mathbb{R}^{4\times1}$ at time point t (corresponding to one object in an image) can be determined by the unbiased linear combination of the measurements $\mathbf{y}_{i,t}$ [55]

$$\mathbf{y}_{f,t} = \sum_{i=1}^{n} \mathbf{K}_i \, \mathbf{y}_{i,t} \tag{4.1}$$

with the gain matrices $\mathbf{K}_i \in \mathbb{R}^{4\times4}$. The $\mathbf{K}_i$ are determined by optimizing a cost function $J(\mathbf{R}_\mathrm{f})$ of the covariance matrix $\mathbf{R}_\mathrm{f} \in \mathbb{R}^{4\times4}$ of the fused measurement

$$\mathbf{R}_\mathrm{f} \;=\; \begin{bmatrix} \mathbf{K}_1 & \cdots & \mathbf{K}_n \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^1 & \cdots & \mathbf{R}_1^n \\ \vdots & \ddots & \vdots \\ \mathbf{R}_n^1 & \cdots & \mathbf{R}_n^n \end{bmatrix} \begin{bmatrix} \mathbf{K}_1 \\ \vdots \\ \mathbf{K}_n \end{bmatrix} \tag{4.2}$$

with $\sum_{i=1}^{n} \mathbf{K}_i = \mathbf{I}$, and $\mathbf{I}$ is the identity matrix. $J(\mathbf{R}_\mathrm{f})$ is a strictly monotonically increasing function such as the trace or determinant of $\mathbf{R}_\mathrm{f}$ [51, 52, 57]. The matrices $\mathbf{R}_i^i$ in (4.2) represent the covariances of measurement $\mathbf{y}_{i,t}$, and the matrices $\mathbf{R}_i^j$ represent the cross-covariances of measurement $\mathbf{y}_{i,t}$ from detector i based on the measurement $\mathbf{y}_{j,t}$ from detector j.

4.2.1 Multi-Detector Fusion with Intensity-based Covariance Intersection (MD-iCI)

In our MD-iCI, all detectors are applied to the same image. Therefore, the measurement errors can be assumed to be correlated due to common image noise, and the measurement processes are not independent, i.e. the cross-covariances in (4.2) are not zero ($\mathbf{R}_i^j \neq \mathbf{0}$ for $i \neq j$) and *unknown*. Fusing multiple measurements $\mathbf{y}_{i,t}$, a *consistent* estimate of the fused measurement $\mathbf{y}_{\mathrm{f},t}$ is represented by the ellipsoid corresponding to the fused covariance matrix $\mathbf{R}_\mathrm{f}$ which bounds the intersection of all ellipsoids of $\mathbf{R}_i^j$ [55, 51]. An estimate is consistent if it converges to the true parameter and the covariance of the estimation error converges to zero as the sample number increases. Using the CI algorithm [55, 57], the intersection for unknown cross-covariances can be characterized by a convex combination of the covariances and allows computing an optimal fused measurement

$$\mathbf{y}_{\mathrm{f},t} = \mathbf{R}_{\mathrm{f},t} \sum_{i=1}^{n} \omega_i \, (\mathbf{R}_i)^{-1} \, \mathbf{y}_{i,t} \qquad\qquad \mathbf{R}_{\mathrm{f},t} = \left(\sum_{i=1}^{n} \omega_i \, (\mathbf{R}_i)^{-1} \right)^{-1} \tag{4.3}$$

$$0 \leq \omega_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} \omega_i = 1$$

Existing CI algorithms (e.g., [196, 191]) are *position-based* (i.e. they use the positions of detections) and determine the optimal weighting coefficients ω_i by minimizing the

trace (or determinant) of $\mathbf{R}_\mathrm{f}$ with respect to ω_i [51]. However, this is a nonlinear optimization problem with constraints, which requires high computation costs [191]. Also, intensity information is not exploited. In contrast, we suggest a different approach to compute ω_i by exploiting image intensities besides positions and uncertainties. Since the weighting coefficients are constrained to $\omega_i \in [0,1]$ and $\sum_{i=1}^{n} \omega_i = 1$, we interpret them as probabilities. We propose using the probabilities of how well the image intensities within a region-of-interest (ROI) around each detection $\mathbf{y}_{i,t}$ represent particle intensities synthesized with the used Gaussian appearance model (cf. Section 3.1). The probabilities are computed based on the image likelihood which was previously used in [48] for computing association weights to solve correspondence finding. Instead, we here use the probabilities based on the image likelihood to incorporate information of the image intensities around the position $\overline{\mathbf{p}}_{i,t} = \left(\overline{p}_x, \overline{p}_y\right)$ of $\mathbf{y}_{i,t}$ to determine the *intensity-based* weighting coefficients for fusing detections. The image likelihood α_i is defined by

$$\alpha_i = \frac{p_o\left(I_\mathrm{ROI}\left(\overline{\mathbf{p}}_{i,t}\right) \middle| I_{g_\mathrm{Gauss}}\left(\overline{\mathbf{p}}_{i,t}\right)\right)}{p_b\left(I_\mathrm{ROI}\left(\overline{\mathbf{p}}_{i,t}\right) \middle| I_{I_b}\left(\overline{\mathbf{p}}_{i,t}\right)\right)} \tag{4.4}$$

where $p_o(\cdot)$ is the image object likelihood defined as the Euclidean distance between the image intensities within the ROI around $\overline{\mathbf{p}}_{i,t}$ and the image intensities synthesized with the Gaussian appearance model of a particle. $p_b(\cdot)$ is the background image likelihood which is defined analogously using the ROI background value I_b. Since we interpret the weighting coefficients as probabilities, we use the normalization

$$\omega_i = \frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i} \tag{4.5}$$

Thus, since the constraints for ω_i of the existing CI algorithms in (4.3) are fullfilled, our MD-iCI yields a consistent estimate. This was proofed in [55] for arbitrary weighting coefficients. MD-iCI determines the fused measurement $\mathbf{y}_{\mathrm{f},t}$ using the computed intensity-based ω_i and incorporates uncertainties by the covariance matrices $\mathbf{R}_i$ of measurements $\mathbf{y}_{i,t}$ by

$$\mathbf{y}_{\mathrm{f},t} = \mathbf{R}_{\mathrm{f,t}} \sum_{i=1}^{n} \left(\frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i}\right) \left(\mathbf{R}_i\right)^{-1} \mathbf{y}_{i,t} \tag{4.6}$$

The ω_i are computed for each time point t, thus the fused measurement covariance matrix

$$\mathbf{R}_{\mathrm{f,t}} = \left(\sum_{i=1}^{n} \left(\frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i}\right) \left(\mathbf{R}_i\right)^{-1}\right)^{-1} \tag{4.7}$$

is changing over time and is a *time-varying* estimate of the measurement noise covariance. In contrast, previous probabilistic tracking approaches (e.g., [125, 47,

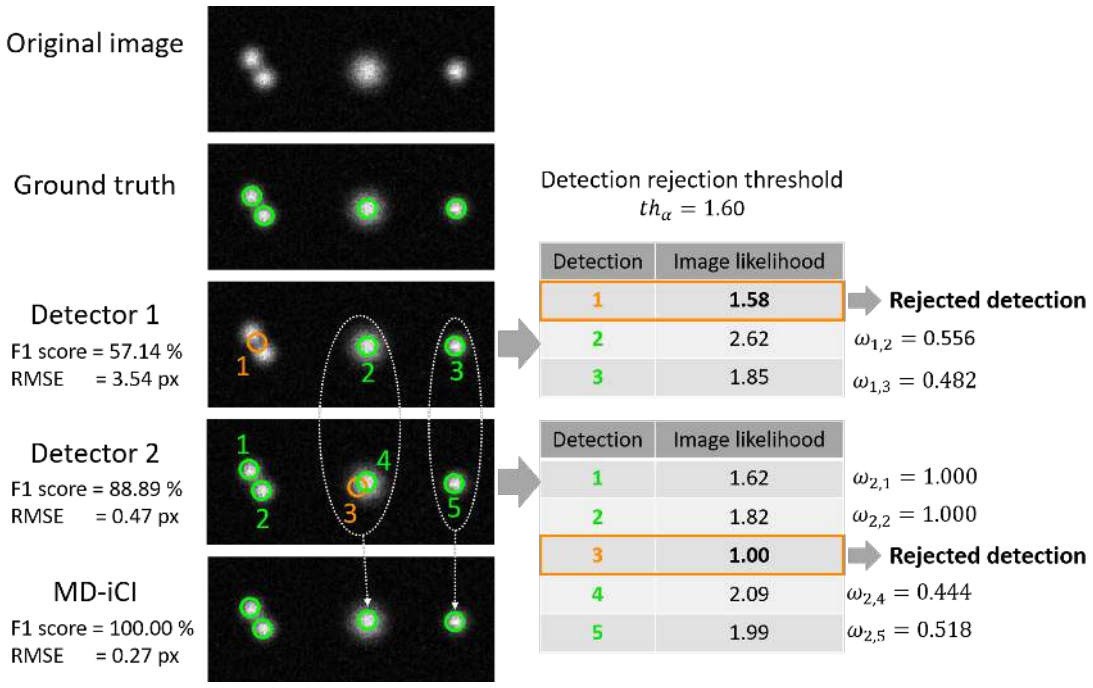


Figure 4.2: MD-iCI and measurement rejection for synthetic data. Ground truth and MD-iCI detections are shown by green circles. For detector 1 and 2, orange circles represent false positive and rejected detections, green circles represent detections used by MD-iCI, and the corresponding table shows the image likelihood and weighting coefficients. The white dotted ellipses depict assigned and fused detections by MD-iCI. Detector 1 is sensitive for large particles leading to a false positive detection for two small particles located close together. Detector 2 is sensitive to small particles and yields a false positive detection for the large particle. MD-iCI fuses the detections of detector 1 and 2 and obtains the best F1 score and $RMSE$ compared to the single detectors.

131, 120, 48, 49, 12]) used a fixed measurement covariance matrix or an uncorrelated matrix (white Gaussian noise), and did not exploit image intensity to determine the measurement noise.

Measurement Rejection and Assignment

In our MD-iCI, we use the image likelihood α_i in (4.4) to identify false positive detections and reject them from fusion. The α_i quantify how well the intensities around measurement $\mathbf{y}_{i,t}$ from detector i agree with the intensities of a particle synthesized with the Gaussian appearance model (cf. Section 3.1). If α_i is below a threshold, $\mathbf{y}_{i,t}$ is classified as false positive detection and rejected for fusion (see Fig. 4.2 for an example). When applying N_d detectors to an image with multiple particles, each detector yields a set of detections. For time point t, the detector with the largest number of detections is identified and a one-to-one correspondence with detections from the other detectors is found by a global nearest neighbor method

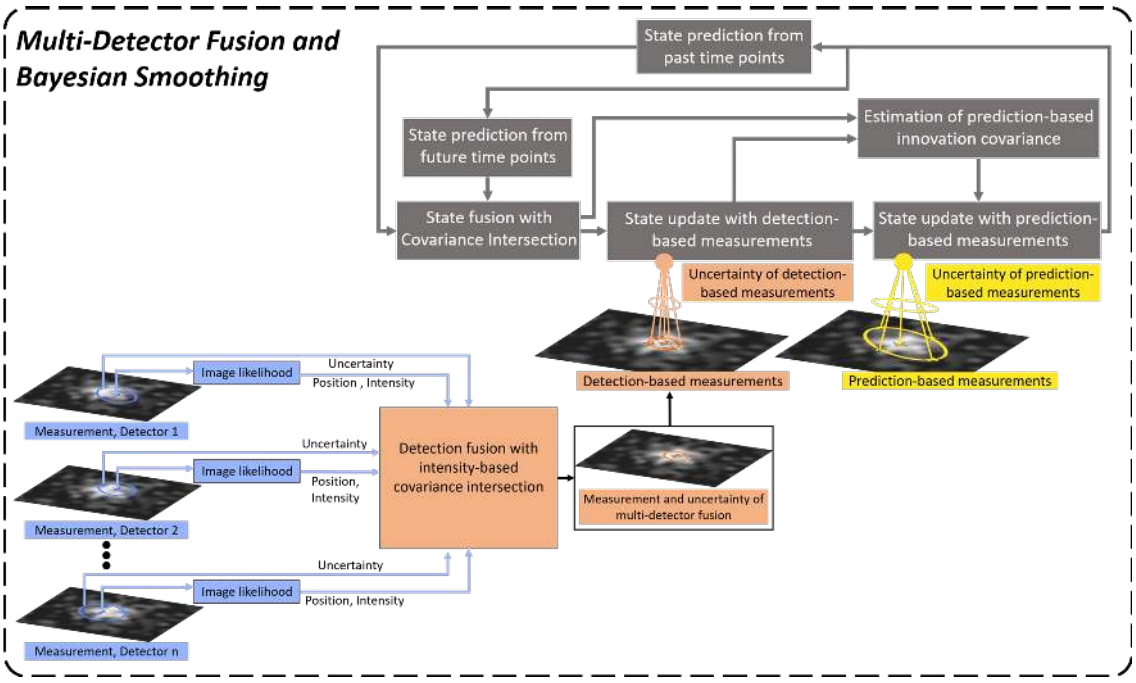


Figure 4.3: Overview of our MD-BS tracking approach. The approach combines our MD-iCI for multi-detector fusion (spatial information) and Bayesian smoothing with covariance intersection for prediction fusion (temporal information).

based on a graph-theoretical approach for the transportation problem [102]. We use the detector with the largest number of detections to obtain the maximum number of one-to-one correspondences. Next, among the $N_d - 1$ detectors with unassigned detections, the detector with the largest number of detections is identified and one-to-one correspondences between detections are determined. This sequential procedure is completed after maximum $N_d - 1$ iterations and the remaining unassigned detections are used for track initialization.

4.3 Bayesian Smoothing with Multi-Detector Fusion

Our MD-iCI approach uses spatial information from multiple detectors. For tracking, we also exploit temporal information using Bayesian smoothing and incorporate past and future time points. We denote our tracking approach as multi-detector Bayesian smoothing (MD-BS). An overview of MD-BS is shown in Fig. 4.3. The posterior probability distribution $p\left(\mathbf{x}_t|\mathbf{y}_{\mathrm{f},1:T}\right)$ is computed conditionally on all fused measurements $\mathbf{y}_{\mathrm{f},1:T} \in \mathbb{R}^{4\times T}$ up to time point T, where $T > t$ is a future time point [65]:

$$p\left(\mathbf{x}_t|\mathbf{y}_{\mathrm{f},1:T}\right) = p\left(\mathbf{x}_t|\mathbf{y}_{\mathrm{f},1:t}\right) \int \frac{p\left(\mathbf{x}_{t+1}|\mathbf{x}_t\right)\, p\left(\mathbf{x}_{t+1}|\mathbf{y}_{\mathrm{f},1:T}\right)}{p\left(\mathbf{x}_{t+1}|\mathbf{y}_{\mathrm{f},1:t}\right)} \mathrm{d}\mathbf{x}_{t+1} \tag{4.8}$$

We solve (4.8) by using two MS-PDAE (Multi-Sensor PDAE, [12]) in conjunction with the proposed MD-iCI running forward and backward in time analogously to the two-filter smoothing approach by Fraser and Potter [71]. At time point t, forward tracking provides a predicted state $\hat{\mathbf{x}}_{t|t-1} \in \mathbb{R}^{6\times 6}$ based on the state at time point $t-1$ due to the Markovian property [65]. Analogously, backward tracking provides a predicted state $\hat{\mathbf{x}}_{t|t+1} \in \mathbb{R}^{6\times 6}$ based on the state at time point $t+1$. The two predicted states are fused as described below.

Prediction Fusion with Covariance Intersection Algorithm

Our MD-BS computes the state $\mathbf{x}_t$ based on the fused predicted state $\hat{\mathbf{x}}_{\mathrm{f},t}$ determined from the forward predicted state $\hat{\mathbf{x}}_{t|t-1}$ and the backward predicted state $\hat{\mathbf{x}}_{t|t+1}$. Two filters are running forward and backward in time and both use the same motion and noise model. Thus, the estimation errors of $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ are correlated due to common noise models [53, 194]. Further, the cross-covariances $\hat{\mathbf{P}}_{t|t-1\,;\,t|t+1}$ and $\hat{\mathbf{P}}_{t|t+1\,;\,t|t-1}$ are not zero and *unknown*. In this case, a *consistent* estimate of $\hat{\mathbf{x}}_{\mathrm{f},t}$ with mean $\hat{\mathbf{m}}_{\mathrm{f},t}$ and covariance $\hat{\mathbf{P}}_{\mathrm{f},t}$ can be obtained by the CI algorithm [55]:

$$\hat{\mathbf{m}}_{\mathrm{f},t} = \hat{\mathbf{P}}_{\mathrm{f},t}\left[\omega\left(\hat{\mathbf{P}}_{t|t-1}\right)^{-1}\hat{\mathbf{m}}_{t|t-1} + (1-\omega)\left(\hat{\mathbf{P}}_{t|t+1}\right)^{-1}\hat{\mathbf{m}}_{t|t+1}\right] \tag{4.9}$$

$$\hat{\mathbf{P}}_{\mathrm{f},t} = \left[\omega\left(\hat{\mathbf{P}}_{t|t-1}\right)^{-1} + (1-\omega)\left(\hat{\mathbf{P}}_{t|t+1}\right)^{-1}\right]^{-1} \tag{4.10}$$

where $\hat{\mathbf{m}}_{t|t-1}$ and $\hat{\mathbf{m}}_{t|t+1}$ are the mean of the predicted states $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$, respectively. In the MD-BS, we equally weight the information from the forward and backward filters using $\omega = 0.5$. Note that the prediction is a coarse estimate of the position of a particle and generally not exactly located at a particle. Thus, exploiting intensity information at the predicted position generally does not improve the result. $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{x}}_{t|t+1}$ are assigned by a global nearest neighbor method [102].

Multi-Sensor PDAE with Multi-Detector Fusion

MD-BS uses forward and backward MS-PDAE filters in conjunction with MD-iCI. Both MS-PDAE filters exploit multiple measurements by generating detection-based and prediction-based measurements within local elliptical regions around the detections and predictions. The two different measurement processes have separate uncertainties and the multiple measurements are integrated by a sequential multi-sensor data fusion approach consisting of two steps to determine the state $\mathbf{x}_t$.

First, the state $\mathbf{x}_{\mathrm{Det},t} \in \mathbb{R}^{6\times 1}$ (with mean $\mathbf{m}_{\mathrm{Det},t} \in \mathbb{R}^{6\times 1}$ and covariance matrix $\mathbf{P}_{\mathrm{Det,t}} \in \mathbb{R}^{6\times 6}$) is determined using the fused predicted state $\hat{\mathbf{x}}_{\mathrm{f},t}$ (with mean $\hat{\mathbf{m}}_{\mathrm{f},t}$ and

covariance matrix $\hat{\mathbf{P}}_{\mathrm{f},t}$) and the *detection-based* measurements $\mathbf{y}_{\mathrm{f},i,\mathrm{Det},t}$ by computing

$$\mathbf{m}_{\mathrm{Det},t} = \hat{\mathbf{m}}_{\mathrm{f,t}} + \mathbf{K}_{\mathrm{Det},t}\,\mathbf{v}_{\mathrm{Det},t} \tag{4.11}$$

$$\mathbf{P}_{\mathrm{Det,t}} = \left(\mathbf{I} - \mathbf{K}_{\mathrm{Det},t}\,\mathbf{H}\right)\hat{\mathbf{P}}_{\mathrm{f},t} \tag{4.12}$$

where $\mathbf{v}_{\mathrm{Det},t} = \sum_{i=1}^{N_{\mathrm{Det}}} \beta_{i,\mathrm{Det},t}\,\mathbf{v}_{i,\mathrm{Det},t}$ is the combined innovation consisting of the association probabilities $\beta_{i,\mathrm{Det},t} \in [0,1]$ (based on a Gaussian appearance model for spot-like particles, see Section 3.1) and the innovations $\mathbf{v}_{i,\mathrm{Det},t} = \mathbf{y}_{\mathrm{f},i,\mathrm{Det},t} - \mathbf{H}\,\hat{\mathbf{m}}_t$. $\mathbf{y}_{\mathrm{f},i,\mathrm{Det},t}$ are determined within elliptical regions around the fused detection $\mathbf{y}_{\mathrm{f},t}$ obtained by our MD-iCI. The Kalman gain matrix for $\mathbf{y}_{\mathrm{f},i,\mathrm{Det},t}$ is given by:

$$\mathbf{K}_{\mathrm{Det},t} = \hat{\mathbf{P}}_{\mathrm{f},t}\,\mathbf{H}^T\left(\mathbf{S}_{\mathrm{Det},t}\right)^{-1} \tag{4.13}$$

$$\mathbf{S}_{\mathrm{Det},t} = \mathbf{H}\,\hat{\mathbf{P}}_{\mathrm{f},t}\,\mathbf{H}^T + \mathbf{R}_{\mathrm{f,Det},t} \tag{4.14}$$

The innovation covariance matrix $\mathbf{S}_{\mathrm{Det},t} \in \mathbb{R}^{4\times4}$ reflects the uncertainty of the innovation $\mathbf{v}_{\mathrm{Det},t}$ and the uncertainty of $\mathbf{y}_{\mathrm{f},i,\mathrm{Det},t}$ is represented by the covariance matrix $\mathbf{R}_{\mathrm{f,Det},t} \in \mathbb{R}^{4\times4}$ obtained by MD-iCI. Note that the covariance matrix $\mathbf{R}_{\mathrm{f,Det},t}$ is changing over time in contrast to [12] where a fixed covariance matrix was used.

Second, the state $\mathbf{x}_t$ (with mean $\mathbf{m}_t$ and covariance matrix $\mathbf{P}_t$) is determined by the estimate $\mathbf{x}_{\mathrm{Det},t}$ (with mean $\mathbf{m}_{\mathrm{Det},t}$ and covariance matrix $\mathbf{P}_{\mathrm{Det,t}}$) and the *prediction-based* measurements $\mathbf{y}_{j,\mathrm{Pred},t}$ with

$$\mathbf{m}_t = \mathbf{m}_{\mathrm{Det},t} + \mathbf{K}_{\mathrm{Pred},t}\,\mathbf{v}_{\mathrm{Pred},t} \tag{4.15}$$

$$\mathbf{P}_t = \left(\mathbf{I} - \mathbf{K}_{\mathrm{Pred},t}\,\mathbf{H}\right)\mathbf{P}_{\mathrm{Det},t} \tag{4.16}$$

where $\mathbf{v}_{\mathrm{Pred},t} = \sum_{j=1}^{N_{\mathrm{Pred}}} \beta_{j,\mathrm{Pred},t}\,\mathbf{v}_{j,\mathrm{Pred},t}$ is the combined innovation with the association probabilities $\beta_{j,\mathrm{Pred},t} \in [0,1]$ (based on a Gaussian appearance model) and the innovations $\mathbf{v}_{j,\mathrm{Pred},t} = \mathbf{y}_{j,\mathrm{Pred},t} - \mathbf{H}\,\mathbf{m}_{\mathrm{Det},t}$. The Kalman gain matrix for $\mathbf{y}_{j,\mathrm{Pred},t}$ is given by:

$$\mathbf{K}_{\mathrm{Pred},t} = \mathbf{P}_{\mathrm{Det},t}\,\mathbf{H}^T\left(\mathbf{H}\,\mathbf{P}_{\mathrm{Det},t}\,\mathbf{H}^T + \mathbf{H}\,\hat{\mathbf{P}}_{\mathrm{f},t}\,\mathbf{H}^T\right)^{-1} \tag{4.17}$$

The detection-based measurements $\mathbf{y}_{\mathrm{f},i,\mathrm{Det},t}$ are determined within an elliptical sampling region centered at the position of the fused detection obtained by MD-iCI. The prediction-based measurements $\mathbf{y}_{j,\mathrm{Pred},t}$ are obtained analogously using the position of the predicted state.

For the assignment between predictions and measurements, we use a displacement-based correspondence finding approach [12], which exploits motion information from past and future time points. We employ displacements obtained by the two MS-PDAE with MD-iCI running in opposite temporal directions and use a graph-theoretical approach [102] to solve the correspondence problem.

5 Deep Learning for Particle Detection and Tracking

In this chapter, novel methods for particle detection and tracking based on deep neural networks are described. First, we present a novel approach which combines a convolutional neural network for particle detection with probabilistic data association for tracking. Second, we have developed a novel approach which uses deep learning for both particle detection and particle association. The approach combines a convolutional neural network for particle detection with an LSTM-based recurrent neural network for tracking. The work has been published in Wollmann/Ritter *et al.* [14] and Ritter *et al.* [15, 16].

5.1 Deep Learning-Based Detection and Bayesian Particle Tracking

We combine deep learning-based particle detection with Bayesian sequential estimation for tracking. Particles are detected by an hourglass-shaped deep learning network denoted as DetNet [14]. Associations between particles are determined by probabilistic data association with elliptical sampling (PDAE) which is based on Kalman filtering and particle filtering [48]. Our method is denoted as DetNet-PDAE. To increase the tracking performance, data association parameters that depend on the detection result are automatically determined using the hyperparameter optimization framework HyperHyper [197] described in Section 7.1.

Deep Learning for Particle Detection

Detection of fluorescent particles is performed by DetNet which uses a Deconvolution Network consisting of a contracting (pooling) and an expanding (unpooling) path and can handle objects at multiple scales naturally by its hourglass-shape. The DetNet architecture is outlined in Figure 5.1. A challenge in particle detection is the small object size and lack of complex shape information, and thus, data augmentation does not significantly increase the training data variability. Therefore, overfitting is likely to occur. With DetNet, we significantly decrease the number of parameters by reducing the number of extracted feature maps and reduce the size of the receptive field by employing pooling only two times instead of five times. Further, we do not

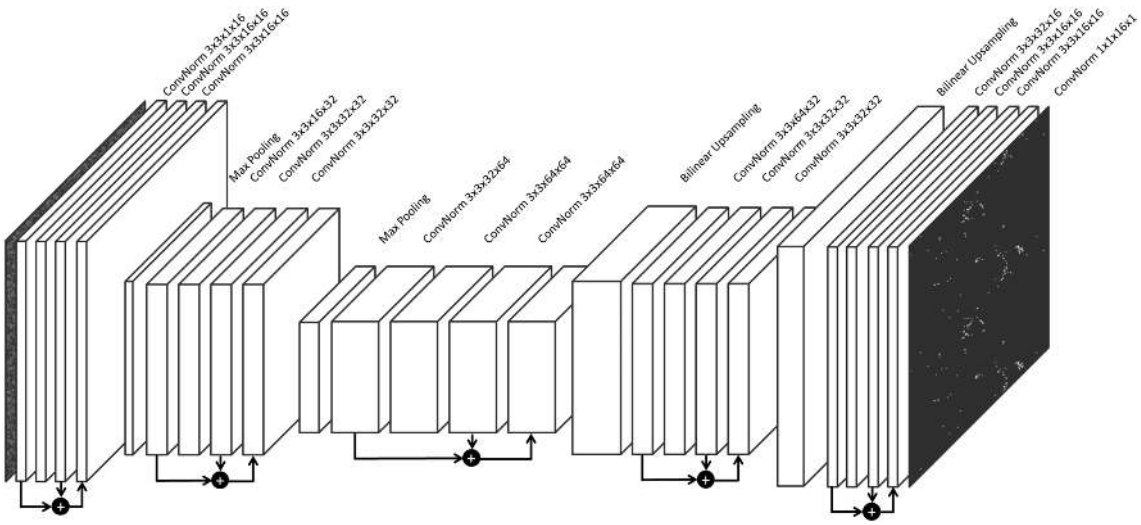


Figure 5.1: Deep neural network architecture of DetNet. The configurations of the specific layers are given above each layer.

exploit skipping connections as in [76] since detailed boundary information is not relevant for particle detection. We use instance normalization [95] since the batch normalization [97] requires a representative dataset, which is hardly available when using only few training samples. The convolutional layers are replaced by residual blocks [97], where the input $\mathbf{x} \in \mathbb{R}^{b \times m \times n \times i}$ of two consecutive convolutional layers $\mathcal{F}$ with parameters $\mathcal{W} \in \mathbb{R}^{k \times k \times i \times j}$ is added to the output to form a residual, where b is the batch size, $m \times n$ the spatial feature map size, i the number of input feature maps, j the number of output feature maps, and k the window size of the convolutional kernel:

$$\mathbf{x}_l = \mathbf{x}_{l-1} + \mathcal{F}\left(\mathbf{x}_{l-1}; \mathcal{W}_l\right) \tag{5.1}$$

We use residual blocks since the problem of gradient vanishing is reduced, which improves the efficiency of training deep architectures. The number of parameters is further reduced and checkerboard artifacts are avoided by employing bilinear upsampling (expanding path) instead of transposed convolutions.

DetNet was trained using a soft Dice loss and early stopping with the AMSGrad optimizer [198] and a learning rate of $l_{\text{init}} = 0.001$ as well as $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The soft Dice loss $\mathcal{L}(\mathbf{X}, \mathbf{Y})$ for the ground truth $\mathbf{Y}$ and prediction $\mathbf{X}$ performs implicit class balancing and penalizes easy samples compared to the Cross-Entropy (CE) loss:

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}) = -\frac{\sum_{i=1}^{N} 2 \cdot X_i \cdot Y_i + \epsilon}{\sum_{i=1}^{N} X_i + Y_i + \epsilon} \tag{5.2}$$

where N is the number of image pixels, and the constant ϵ is employed to avoid division by zero (we used $\epsilon = 10^{-6}$). The stability of the training was found to be improved by calculating the Dice loss over all N pixels in a batch instead of averaging the Dice loss over the single images. In contrast, training was not successful when

using standard CE due to the heavy class imbalance. We augmented the training data using random flipping, rotation, and cropping.

Bayesian Estimation for Particle Tracking

After particle detection, tracking of particles in our method is formulated as Bayesian sequential estimation problem. A particle at time point t is represented by a state $\mathbf{x}_t = (p_x, \dot{p}_x, p_y, \dot{p}_y, I_{\max}, \sigma_{xy})^T$ with position $\mathbf{p}_t = (p_x, p_y)^T$, velocity $\mathbf{v}_t = (\dot{p}_x, \dot{p}_y)^T$, and intensity information $\mathbf{I}_t = (I_{\max}, \sigma_{xy})^T$, where $I_{\max}$ is the maximum intensity and σ_{xy} the width of a spot-like structure. The state $\mathbf{x}_t$ is reflected by a noisy measurement $\mathbf{y}_t = \left(\overline{p}_x, \overline{p}_y, \overline{I}_{\max}, \overline{\sigma}_{xy}\right)^T$. Using a dynamic model $p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)$ and a measurement model $p\left(\mathbf{y}_t|\mathbf{x}_t\right)$, the aim of Bayesian sequential estimation is to calculate the posterior distribution $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ by Bayes' theorem [65]. The posterior distribution can be resolved in a closed form via the Kalman filter by assuming linear and Gaussian models. For time point t, DetNet provides a measurement $\mathbf{y}_t$. The uncertainty of $\mathbf{y}_t$ is reflected by the covariance matrix $\mathbf{R} \in \mathbb{R}^{4\times4}$. For particles that undergo directed motion or random walk, we use the dynamic models as in [48] with the covariance matrix $\mathbf{Q} \in \mathbb{R}^{6\times6}$ to compute the prediction $\hat{\mathbf{x}}_t$. An estimate for the mean $\mathbf{x}_t$ of $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ is calculated by:

$$\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t\, \nu_t \tag{5.3}$$

with the combined innovation $\boldsymbol{\nu}_t = \sum_{i=1}^{N_m} \beta_{i,t}\, \boldsymbol{\nu}_{i,t}$ and the Kalman gain

$$\mathbf{K}_t = \hat{\mathbf{P}}_t\, \mathbf{H}^T \left(\mathbf{H}\, \hat{\mathbf{P}}_t\, \mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{5.4}$$

The combined innovation $\boldsymbol{\nu}_t$ includes the association probabilities $\beta_{i,t}$ and innovations $\boldsymbol{\nu}_{i,t} = \mathbf{y}_{i,t} - \hat{\mathbf{y}}_t$ for N_m samples. The predicted measurement $\hat{\mathbf{y}}_t = \mathbf{H}\, \hat{\mathbf{x}}_t$ is based on the predicted state $\hat{\mathbf{x}}_t$ and the measurement model $\mathbf{H} \in \mathbb{R}^{4\times6}$. The measurements $\mathbf{y}_{i,t}$ are determined by elliptical sampling. For the association probabilities, we have $\beta_{0,t} + \sum_{i=1}^{N_m} \beta_{i,t} = 1$, where $\beta_{0,t}$ reflects the probability that no measurement corresponds to a tracked particle. The association probabilities $\beta_{i,t}$ can be interpreted as weights for the measurements $\mathbf{y}_{i,t}$ and depend on the expected degree of image noise σ_n. Therefore, σ_n depends on the detection performance of DetNet and has an influence on the tracking performance. In our approach, σ_n is automatically determined by hyperparameter optimization. The covariance matrix of the prediction $\hat{\mathbf{P}}_t \in \mathbb{R}^{6\times6}$ in (5.4) can be computed using a directed or random walk motion model $\mathbf{F} \in \mathbb{R}^{6\times6}$:

$$\hat{\mathbf{P}}_t = \mathbf{F}\, \mathbf{P}_{t-1}\, \mathbf{F}^T + \mathbf{Q} \tag{5.5}$$

Combining (5.4) and (5.5) shows that the update state in (5.3) and therefore the tracking performance depends on the covariance matrix $\mathbf{Q}$ of the motion model

and the covariance matrix $\mathbf{R}$ of the measurement model. In our approach, we use the elements of $\mathbf{Q}$ and $\mathbf{R}$ as hyperparameters and optimize them automatically. To optimize the hyperparameters σ_n, $\mathbf{Q}$, and $\mathbf{R}$, we use the distributed black-box optimization framework HyperHyper [197] (cf. Section 7.1). HyperHyper consists of a hyperparameter space definition, an optimizer that includes a hyperparameter candidate sampler and an optimization strategy, and an evaluation loop.

5.2 Deep Learning for Particle Detection and Tracking

We introduce a novel tracking approach for fluorescent particles in microscopy images which uses deep learning for both particle detection and particle association. We combine a convolutional neural network for particle detection with a long short-term memory (LSTM)-based recurrent neural network (RNN) for tracking. To the best of our knowledge, we are the first that suggest a deep learning approach for both particle detection and tracking. Particle detection is performed with the domain adapted Deconvolution Network DetNet [14]. This network has the advantages that it does not require a sliding window scheme (compared to [83]), all particles are detected at once, and it has a much lower number of parameters (compared to [84]). For correspondence finding we use a deep RNN with bidirectional LSTMs, denoted as DPHT [145]. This network exploits past and future information in both forward and backward direction as well as uses multiple track hypotheses. Assignment probabilities are determined jointly across multiple detections. Manual tuning of model parameters is not required, and manually labeled data is not needed for network training.

DetNet-DPHT combines a convolutional neural network for particle detection with a bidirectional RNN for tracking. Fluorescently labeled particles in microscopy images are detected using DetNet [14]. This network has an hourglass-shape architecture and can handle objects at multiple scales and does not require a sliding window scheme. All particles share full-image convolutional features and can be detected at once within an image. We trained DetNet using a soft Dice loss

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}) = -\frac{\sum_{i=1}^{N} 2 \cdot X_i \cdot Y_i + \epsilon}{\sum_{i=1}^{N} X_i + Y_i + \epsilon} \tag{5.6}$$

where $\mathbf{Y}$ is the ground truth, $\mathbf{X}$ the prediction, ϵ is a constant used to avoid division by zero ($\epsilon = 10^{-6}$), and N denotes the number of pixels in an image. With the soft Dice loss, we perform implicit class balancing and penalize easy samples. We used the AMSGrad optimizer [198] with early stopping and a learning rate of $l_{\text{init}} = 0.001$ with $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

The position information of the detected particles is used for association finding using a deep RNN denoted as Deep Particle Hypotheses Tracker (DPHT) [145]. Assignment probabilities are determined jointly across multiple detections, and

probabilities of missing detections are also computed. The network employs track hypotheses in consecutive frames and exploits information from past and future time points in both forward and backward direction by stacked bidirectional LSTMs. In addition, the network consists of fully-connected layers, 1D max-pooling layers, and temporal convolution layers. Training is performed using the loss function

$$\mathcal{L} = \mathcal{L}_{\mathbf{a}}(\mathbf{a}, \tilde{\mathbf{a}}) + \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}}) \tag{5.7}$$

with the cross-entropy loss $\mathcal{L}_{\mathbf{a}}(\mathbf{a}, \tilde{\mathbf{a}}) = -\sum_{j=0}^{M} \tilde{a}_j \cdot \log(a_j)$ which measures the deviation between the computed assignment probabilities $\mathbf{a}$ of M detections and ground truth $\tilde{\mathbf{a}}$, and the cross-entropy loss $\mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}}) = -(\tilde{\rho}_{\text{alive}} \log(\rho_{\text{alive}}) + \tilde{\rho}_{\text{dead}} \log(\rho_{\text{dead}}))$ which quantifies the deviation between the predicted existence probabilities ρ_{alive} and ρ_{dead} and the ground truth $\tilde{\rho}_{\text{alive}}$ and $\tilde{\rho}_{\text{dead}}$, respectively. We used the Adam optimizer [199] with early stopping, Gaussian dropout with a rate of $p = 0.25$, and a learning rate of $l_{\text{init}} = 0.001$ with $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

6 Experimental Results

In this chapter, experimental results of extensive evaluations of the detection and tracking methods in Chapter 3, 4, and 5 are presented. The developed methods are benchmarked on data of the Particle Tracking Challenge as well as live cell fluorescence microscopy image sequences displaying HCV proteins, HIV-1 particles, and chromatin structures. The experimental results show that the developed methods obtain state-of-the-art results or outperform existing methods.

6.1 Bayesian Smoothing for Particle Tracking

This section presents experimental results of the particle tracking methods proposed in Chapter 3 which are a two-filter probabilistic data association approach (S-PDAE, Section 3.2) and a Bayesian smoothing approach with multi-sensor data fusion (SMS-PDAE, Section 3.3).

6.1.1 Two-Filter Probabilistic Data Association for Particle Tracking

We evaluated our two-filter probabilistic data association approach (S-PDAE) (cf. Section 3.2) based on data of the Particle Tracking Challenge [185] for low signal-to-noise ratios. We also applied our approach to live cell microscopy images of human immunodeficiency virus type 1 (HIV-1) particles and hepatitis C virus (HCV) proteins.

6.1.1.1 Particle Tracking Challenge Data

We assessed the performance of the proposed S-PDAE approach based on data from the Particle Tracking Challenge [185] and performed a comparison with the overall top-three methods (Methods 5, 1, 2). Method 5 uses SEF [106] for particle localization and probabilistic data association with elliptical sampling (PDAE) [48]. Method 1 localizes particles by iterative centroid calculation and particle assignment is performed by combinatorial optimization [102]. Method 2 localizes particles by convolution with a disk shaped object model and performs track linking by MHT [131]. For our S-PDAE approach we use SEF [106] for particle localization.

We evaluated the tracking methods on data of the vesicle scenario for low signal-to-noise ratios of SNR = 1 and SNR = 2 as well as low and medium particle densities. Note that in [49] only high SNRs were considered. The time-resolved image sequences

Table 6.1: Performance metrics for vesicle scenario with SNR = 1 and different particle densities.

Metric	α	β	JSC_θ	JSC	$RMSE$
Low					
Method 5	0.211	**0.112**	0.185	**0.179**	2.119
Method 1	0.022	0.018	0.115	0.024	**1.568**
Method 2	0.175	0.104	**0.281**	0.178	2.423
S-PDAE	**0.225**	0.102	0.155	0.166	2.150
Medium					
Method 5	0.162	0.142	**0.458**	0.225	2.172
Method 1	0.027	0.026	0.300	0.034	**1.533**
Method 2	**0.198**	0.111	0.335	0.192	2.386
S-PDAE	*0.174*	**0.148**	0.450	**0.240**	2.188

consist of 100 images with 512×512 pixels. We computed the performance metrics α, β, JSC_θ, JSC, and $RMSE$ as in [185]. $\alpha \in [0,1]$ reflects the overall matching of ground truth with computed trajectories. Spurious tracks are quantified with $\beta \in [0,\alpha]$. The Jaccard similarity coefficient $JSC \in [0,1]$ represents the overall performance of particle detection, and $JSC_\theta \in [0,1]$ reflects the similarity of the entire tracks. $RMSE$ indicates the overall localization accuracy.

The results for SNR = 1 are shown in Table 6.1 and for SNR = 2 in Table 6.2. Bold numbers highlight the best result among the evaluated approaches. An italic score indicates an improvement of S-PDAE compared to PDAE (Method 5). It can be seen that our S-PDAE approach yields state-of-the-art results and best results in several cases. Considering the overall matching performance α for SNR = 1 and 2 and low particle density, S-PDAE yields the best result compared to all other methods. For SNR = 1 with medium density, S-PDAE achieves a better result for α than PDAE. Regarding β and JSC_θ, S-PDAE yields comparable results as PDAE for SNR = 1 and 2. $RMSE$ is slightly larger for S-PDAE compared to PDAE.

6.1.1.2 Evaluation on Live Cell Microscopy Images

We also evaluated the proposed S-PDAE approach based on live cell microscopy image sequences displaying fluorescently labeled human immunodeficiency virus type 1 (HIV-1) particles and hepatitis C virus (HCV) proteins. We used two image sequences with labelled HIV-1 particles [24] acquired from a focal plane in a 3D collagen structure [5] with an Ultra-View ERS spinning disk confocal microscope using a Hamamatsu C9100-50 camera. Both image sequences consist of 50 images with 1000×1000 pixels (Sequence 1 and 2). We also used one HCV image sequence with ApoE particles comprising 34 live cell images with 512×512 pixels, which was acquired with an Ultra-View ERS spinning disk confocal microscope mounted on a Nikon TE2000-E (Sequence 3).

Table 6.2: Performance metrics for vesicle scenario with SNR = 2 and different particle densities.

Metric	α	β	JSC_θ	JSC	$RMSE$
Low					
Method 5	0.652	0.590	0.763	0.717	1.118
Method 1	0.225	0.155	0.232	0.178	**0.796**
Method 2	0.661	0.607	**0.792**	0.742	1.219
S-PDAE	**0.706**	**0.652**	0.750	**0.800**	1.163
Medium					
Method 5	0.448	0.391	**0.664**	0.489	1.325
Method 1	0.398	0.298	0.411	0.340	**0.840**
Method 2	**0.517**	**0.417**	0.629	**0.510**	1.254
S-PDAE	0.443	0.383	0.653	0.489	1.368

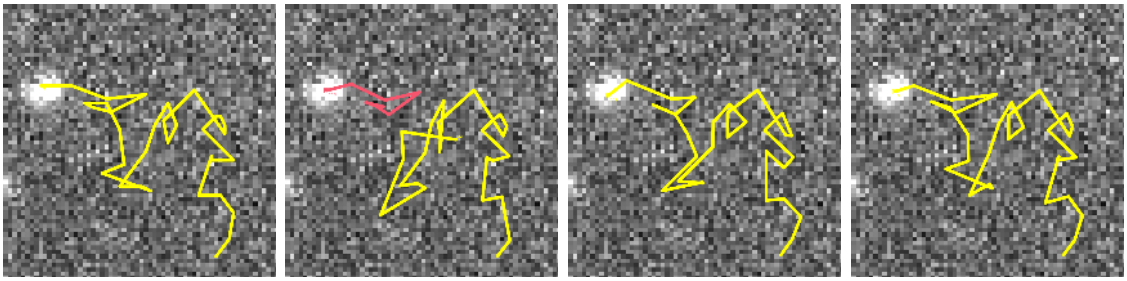
Table 6.3: Results for P_{Track} for live cell microscopy data.

Seq.	PT	KF	PDAE	S-PDAE
1	0.075	0.087	0.110	**0.119**
2	0.137	0.129	0.156	**0.200**
3	0.309	0.411	0.623	**0.630**

We compared the S-PDAE approach with ParticleTracker (PT) [102], a Kalman filter approach (KF) [200], and PDAE [48]. KF uses an LoG filter for particle localization and links tracks by linear assignment similar to u-track [103]. For all approaches, optimal parameters were determined by grid search.

For performance evaluation we calculated the tracking accuracy P_{track}, which quantifies the percentage of correctly computed trajectories relative to the number of true trajectories as in [48]. Ground truth trajectories were determined by manual annotation using the Manual Tracking plug-in in ImageJ. Sequences 1 and 2 contain 16 ground truth trajectories, and Sequence 3 comprises 108 ground truth trajectories.

Table 6.3 shows the result for P_{track} for all evaluated approaches. It turns out that S-PDAE outperforms the other methods for all image sequences. The improvement is clearly visible for an example trajectory from Sequence 1 with complex motion shown in Fig. 6.1. The data is challenging due to image noise, low contrast, and clutter. It can be seen that KF yields a broken trajectory, whereas PDAE and S-PDAE can track the particle over all time points. S-PDAE yields the best result compared to ground truth.



(a) Ground truth (b) KF (c) PDAE (d) S-PDAE

Figure 6.1: Ground truth and computed trajectories for a HIV-1 particle from Sequence 1 using different tracking methods.

Summary For data of the Particle Tracking Challenge with low SNR values, our approach achieved state-of-the-art results and best results for several cases. For live cell HIV-1 and HCV microscopy image data it turned out that S-PDAE outperformed existing methods.

6.1.2 Bayesian Smoothing and Multi-Sensor Data Fusion for Particle Tracking

We evaluated our SMS-PDAE (Section 3.3) approach on the Particle Tracking Challenge data as well as using different time-lapse fluorescence microscopy images of human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) proteins. SMS-PDAE consists of multi-sensor data fusion with probabilistic data association (cf. Section 3.3.1) and Bayesian smoothing using covariance intersection (cf. Section 3.3.2).

6.1.2.1 Particle Tracking Challenge Data

We used data from the Particle Tracking Challenge [185] and performed a comparison of the SMS-PDAE approach with the overall top-three performing methods (Method 5, 1, 2) of the challenge. Method 5 uses the spot-enhancing filter (SEF) for particle detection and PDAE for particle linking [48]. Method 1 localizes particles by iterative centroid calculation and assigns particles by combinatorial optimization [102]. Method 2 detects particles by convolution with a disk shaped object and uses multiple-hypothesis tracking for correspondence finding [131]. We also compared our approach with the piecewise-stationary motion model smoother (PMMS) [49] and our previous two-filter smoothing approach based on PDAE (S-PDAE) [11]. The PMMS approach detects particles by a blob detector using the difference of Gaussians and employs piecewise-stationary motion modeling for correspondence finding [49]. The S-PDAE detects particles by SEF and uses a two-filter smoothing approach for correspondence finding [11]. In addition, we compared the SMS-PDAE with different variants of it, the (non-smoothed) MS-PDAE with position-based correspon-

dence finding (MS-PDAE (Pos)) and with displacement-based correspondence finding (MS-PDAE (Displ)). For all variants we used SEF [106] for particle detection. SEF consists of applying a Laplacian-of-Gaussian (LoG) filter with standard deviation σ_{LoG}, followed by thresholding using the mean of the absolute values of the filter responses plus a factor c times the standard deviation and determination of local maxima. For the approaches based on PDAE, we used the same parameters for σ_{LoG} (typically 3 pixels for low SNR levels and 1.5 pixels for high SNR levels) and c (typically 3 for low SNR levels and 5 for high SNR levels).

We assessed the different tracking approaches for the vesicle scenario from the Particle Tracking Challenge data consisting of 12 time-lapse image sequences. We considered all SNR levels (SNR=$1, 2, 4, 7$) and all object densities (low, medium, high). The number of particles ranges from around 100 for low object density to around 1000 for high object density. Each of the 12 image sequences consists of 100 images with an image size of 512×512 pixels.

The performance was quantified using the metrics α, β, JSC_θ, JSC, and $RMSE$ from the Particle Tracking Challenge [185]. The metrics α and β account for association errors and localization errors at position level. $\alpha \in [0, 1]$ quantifies the overall matching quality between ground truth and computed trajectories. A perfect matching is reflected by $\alpha = 1$, whereas $\alpha = 0$ indicates no matching. $\beta \in [0, \alpha]$ considers additionally spurious trajectories. The performance at track level is evaluated by the Jaccard similarity coefficient $JSC_\theta \in [0, 1]$. The detection performance is quantified by the Jaccard similarity coefficient $JSC \in [0, 1]$, whereas the localization performance is evaluated by the root mean-square error ($RMSE$). Note that for the PMMS [49] the results are provided for SNR=4 and 7, but not for SNR=2 and 1, and $RMSE$ has not been given. Further, all performance values for the PMMS are given only up to two decimal digits, whereas here the values are reported up to three decimal digits.

Table 6.4 provides for the different tracking methods the mean values of all performance metrics over all object densities for low SNR levels (SNR = 1 and 2) and high SNR levels (SNR = 4 and 7). The best performing method is marked in bold. The SMS-PDAE outperforms all methods for all SNR levels for the tracking score β and the detection score JSC. For low SNR levels, SMS-PDAE yields the best result for three of the five metrics (β, JSC_θ, JSC). For high SNR levels, SMS-PDAE outperforms PMMS, Method 2, and S-PDAE for all metrics. MS-PDAE (Displ) yields improved results compared to Method 5 (PDAE) for β, JSC_θ, and JSC for low SNR levels, and for β and $RMSE$ for high SNR levels. Further, MS-PDAE (Displ) yields better results than MS-PDAE (Pos) for the tracking scores α and β for all SNR levels.

Table 6.5 shows the mean value of all metrics over all SNR levels for different object densities. The mean values for PMMS were not computed since values are only available for high SNR levels (SNR = 4 and 7). The SMS-PDAE outperforms all methods for all object densities for β. For medium and high object density,

Table 6.4: Mean performance metrics over all object densities for low SNR levels (SNR=1 and 2) and high SNR levels (SNR=4 and 7) for the Particle Tracking Challenge data.

Metric	α	β	JSC_θ	JSC	$RMSE$
Low SNR levels					
Method 5	0.327	0.275	0.523	0.365	1.752
Method 1	0.176	0.130	0.278	0.154	**1.281**
Method 2	**0.345**	0.261	0.479	0.350	1.899
PMMS	–	–	–	–	–
S-PDAE	0.334	0.276	0.489	0.372	1.797
MS-PDAE (Pos)	0.316	0.272	**0.530**	0.359	1.757
MS-PDAE (Displ)	0.323	0.277	0.527	0.367	1.777
SMS-PDAE	0.335	**0.286**	**0.530**	**0.376**	1.699
High SNR levels					
Method 5	0.685	0.627	**0.796**	0.677	0.694
Method 1	**0.698**	0.625	0.758	0.664	**0.561**
Method 2	0.610	0.555	0.778	0.627	0.900
PMMS	0.688	0.632	0.783	0.667	-
S-PDAE	0.648	0.588	0.775	0.645	0.788
MS-PDAE (Pos)	0.683	0.630	0.791	0.672	0.649
MS-PDAE (Displ)	0.685	0.631	0.781	0.672	0.674
SMS-PDAE	0.691	**0.637**	0.785	**0.678**	0.663

SMS-PDAE yields the best results for three of the five metrics (α, β, JSC). MS-PDAE (Displ) yields better results than MS-PDAE (Pos) for all object densities for α, β, and JSC. For low object densities, MS-PDAE (Displ) and MS-PDAE (Pos) obtain improved results for β, JSC_θ, and JSC compared to Method 5 (PDAE). For medium and high object densities, MS-PDAE (Displ) and MS-PDAE (Pos) outperform Method 5 for $RMSE$.

The overall performance as mean value of the performance metrics over all SNR levels and all object densities is provided in Table 6.6. The mean values for PMMS were not computed since values are only available for high SNR levels (SNR = 4 and 7). As can be seen in Table 6.6, SMS-PDAE yields the best results for three of the five metrics. SMS-PDAE outperforms all methods for the tracking metrics α and β, which represent the performance of correspondence finding, and the metric JSC, which reflects the detection performance. MS-PDAE (Displ) yields better results than MS-PDAE (Pos) for α, β, and JSC. MS-PDAE (Pos) yields the best result for JSC_θ.

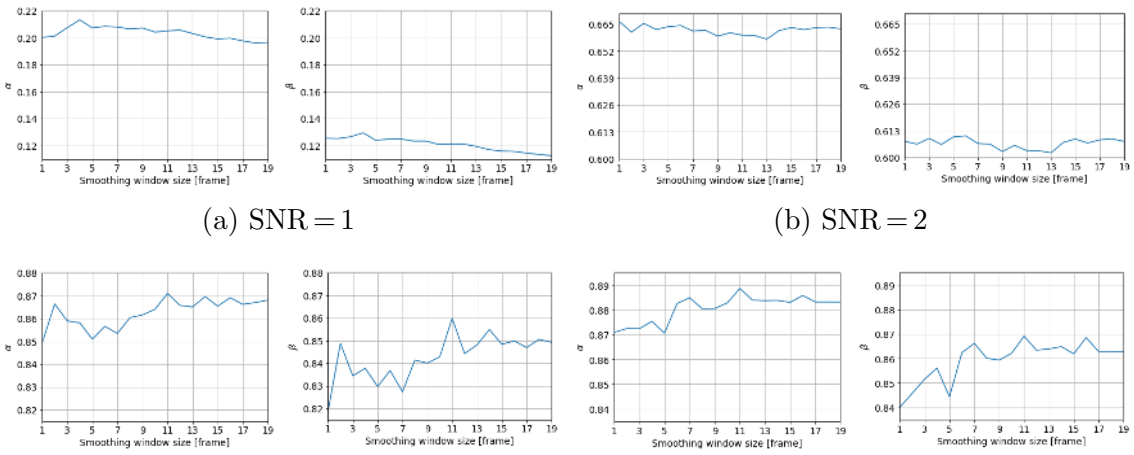
We also determined the computation time of our approach. We used an image sequence of the vesicle scenario consisting of 100 images (512×512 pixels) with SNR = 4 and medium object density. PDAE, MS-PDAE (Pos), and MS-PDAE (Displ) took 2.11 min, 2.13 min, and 2.04 min, respectively. For S-PDAE and SMS-PDAE the computation time was 3.46 min and 7.23 min, respectively. Computations were

Table 6.5: Mean performance metrics over all SNR levels for different object densities (low, medium, high) for the Particle Tracking Challenge data.

Metric	α	β	JSC_θ	JSC	$RMSE$
Low					
Method 5	0.654	0.603	0.699	0.675	1.011
Method 1	0.483	0.439	0.512	0.467	**0.774**
Method 2	0.602	0.557	**0.723**	0.650	1.252
PMMS	–	–	–	–	–
S-PDAE	**0.664**	0.609	0.689	**0.685**	1.050
MS-PDAE (Pos)	0.649	0.608	0.705	0.675	1.018
MS-PDAE (Displ)	0.649	0.608	0.714	0.678	1.034
SMS-PDAE	0.658	**0.615**	0.719	0.683	0.984
Medium					
Method 5	0.486	0.432	0.670	0.500	1.752
Method 1	0.453	0.388	0.559	0.419	**0.881**
Method 2	0.477	0.397	0.620	0.475	1.360
PMMS	–	–	–	–	–
S-PDAE	0.470	0.412	0.659	0.488	1.288
MS-PDAE (Pos)	0.477	0.427	**0.671**	0.490	1.217
MS-PDAE (Displ)	0.479	0.428	0.662	0.493	1.250
SMS-PDAE	**0.490**	**0.438**	0.665	**0.503**	1.194
High					
Method 5	0.378	0.319	**0.609**	0.387	1.429
Method 1	0.375	0.304	0.484	0.340	**1.108**
Method 2	0.354	0.270	0.538	0.341	1.587
PMMS	–	–	–	–	–
S-PDAE	0.340	0.273	0.548	0.351	1.540
MS-PDAE (Pos)	0.373	0.319	0.606	0.381	1.373
MS-PDAE (Displ)	0.385	0.326	0.586	0.388	1.394
SMS-PDAE	**0.391**	**0.332**	0.588	**0.395**	1.365

Table 6.6: Mean performance metrics over all SNR levels and all object densities for the Particle Tracking Challenge data.

Metric	α	β	JSC_θ	JSC	$RMSE$
Method 5	0.506	0.451	0.659	0.521	1.223
Method 1	0.437	0.377	0.518	0.409	**0.921**
Method 2	0.478	0.408	0.629	0.489	1.399
PMMS	–	–	–	–	–
S-PDAE	0.491	0.432	0.632	0.508	1.292
MS-PDAE (Pos)	0.500	0.451	**0.660**	0.516	1.203
MS-PDAE (Displ)	0.504	0.454	0.654	0.520	1.226
SMS-PDAE	**0.513**	**0.461**	0.657	**0.527**	1.181



(a) SNR = 1 (b) SNR = 2

(c) SNR = 4 (d) SNR = 7

Figure 6.2: Impact of the smoothing window size of SMS-PDAE on the performance metrics α and β for different SNR levels.

performed on a workstation with Intel Core i7-6700K CPU (four cores with 4 GHz) and 32 GB RAM running Linux.

6.1.2.2 Impact of Smoothing Window Size on Tracking Performance

The tracking performance of the SMS-PDAE depends on the smoothing window size, which defines the number of used future time points (equal to the number of past points) for state estimation and correspondence finding. We analyzed the impact of the window size on the tracking performance metrics α and β for the vesicle scenario for all SNR levels (SNR=$1, 2, 4, 7$) with low object density which we choose to focus on state estimation using covariance intersection. For each SNR level we varied the window size from 1 to 19 frames.

The results are shown in Fig. 6.2. For SNR = 1, best results for α and β are obtained with a window of 4 frames. Using more frames leads to a decrease in tracking performance. For SNR = 2, best results for α are obtained with a window of 1 or 3 frames while best results for β are achieved using 5 or 6 frames. For high SNR levels (SNR = 4 and 7), the best results are obtained with a window size of 11 frames. In our experiments, we set the smoothing window size mainly according to the SNR level in the image data. For low SNR levels (SNR = $1, 2$) we typically used window sizes of 4 or 5. For high SNR levels (SNR = $4, 7$) we typically employed larger window sizes of 7, 10, or 11. For the HIV-1 and HCV time-lapse microscopy data we typically used window sizes of 3, 5, and 7.

6.1.2.3 Evaluation on Time-Lapse Fluorescence Microscopy Images

We also evaluated the SMS-PDAE using challenging time-lapse microscopy data sets containing human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) proteins imaged with different types of fluorescence microscopes and different

spatial-temporal resolutions. In total, we used 12 microscopy image sequences and compared the results with those of previous methods.

HIV-1 Live Cell Widefield Microscopy Images We applied the SMS-PDAE to nine HIV-1 live cell microscopy image sequences displaying double labeled HIV-1 particles within incubated HeLa cells [24]. All image sequences were acquired using a fluorescence widefield setup based on a Zeiss Axiovert 200 M microscope with a Roper Scientific Cascade II EM-CCD camera. The image size for the first four sequences is 256×256 pixels and for the last five sequences 512×512 pixels. The spatial resolution is 160 nm/pixel and the temporal resolution is 100 ms/frame. The videos comprise 150 up to 400 images. The data is challenging due to low SNR, clutter, cellular autofluorescence, and out of focus movement.

We performed a comparison with the ParticleTracker (PT) [102], the uTrack approach [103], a multiple-hypothesis tracking approach (MHT) [201], and a Kalman filter approach (Kalman) [48]. PT assigns particles by combinatorial optimization, and uTrack is based on a linear assignment method. MHT is based on a Kalman filter with multiple motion models [185, 201]. "Kalman" uses a global nearest neighbor (GNN) scheme for correspondence finding [102]. Further, we compared the SMS-PDAE with the PDAE using a random walk motion model and the PDAE using an interacting motion model (IMM) including both a random walk motion model and a directed motion model. We also included the S-PDAE [11], MS-PDAE (Pos) and MS-PDAE (Displ) in the comparison.

We quantified the methods by the tracking accuracy measure $P_{\text{track}} \in [0, 1]$

$$P_{\text{track}} = \frac{n_{\text{track,correct}}}{n_{\text{track,total}}} \tag{6.1}$$

where $n_{\text{track,correct}}$ is the number of correctly computed trajectories determined as Gaussian weighted sum of the percentage of correctly tracked time steps (positions within an Euclidean distance of five pixels to the ground truth). $n_{\text{track,total}}$ represents the number of ground truth trajectories [48]. Ground truth was obtained by manual annotation between five and 43 ground truth trajectories for each image sequence using MetaMorph. Since annotating all trajectories in all sequences is not feasible, the performance metrics of the Particle Tracking Challenge cannot be computed. The results of the different approaches for all nine image sequences is shown in Table 6.7. Besides P_{track}, we also determined its standard deviation as a measure of robustness. The best performing method is marked in bold. It turns out that the SMS-PDAE yields the overall best performance with a mean P_{track} of 90 %. SMS-PDAE outperforms all methods for four image sequences and yields the best performance together with other methods for additionally four image sequences. Further, SMS-PDAE, MS-PDAE (Displ), PDAE, and PDAE (IMM) yield the lowest standard deviation of 8 % (highest robustness). The second best overall result is obtained by PDAE (IMM) with a mean P_{track} of 89 %, however, this method uses a

Table 6.7: Tracking results for HIV-1 live cell widefield image sequences in terms of the tracking accuracy P_{track} [%].

Seq.	1	2	3	4	5	6	7	8	9	Mean	Std. dev.
PT	73	55	63	67	**94**	64	82	74	70	71	11
Kalman	96	81	90	82	**94**	68	86	77	81	84	9
uTrack	75	64	52	66	93	66	85	75	66	71	12
MHT	91	68	91	78	**94**	71	87	77	83	82	9
PDAE	97	90	98	88	**94**	76	87	**79**	83	88	**8**
PDAE (IMM)	98	**92**	99	89	**94**	75	88	**79**	**84**	89	**8**
S-PDAE	98	**92**	90	89	**94**	74	84	71	80	86	9
MS-PDAE (Pos)	96	88	97	87	**94**	73	82	68	70	84	12
MS-PDAE (Displ)	**99**	**92**	97	89	**94**	78	88	76	82	88	**8**
SMS-PDAE	**99**	**92**	**100**	**91**	**94**	**79**	**92**	78	**84**	**90**	**8**

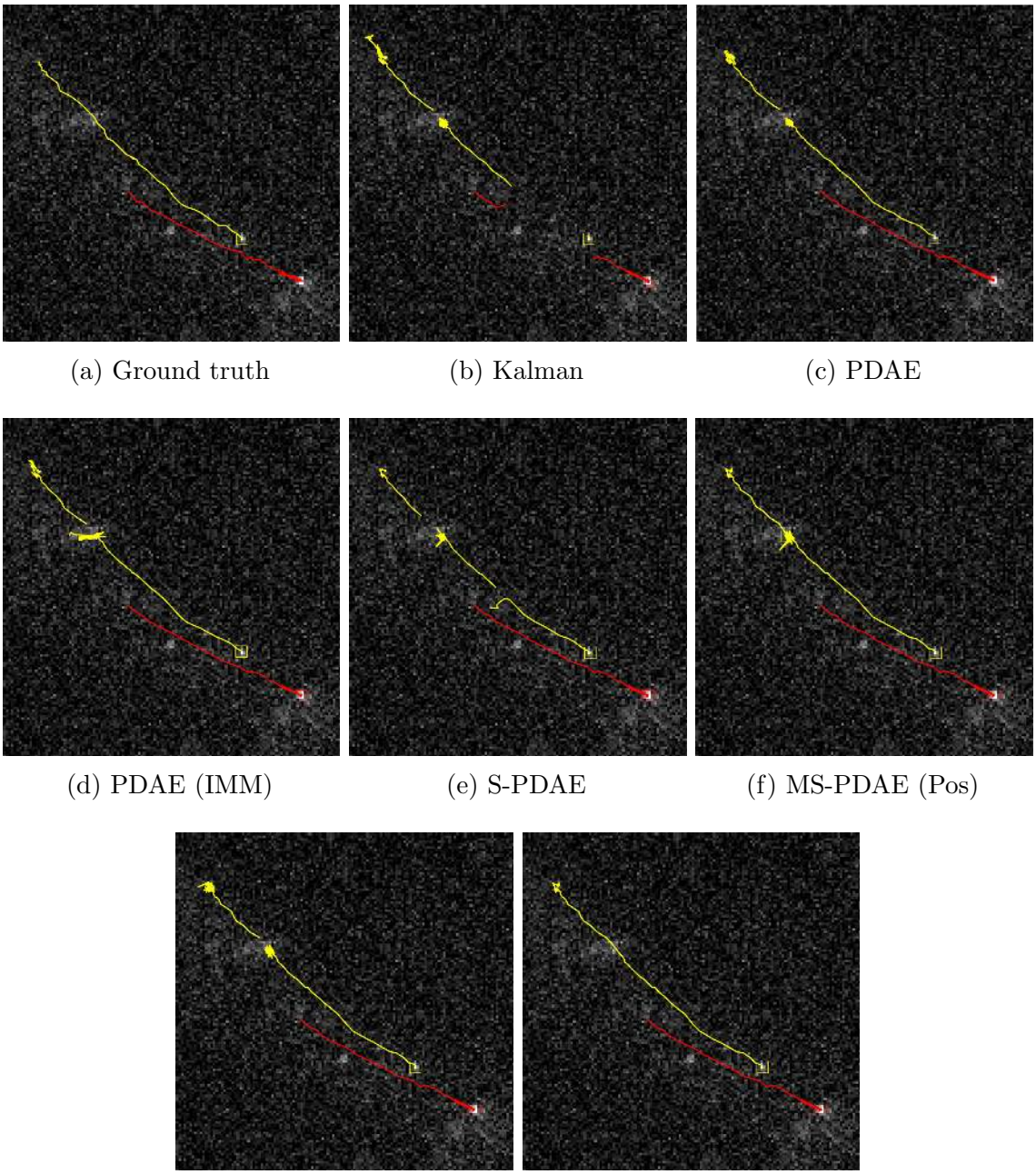
multiple motion model compared to the other PDAE approaches. MS-PDAE (Displ) yields a same mean P_{track} of 88 % as PDAE, but outperforms PDAE for five of nine image sequences. MS-PDAE (Displ) performs better than MS-PDAE (Pos) (mean P_{track} of 84 %).

In Fig. 6.3, two example trajectories for different tracking approaches are shown. Only the SMS-PDAE yields two complete trajectories. By using information from both past and future time points, the SMS-PDAE is able to correctly determine the correspondence for the yellow trajectory. The other approaches (except Kalman) yield a complete red trajectory, but the yellow trajectory has gaps.

HIV-1 Time-Lapse Spinning Disc Confocal Microscopy Images The second HIV-1 time-lapse imaging data set consists of two image sequences with labeled HIV-1 particles imaged from a confocal plane in a 3D collagen matrix [19, 5] with an UltraView ERS spinning disc confocal microscope using a Hamamatsu C9100-50 camera. Both image sequences comprise 50 images with an image size of 1000×1000 pixels. The spatial resolution is 167 nm/pixel and the temporal resolution is 200 ms/frame. The image sequences are challenging due to low SNR, out of focus movement, and complex motion of the viral particles.

We compared the SMS-PDAE with PT, uTrack, MHT, KF [200], PDAE, S-PDAE, MS-PDAE (Pos), and MS-PDAE (Displ). KF is a Kalman filter approach with an LoG filter for particle detection and linear assignment for correspondence finding. 16 ground truth trajectories for each image sequence were manual annotated using the Manual Tracking plug-in in ImageJ [202].

The tracking results for all approaches are shown in Table 6.8, and the best performing method is marked in bold. It can be seen that the SMS-PDAE obtains the best result for both image sequences. MS-PDAE (Displ) yields a better result than MS-PDAE (Pos), and both approaches outperform the previous PDAE method. An example trajectory is shown in Fig. 6.4. Only the SMS-PDAE obtains a complete



(a) Ground truth (b) Kalman (c) PDAE

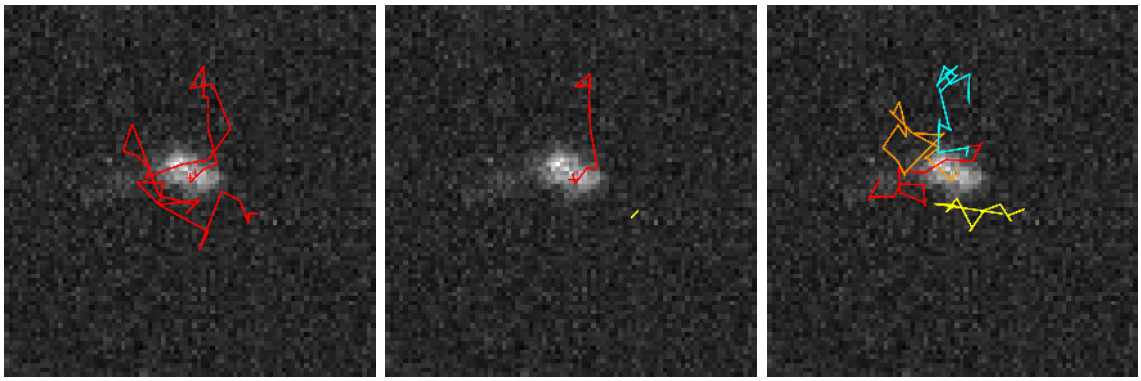
(d) PDAE (IMM) (e) S-PDAE (f) MS-PDAE (Pos)

(g) MS-PDAE (Displ) (h) SMS-PDAE

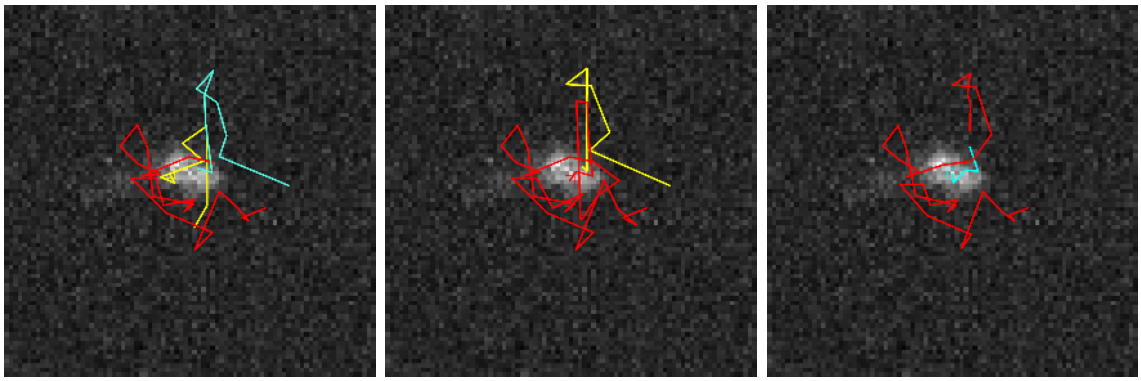
Figure 6.3: Ground truth and tracking results for HIV-1 live cell widefield microscopy images (time point $t = 297$). Only the SMS-PDAE yields a complete yellow trajectory and correct correspondences.

trajectory.

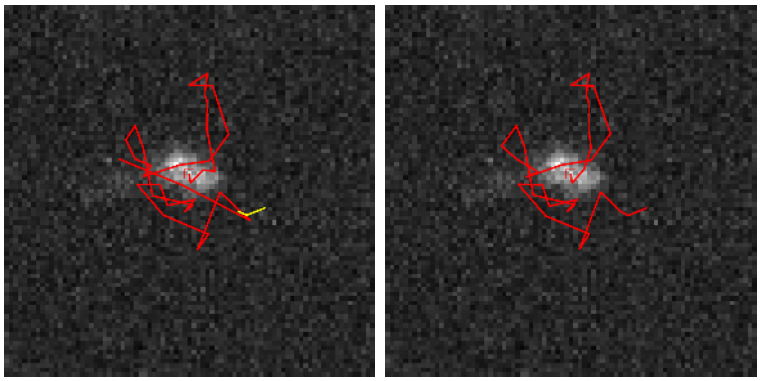
HCV Live Cell Spinning Disc Confocal Microscopy Images We also evaluated the SMS-PDAE for challenging live cell microscopy data displaying fluorescently labeled ApoE proteins within HCV proteins expressing Huh7/LunetCD81H cells [18]. The considered image sequence comprises 34 images with an image size of 512×512



(a) Ground truth (b) PT (c) MHT



(d) PDAE (e) S-PDAE (f) MS-PDAE (Pos)



(g) MS-PDAE (Displ) (h) SMS-PDAE

Figure 6.4: Ground truth and tracking results for HIV-1 time-lapse spinning disc confocal microscopy images of fluorescently labeled HIV-1 particles embedded in a 3D collagen matrix (time point $t = 34$). Only the SMS-PDAE yields a complete trajectory and correct correspondences.

Table 6.8: Tracking results for HIV-1 time-lapse spinning disk confocal image sequences in terms of the tracking accuracy P_{track} [%].

Seq.	PT	KF	uTrack	MHT	PDAE	S-PDAE	MS-PDAE (Pos)	MS-PDAE (Displ)	SMS-PDAE
1	7.5	8.7	2.5	9.8	11.0	11.9	11.4	11.5	**13.3**
2	13.7	12.9	10.3	19.8	15.6	20.0	17.0	18.4	**21.0**

Table 6.9: Tracking results for HCV live cell spinning disk confocal image data in terms of the tracking accuracy P_{track} [%].

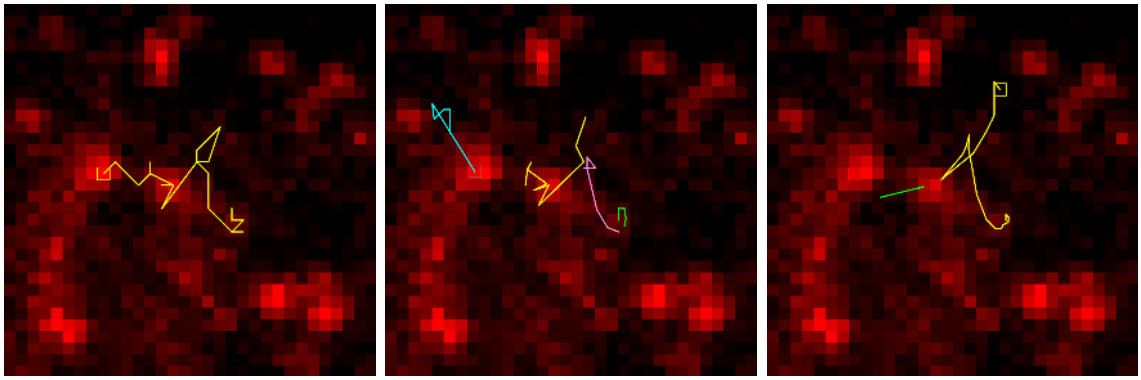
PT	KF	uTrack	MHT	PDAE	S-PDAE	MS-PDAE (Pos)	MS-PDAE (Displ)	SMS-PDAE
30.9	41.1	39.6	61.3	62.3	63.0	64.9	69.0	**69.3**

pixels, and was acquired with an Ultra-View ERS spinning disk confocal microscope mounted on a Nikon TE2000-E. The spatial resolution is 220 nm/pixel and the temporal resolution is 37 s/frame. The data is challenging due to clutter, high object density, and low SNR.

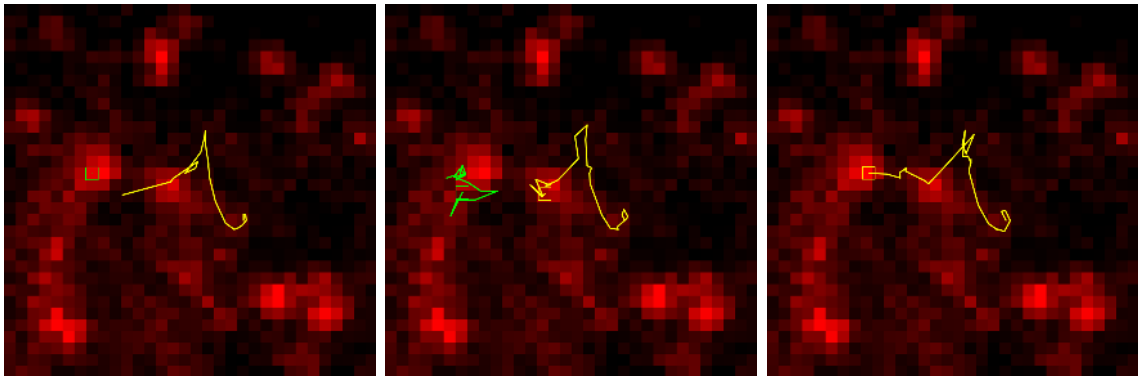
We compared the SMS-PDAE with PT, uTrack, MHT, KF, PDAE, S-PDAE, MS-PDAE (Pos), and MS-PDAE (Displ). We manually annotated 108 ground truth trajectories using the Manual Tracking plug-in in ImageJ [202].

The tracking results for all approaches are shown in Table 6.9. The SMS-PDAE outperforms all other methods. MS-PDAE (Displ) yields a better result than MS-PDAE (Pos), and both approaches outperform PDAE and S-PDAE. Example results for different tracking approaches are shown in Fig. 6.5. Only SMS-PDAE yields a complete trajectory due to incorporating information from both past and future time points. The S-PDAE yields an incomplete trajectory with a false correspondence at the end of the trajectory (left side in the image). All other methods obtain broken trajectories. In Fig. 6.6, tracking results for a 96×96 pixels image section for MS-PDAE (Displ) and SMS-PDAE are shown. SMS-PDAE yields more complete trajectories which are less often broken.

Summary A quantitative evaluation on the Particle Tracking Challenge data revealed that the SMS-PDAE obtains state-of-the-art results or outperforms previous methods. In particular, we demonstrated that the sequential multi-sensor data fusion approach combined with Bayesian smoothing and displacement-based correspondence finding improves tracking compared to the previous PDAE method. We also benchmarked the SMS-PDAE approach on 12 challenging time-lapse fluorescence microscopy image sequences acquired with different types of microscopes, different spatial-temporal resolutions, and different types of viruses. We found that our approach outperforms previous methods for both HIV-1 and HCV microscopy data including high object density, low SNR, clutter, complex motion, and out of focus motion.

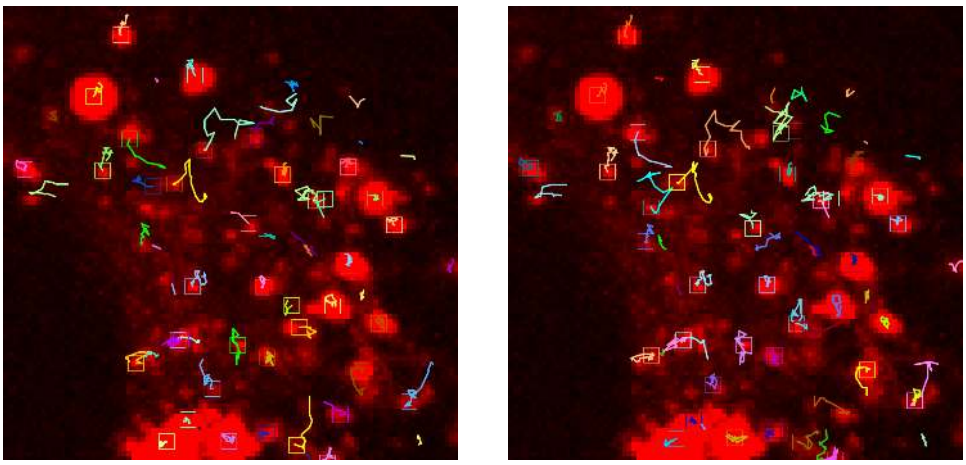


(a) Ground truth (b) MHT (c) PDAE



(d) S-PDAE (e) MS-PDAE (Displ) (f) SMS-PDAE

Figure 6.5: Ground truth and tracking results for different methods for HCV live cell spinning disk confocal microscopy images of fluorescently labeled ApoE proteins (time point $t = 27$). Only SMS-PDAE yields a complete trajectory and correct correspondences.



(a) MS-PDAE (Displ) (b) SMS-PDAE

Figure 6.6: Tracking results for MS-PDAE (Displ) and SMS-PDAE for a 96×96 pixels section of the HCV live cell spinning disk confocal microscopy images of fluorescently labeled ApoE proteins (time point $t = 21$).

6.2 Multi-Detector Fusion and Bayesian Smoothing for Particle Tracking

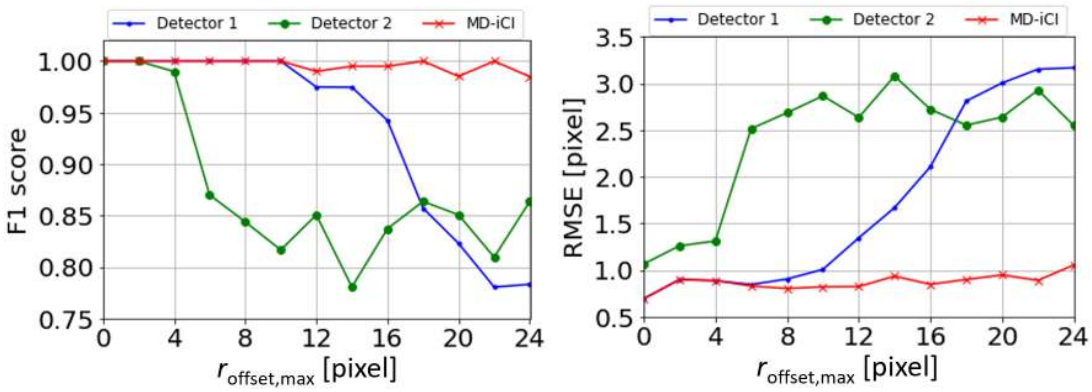
This section presents experimental results of the novel particle tracking method MD-BS which is based on multi-detector fusion and Bayesian smoothing. MD-BS is described in Chapter 4 and employs multi-detector fusion with the novel intensity-based covariance intersection approach MD-iCI (cf. Section 4.2). We assessed our MD-iCI on synthetic image data displaying particles with heterogeneous size and demonstrated that MD-iCI improves particle detection and localization compared to existing detection methods. Our MD-BS for particle tracking is evaluated using data from the Particle Tracking Challenge and yielded state-of-the-art results or better results than previous methods. Further, we performed a quantitative evaluation for challenging live cell time-lapse microscopy image data showing different subcellular structures, namely hepatitis C virus (HCV) associated proteins and fluorescently labeled chromatin structures acquired at different spatial-temporal resolutions. Our approach obtained superior results compared to existing methods.

Synthetic Data

We assessed the detection and localization performance of MD-iCI on synthetic image data showing particles with heterogeneous size. We generated synthetic images with 512×512 pixels that display 100 particles represented by a 2D Gaussian function. For each image, a particle has a radius (standard deviation of the 2D Gaussian function) between r_{min} and r_{max}, sampled from a uniform distribution. For r_{min} we used 4 pixels, while $r_{\text{max}} = r_{\text{min}} + r_{\text{offset, max}}$ and $r_{\text{offset, max}}$ varies for the different images (from 0 to 24 pixels) and defines the variation of particle sizes (heterogeneity) in an image. To simulate camera noise (e.g., CCD) the images are corrupted by Poisson noise [102].

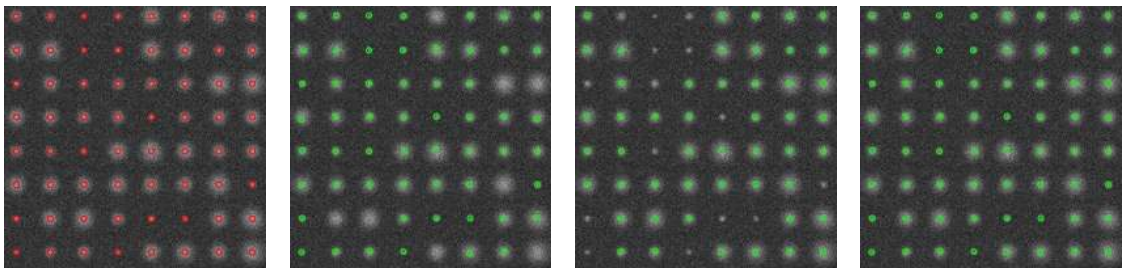
We compared MD-iCI with two different single-scale spot-enhancing filters (SEF) [106]. SEF uses a Laplacian-of-Gaussian (LoG) filter with standard deviation σ_{LoG} followed by thresholding. The threshold is determined by the mean of the absolute values of the filter responses plus a factor c times the standard deviation. This method is often used for particle detection.

The detection performance is quantified by the F1 score $\in [0, 1]$, where a value of 1 represents a perfect result. The localization performance is quantified by the mean $RMSE \in [0, 5]$ between detected particles and ground truth. The assignment between detected particles and ground truth was determined by the Munkres algorithm [122] with a maximal gating distance of 5 pixels. To compare the detection approaches, we considered the $RMSE$ of true positive as well as false negative detections. This has the advantage that the number of considered detections is the same for all methods. If a ground truth particle is not matched with a computed detection, a localization error of 5 pixels (maximal gating distance) is used.



(a) Detection performance (b) Localization performance

Figure 6.7: Performance of MD-iCI (red curve) compared to two SEFs with $\sigma_{\text{LoG}} = 3.0$ (blue) and $\sigma_{\text{LoG}} = 8.0$ (green) for images with increasing $r_{\text{offset, max}}$ (maximum particle radii offsets) corresponding to increasing particle size heterogeneity. MD-iCI fuses the detections of the two SEFs. a) Detection performance by the F1 score. b) Localization performance by $RMSE$.

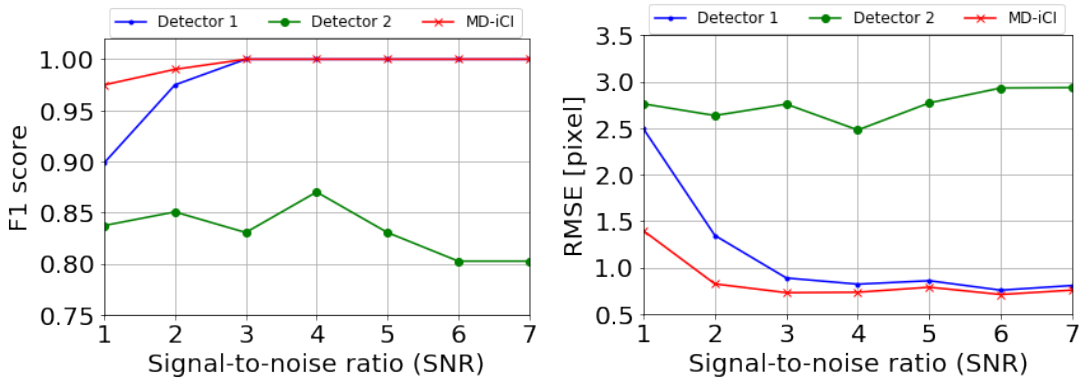


(a) Ground truth (b) SEF $\sigma_{\text{LoG}} = 3.0$ (c) SEF $\sigma_{\text{LoG}} = 8.0$ (d) MD-iCI

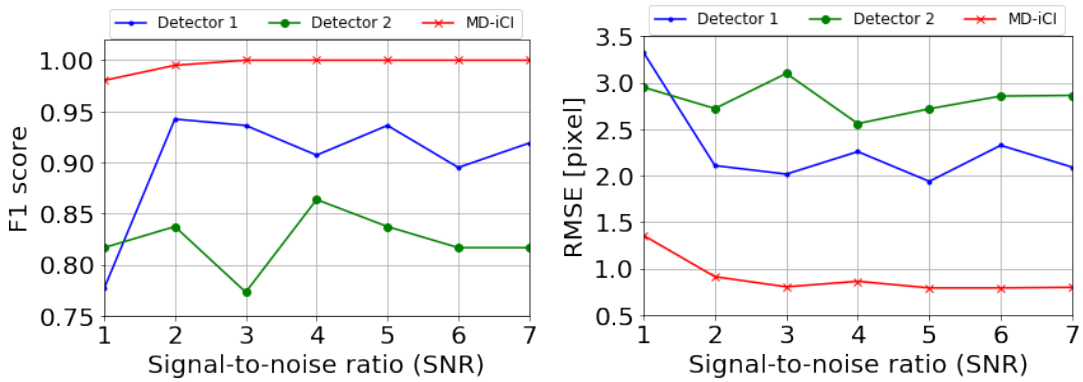
Figure 6.8: Ground truth and detection results for two SEFs ($\sigma_{\text{LoG}} = 3.0$ and $\sigma_{\text{LoG}} = 8.0$) and MD-iCI for a synthetic image with $r_{\text{offset, max}} = 16$ pixels. Only MD-iCI correctly detects all particles.

First, we studied images with an increasing $r_{\text{offset, max}}$ (maximum particle radii offset) from 0 to 24 pixels. We used SNR = 1.89. The detection performance is shown in Fig. 6.7 a and the localization performance is displayed in Fig. 6.7 b. It turns out, that MD-iCI performs best. MD-iCI yields for all images an F1 score above 98 % compared to two single-scale SEFs ($\sigma_{\text{LoG}} = 3.0$ and 8.0) which obtain F1 scores above 78 %. MD-iCI yields for all images an $RMSE$ below 1.1 pixels, whereas both SEFs obtain $RMSE$s up to 3.08 and 3.17 pixels. An example detection result is shown in Fig. 6.8 for $r_{\text{offset, max}} = 16$ pixels (yielding particle radii between 4 and 20 pixels). It can bee seen, that MD-iCI detects all particles with various sizes compared to both SEFs.

Second, we considered different SNR levels (SNR = 1 to 7) and used $r_{\text{offset, max}} = 12$ pixels as well as $r_{\text{offset, max}} = 16$ pixels. The result is shown in Fig. 6.9. For both $r_{\text{offset, max}}$, MD-iCI yields for all SNR levels an F1 score above 97 % and an $RMSE$ below 1.4 pixels. In contrast, the first SEF (with $\sigma_{\text{LoG}} = 3.0$) obtains an F1 score



(a) $r_{\text{offset, max}} = 12$ pixels



(b) $r_{\text{offset, max}} = 16$ pixels

Figure 6.9: Performance of MD-iCI (red curve) compared to two SEFs with $\sigma_{\text{LoG}} = 3.0$ (blue) and $\sigma_{\text{LoG}} = 8.0$ (green) for images with different SNR levels for (a) $r_{\text{offset, max}} = 12$ and (b) $r_{\text{offset, max}} = 16$ pixels.

above 90 % for $r_{\text{offset, max}} = 12$ pixels, and above 78 % for $r_{\text{offset, max}} = 16$ pixels. The second SEF ($\sigma_{\text{LoG}} = 8.0$) yields an F1 score above 78 % both $r_{\text{offset, max}}$. Further, the first SEF yields an $RMSE$ below 2.5 pixels for $r_{\text{offset, max}} = 12$ pixels and below 3.4 pixels for $r_{\text{offset, max}} = 16$, whereas the second SEF yields an $RMSE$ below 3.1 pixels for both $r_{\text{offset, max}}$. It also turns out that for a larger $r_{\text{offset, max}}$ (larger size heterogeneity) the performance of the SEFs is reduced while the performance of MD-iCI is not affected.

Particle Tracking Challenge Data

We assessed our multi-detector fusion approach MD-iCI and our tracking approach MD-BS on image data from the Particle Tracking Challenge [185]. We used time-lapse image sequences of the vesicle scenario with all object densities and all low SNR levels. We focus on low SNR levels, since it is known from previous studies that these SNR levels are challenging for particle detection and tracking [185]. The number of particles ranges from around 100 for low object density to around 1000

Table 6.10: Detection performance for all object densities for low SNR levels for the Particle Tracking Challenge data in terms of the F1 score [%].

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
SNR = 1					
Low	26.06	18.53	25.90	26.19	**26.82**
Medium	36.55	21.18	49.75	38.68	**50.42**
High	38.40	34.82	48.56	44.61	**55.79**
SNR = 2					
Low	79.96	72.50	**90.18**	80.08	82.49
Medium	80.82	78.33	90.61	85.74	**90.94**
High	79.27	81.69	84.76	86.40	**87.39**
Mean	56.84	51.18	64.96	60.28	**65.64**
Std. dev.	11.51	13.20	12.19	11.99	**11.37**

for high object density. Each of the six image sequences consists of 100 images with 512×512 pixels.

For particle detection and localization, we compared MD-iCI with two single-scale SEFs [106]. For SEF 1 we used a standard deviation $\sigma_{\text{LoG}} = 3$ pixels for low and medium object density as well as for high object density and SNR = 1, and we employed $\sigma_{\text{LoG}} = 2$ pixels for high object density and SNR = 2. For SEF 2 we used $\sigma_{\text{LoG}} = 4$ pixels for low object density, $\sigma_{\text{LoG}} = 2$ pixels for medium object density, and $\sigma_{\text{LoG}} = 4$ pixels for high object density and SNR = 1 and $\sigma_{\text{LoG}} = 2$ pixels for SNR = 2. These parameter choices yielded the best result. Besides the classical method SEF, we also compared MD-iCI with the multi-scale deep learning method DetNet [14]. DetNet is an adapted Deconvolution Network which naturally handles objects at multiple scales by the hourglass-shape of the network. For each image sequence, we used the last image for training DetNet and the first image for validation. We investigated two variants of MD-iCI, one which fuses the detections of two SEFs (MD-iCI 1) and one which fuses the detections of DetNet and two SEFs (MD-iCI 2).

The detection performance is quantified by the F1 score $\in [0, 1]$ computed as the mean F1 score over all images of a sequence. The localization performance is quantified by the mean $RMSE \in [0, 5]$ (cf. Section 6.2). The assignment between detected particles and ground truth was determined by the Munkres algorithm [122] with a maximal gating distance of 5 pixels. The detection performance for all object densities and all SNR levels is shown in Table 6.10, and the best performing method is marked in bold. MD-iCI 2 outperforms the other methods for SNR = 1 for all object densities, and for SNR = 2 for the challenging high and medium density. For SNR = 2 and low density, DetNet yields the best result. Further, MD-iCI 1 performs better for all SNR levels and all object densities than the single-scale SEFs. The localization performance is given in Table 6.11. MD-iCI 2 obtains the best localization results for all SNR levels and all object densities. Further, MD-iCI 1 yields better localization

Table 6.11: Localization performance for all object densities for low SNR levels for the Particle Tracking Challenge data in terms of the $RMSE$.

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
SNR = 1					
Low	4.38	4.76	4.67	4.37	**4.36**
Medium	4.43	4.70	4.24	4.36	**4.15**
High	4.42	4.52	4.24	4.30	**4.00**
SNR = 2					
Low	2.29	3.37	2.40	2.28	**2.16**
Medium	2.93	3.03	2.45	2.55	**2.23**
High	2.99	2.92	2.84	2.54	**2.47**
Mean	3.57	3.88	3.47	3.40	**3.23**
Std. dev.	0.42	**0.39**	0.46	0.46	0.47

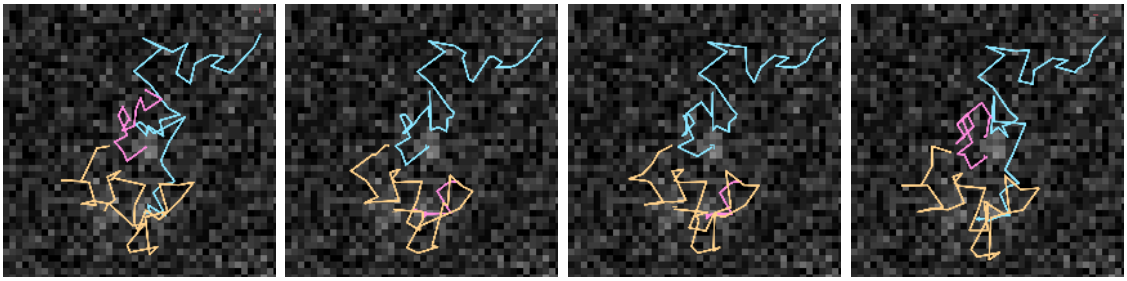
results compared to both single-scale SEFs for all SNR levels and all object densities, and performs better than DetNet for all SNR levels and low object density.

For particle tracking, we performed a comparison of MD-BS with the overall top-three performing methods (Method 5, 1, 2) of the Particle Tracking Challenge. Method 5 uses SEF for particle detection and PDAE for particle linking [48]. Method 1 employs iterative centroid calculation for particle localization and assigns particles by combinatorial optimization [102]. Method 2 detects particles by convolution with a disk shaped object and finds correspondences by multiple-hypothesis tracking [131]. We also compared MD-BS with the recent SMS-PDAE method [12]. SMS-PDAE uses SEF for particle detection and temporal multi-sensor data fusion with covariance intersection for tracking. For SMS-PDAE and Method 5 we used SEF 1, since it yielded a better detection and localization performance than SEF 2 (cf. Table 6.10 and 6.11). We investigated two variants of MD-BS, one which fuses the detections of two SEFs (SEF 1 and SEF 2) denoted as MD-BS 1 and one which fuses the detections of DetNet and two SEFs (SEF 1 and SEF 2) denoted as MD-BS 2.

The tracking performance is evaluated by the metrics α, β, JSC_θ, JSC, and $RMSE$ from the Particle Tracking Challenge [185]. α and β quantify association and localization errors. $\alpha \in [0, 1]$ provides the overall matching quality between ground truth and computed trajectories where a perfect matching is given by $\alpha = 1$ and no matching is reflected by $\alpha = 0$. $\beta \in [0, \alpha]$ considers additionally spurious trajectories compared to α. The similarity at the track level is evaluated by the Jaccard similarity coefficient $JSC_\theta \in [0, 1]$. The Jaccard similarity coefficient $JSC \in [0, 1]$ quantifies the detection performance, and the root mean-square error $RMSE$ evaluates the localization performance. Table 6.12 shows the mean values of all tracking performance metrics over all object densities and all low SNR levels. MD-BS 2 outperforms all methods for three out of five metrics, namely α, β, and JSC. Further, MD-BS 1 shows improved results for α, β, and JSC compared to

Table 6.12: Mean performance metrics over all object densities for low SNR levels for the Particle Tracking Challenge data.

Metric	α	β	JSC_θ	JSC	$RMSE$
Method 5	0.327	0.275	0.523	0.365	1.752
Method 1	0.176	0.130	0.278	0.154	**1.281**
Method 2	0.345	0.261	0.479	0.350	1.899
SMS-PDAE	0.338	0.288	**0.527**	0.376	1.674
MD-BS 1	0.351	0.293	0.526	0.391	1.695
MD-BS 2	**0.360**	**0.296**	0.523	**0.404**	1.833



(a) Ground truth (b) PDAE (c) SMS-PDAE (d) MD-BS

Figure 6.10: Ground truth and tracking results for different methods for the vesicle scenario from the Particle Tracking Challenge data with SNR = 2 and medium density (time point $t = 70$). Only MD-BS yields three complete trajectories.

SMS-PDAE and Method 5. In Fig. 6.10, three example trajectories for SNR = 2 and medium object density for different tracking approaches are shown. Only MD-BS (we used MD-BS 2) yields three complete trajectories. By fusing multiple detection results and exploiting information from both past and future time points, MD-BS is able to obtain three complete trajectories (orange, blue, and purple). The other approaches yield one complete trajectory (orange), but other trajectories are not complete (blue) or are missing (purple).

Evaluation on Time-Lapse Fluorescence Microscopy Images

We also evaluated our multi-detector fusion approach MD-iCI as well as our tracking approach MD-BS on challenging time-lapse microscopy data sets containing hepatitis C virus (HCV) associated proteins and chromatin structures. The images are acquired with confocal fluorescence microscopes using different spatial-temporal resolutions. In total, we used nine live cell microscopy image sequences and compared the results for both object detection and localization as well as for object tracking with those of previous methods.

Table 6.13: Detection results for HCV live cell image sequences, F1 score [%].

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
1	75.40	39.00	82.35	77.50	**83.38**
2	56.63	3.75	66.67	57.14	**69.83**
3	69.84	78.62	81.48	75.24	**84.08**
Mean	67.29	40.46	76.83	69.96	**79.10**
Std. dev.	6.82	26.49	6.23	7.89	**5.68**

HCV Live Cell Confocal Microscopy Images

We applied our MD-iCI to three live cell microscopy image sequences displaying HCV-infected cells with mCherry-labeled host cell protein ApoE, mCherry-labeled viral protein NS5A, or mTurquoise2-labeled ApoE, respectively [18]. The first two image sequences consist of 34 images with 225×194 and 177×249 pixels. The spatial resolution is 220 nm/pixel and the temporal resolution is 37 s/frame and 1.5 s/frame, respectively. The third image sequence consists of 91 images with 726×396 pixels. The spatial resolution is 90 nm/pixel and the temporal resolution is 2 s/frame. The image data was acquired with a PerkinElmer UltraVIEW ERS or VoX spinning disk confocal microscope mounted on a Nikon TE2000-E or TiE. The data is challenging due to heterogeneous object size, clutter, high object density, and complex motion.

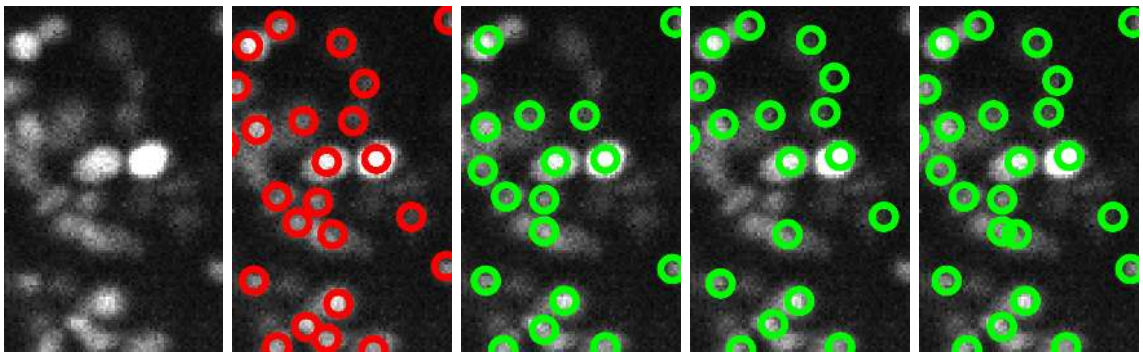
For particle detection and localization, we performed a comparison of MD-iCI with SEF [106] and DetNet [14]. For SEF we used two variants with different values for σ_{LoG}. For SEF 1 we used $\sigma_{\text{LoG}} = 2$ pixels for all videos, and for SEF 2 we used $\sigma_{\text{LoG}} = 2.6$ pixels for the first two videos and $\sigma_{\text{LoG}} = 3$ pixels for the third video. These parameter choices yielded the best result. We trained DetNet on the last image of a video and used the first image for validation. Further, we investigated two variants of MD-iCI, one which fuses detections of two SEFs (MD-iCI 1) and one which fuses the detections of DetNet and two SEFs (MD-iCI 2).

For the detection performance, we computed the F1 score $\in [0, 1]$ for time point five for each image sequence. The localization performance is evaluated by the mean $RMSE \in [0, 5]$ between detected particles and ground truth (cf. Section 6.2). The assignment between detected particles and ground truth was determined by the Munkres algorithm [122] with a maximal gating distance of 5 pixels. Each image sequence contains between 157 and 194 ground truth detections manually annotated using the Manual Tracking plug-in in ImageJ [202]. The detection performance is shown in Table 6.13 and the localization results are given in Table 6.14. For both detection and localization, MD-iCI 2 yields the best results for all three image sequences. Further, MD-iCI 1 outperforms both single-scale SEFs. Example detection results for MD-iCI 2 are displayed in Fig. 6.11 which show that our approach yields better results than SEF and DetNet.

For tracking of particles, we performed a comparison of MD-BS with a Kalman

Table 6.14: Localization results for HCV live cell image sequences, $RMSE$.

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
1	3.21	4.37	2.93	3.10	**2.71**
2	3.63	4.95	3.55	3.62	**2.64**
3	3.23	3.25	3.36	3.08	**2.90**
Mean	3.36	4.19	3.28	3.27	**2.75**
Std. dev.	0.17	0.61	0.22	0.22	**0.10**



(a) Orig. image (b) Ground truth (c) SEF 2 (d) DetNet (e) MD-iCI

Figure 6.11: Original image and detection results for HCV live cell microscopy data of fluorescently labeled ApoE proteins (time point $t = 5$).

filter tracking approach (KF) [200], the ParticleTracker (PT) [102], a multiple-hypothesis tracking approach (MHT) [201, 185], and the Smoothing Multi-Sensor PDAE (SMS-PDAE) [12]. KF uses an LoG filter for particle detection and linear assignment for correspondence finding. PT employs intensity-weighted centroid calculation for particle localization and assigns particles by combinatorial optimization. MHT localizes particles by a wavelet-based detection scheme and uses a Kalman filter with multiple motion models.SMS-PDAE employs a single-scale SEF, and we used SEF 1 since it yielded better results than SEF 2 (cf. Tables 6.13 and 6.14). We also investigated two variants of MD-BS, one which fuses the detections of two SEFs (SEF 1 and SEF 2) denoted as MD-BS 1 and one which fuses the detections of DetNet and two SEFs (SEF 1 and SEF 2) denoted as MD-BS 2.

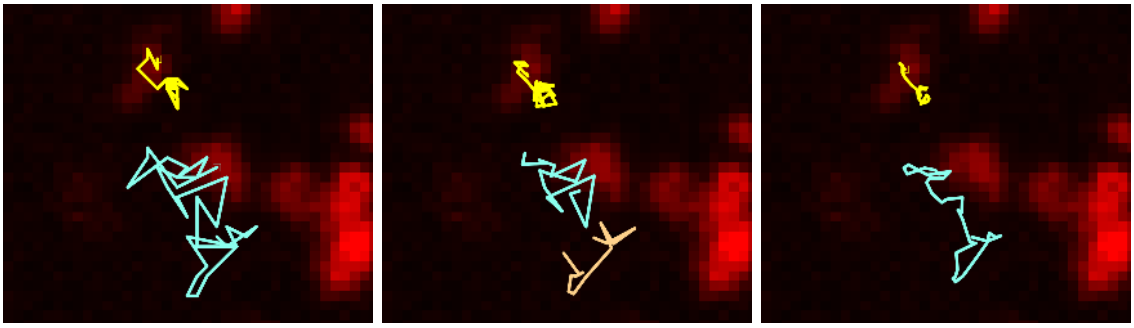
The tracking accuracy is evaluated by the measure $P_{\text{track}} \in [0, 1]$:

$$P_{\text{track}} = \frac{n_{\text{track,correct}}}{n_{\text{track,total}}} \tag{6.2}$$

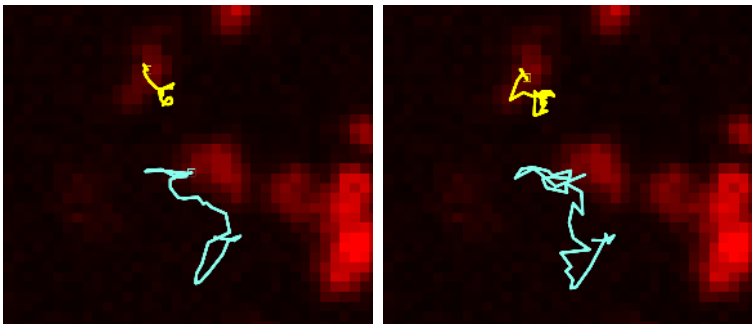
$n_{\text{track,total}}$ is the number of ground truth trajectories and $n_{\text{track,correct}}$ the number of correctly computed trajectories. $n_{\text{track,correct}}$ is determined as Gaussian weighted sum of the percentage of correctly tracked time steps, and we used a maximal gating distance of 5 pixels [48]. Ground truth was obtained by manual annotation using the Manual Tracking plug-in in ImageJ (29, 32, and 108 trajectories for the different

Table 6.15: Tracking accuracy for HCV live cell image sequences, P_{track} [%].

Seq.	1	2	3	Mean	Std. dev.
KF	41.14	40.03	48.93	43.57	3.43
PT	30.88	33.35	29.22	31.15	**1.47**
MHT	61.46	72.32	41.51	58.43	11.05
SMS-PDAE	69.35	74.67	51.68	65.23	8.51
MD-BS 1	70.00	75.10	52.74	65.95	8.29
MD-BS 2	**70.25**	**77.34**	**54.42**	**67.34**	8.30



(a) Ground truth (b) MHT (c) SMS-PDAE



(d) MD-BS 1 (e) MD-BS 2

Figure 6.12: Ground truth and tracking results for different methods for HCV live cell confocal microscopy images of fluorescently labeled ApoE proteins (time point $t = 34$).

videos). The results are shown in Table 6.15. We also determined the mean P_{track} value as a measure of the overall performance. It turned out, that MD-BS 2 yields the best result for all image sequences and achieves a mean P_{track} of 67.34 %. MD-BS 1 is somewhat worse and obtains better results than SMS-PDAE for all image sequences. In Fig. 6.12, example results are shown. MD-BS 2 yields more complete and correct trajectories than SMS-PDAE and MD-BS 1.MHT obtains a broken trajectory.

Table 6.16: Detection results for chromatin live cell image sequences, F1 score [%].

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
1	82.18	83.08	85.71	82.35	**86.49**
2	82.19	75.76	88.31	86.84	**89.74**
3	87.80	64.00	89.60	89.60	**91.34**
4	72.09	70.11	76.36	78.35	**79.30**
5	63.67	31.63	79.60	64.78	**80.67**
6	66.67	44.44	74.07	71.58	**74.47**
Mean	75.77	61.50	82.28	78.92	**83.67**
Std. dev.	9.69	8.79	**6.50**	9.41	6.58

Chromatin Live Cell Confocal Microscopy Images

We also evaluated MD-iCI and MD-BS using six challenging live cell microscopy image sequences with fluorescently labeled chromatin structures of HeLa Kyoto cells. The image sequences consist of 63 images with 512×512 pixels. The spatial resolution is 48 nm/pixel for four videos, and 31 nm/pixel and 49 nm/pixel for the remaining videos. The temporal resolution is 0.985 s/frame for all videos. The images were acquired with a PerkinElmer UltraVIEW VoX confocal microscope. The data is challenging due to high heterogeneity in object size, high object density, and low SNR.

For chromatin detection and localization, we performed a comparison of MD-iCI with SEF and DetNet. For SEF we used two different variants, SEF 1 with $\sigma_{\text{LoG}} = 2.7$ pixels for all videos, and SEF 2 with $\sigma_{\text{LoG}} = 3.5$ pixels for two videos (videos 3 and 6) and $\sigma_{\text{LoG}} = 3$ pixels for the remaining videos. For DetNet, for each video we randomly splitted the remaining five videos into three videos for testing and two for validation. DetNet was trained on the last image of the training videos and evaluated on the first image of the validation videos. We assessed two variants of MD-iCI as for the HCV data above. The results are given in Table 6.16. It turned out that MD-iCI 2 outperforms the other methods for all image sequences and obtains a mean F1 score of 83.67 %. MD-iCI 1 performs better than both singe-scale SEFs and obtains a mean F1 score of 78.92 % compared to SEF 1 with 75.77 % and SEF 2 with 61.50 %. DetNet and MD-iCI 2 yield the lowest standard deviation for the F1 score (highest robustness). The localization performance is given in Table 6.17. MD-iCI 2 outperforms all methods for five out of six videos. MD-iCI 1 obtains the best result for one video and better results than both single-scale SEFs for all videos.

For chromatin tracking, we compared MD-BS with the same methods as for the HCV data above. We manually annotated for each image sequence between 20 and 22 ground truth trajectories. The tracking results for all approaches are shown in Table 6.18. MD-BS 2 outperforms all other methods for all image sequences and achieves a mean P_{track} of 45.63 %. MD-BS 1 obtains for all image sequences a better

Table 6.17: Localization results for chromatin live cell image sequences, $RMSE$.

Seq.	SEF 1	SEF 2	DetNet	MD-iCI 1	MD-iCI 2
1	2.58	2.70	2.29	2.55	**1.89**
2	2.94	3.31	2.36	2.60	**2.19**
3	2.67	3.71	2.39	2.46	**2.22**
4	3.40	3.47	2.95	2.96	**2.69**
5	3.77	4.55	3.09	3.73	**3.01**
6	3.33	4.04	3.26	**2.64**	2.72
Mean	3.12	3.63	2.72	2.82	**2.45**
Std. dev.	1.04	**0.28**	0.95	1.06	0.94

Table 6.18: Tracking accuracy for chromatin live cell image sequences, P_{track}[%].

Seq.	1	2	3	4	5	6	Mean	Std. dev.
KF	22.01	23.47	35.39	11.23	23.74	24.50	23.39	**7.68**
PT	39.68	24.76	31.09	6.37	36.99	20.00	26.48	12.29
MHT	38.06	26.87	48.87	24.30	40.99	42.64	36.96	9.53
SMS-PDAE	47.85	30.16	49.60	13.29	44.67	30.31	35.98	14.02
MD-BS 1	48.71	36.98	49.85	22.37	46.85	38.12	40.48	10.41
MD-BS 2	**53.94**	**47.21**	**55.64**	**24.44**	**49.72**	**42.80**	**45.63**	11.36

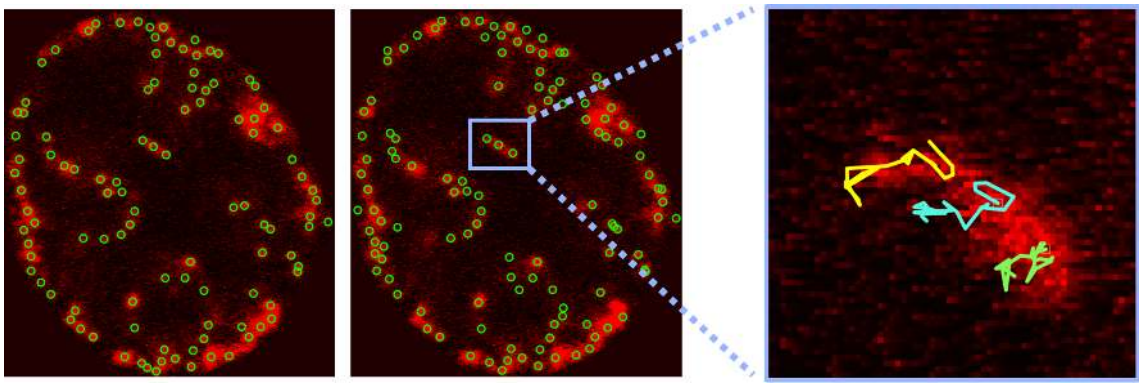


Figure 6.13: Results for chromatin live cell microscopy data. (Left) Detection results of MD-iCI 2 (time point $t = 5$). (Right) Tracking results of MD-BS 2 for a region-of-interest (time point $t = 25$).

result than SMS-PDAE. Example detection results of MD-iCI 2 and tracking results of MD-BS 2 are shown in Fig. 6.13. It can bee seen, that the detections are close to the ground truth for the challenging data with heterogeneous object size and that complete trajectories are obtained.

Summary We quantitatively evaluated MD-BS using data from the Particle Tracking Challenge and obtained state-of-the-art results or outperformed previous methods. We also demonstrated that MD-iCI improves particle detection and localization compared to existing detection methods. In addition, we benchmarked MD-BS and MD-iCI on nine challenging live cell fluorescence microscopy image sequences acquired with microscopes using different spatial-temporal resolutions, and different types of subcellular structures. We found, that our approach outperforms previous methods for microscopy data of HCV associated proteins and chromatin structures including high object density, low SNR, heterogeneous object size, and complex motion.

6.3 Deep Learning for Particle Detection and Tracking

This section presents experimental results of the particle tracking methods proposed in Chapter 5. First, DetNet-PDAE (Section 5.1) is evaluated which combines deep learning-based particle detection with Bayesian sequential estimation for tracking. Second, DetNet-DPHT (Section 5.2) is evaluated which exploits deep learning for both particle detection and tracking.

6.3.1 Deep Learning-Based Detection and Bayesian Particle Tracking

We provide insight on the interaction of particle detection and association by comparing the optimized and non-optimized DetNet-PDAE method. We evaluated DetNet-PDAE (Section 5.1) using data of the Particle Tracking Challenge [185] with different SNR levels. Further, we assessed the method on live cell microscopy images of hepatitis C virus (HCV) proteins. It turned out that our method improves the results compared to existing methods.

6.3.1.1 Particle Tracking Challenge Data

We evaluated DetNet-PDAE using the microtubule data from the Particle Tracking Challenge [185]. We compared DetNet-PDAE with the overall top-three tracking methods (Methods 5, 1, 2) and a spot-enhancing filter (SEF)-based tracking approach. Method 5 consists of applying Gaussian fitting (GF) for particle detection and PDAE for tracking [48]. Method 1 uses iterative centroid calculation for particle detection and performs combinatorial optimization for tracking [102]. Method 2 detects particles using a disk shaped object and employes MHT for tracking [131]. The SEF-based tracking approach (SEF-PDAE) applies a LoG filter followed by thresholding for particle detection and uses PDAE for tracking. To show the influence of the data association hyperperparameters σ_n, $\mathbf{Q}$, and $\mathbf{R}$, for our DetNet-PDAE method, we computed the results without and with hyperparameter optimization with Covariance Matrix Adaption Evolution Strategy (CMA-ES) in HyperHyper [197] (DetNet-PDAE vs. DetNet-PDAE (opt)). DetNet uses optimized detection hyperparameters from training as in [14]. For Method 5, SEF-PDAE, and DetNet-PDAE, a directed motion model is used for $\mathbf{F}$.

First, we evaluated the detection performance of DetNet, GF, and SEF. We computed the $F1\,\text{score} \in [0, 1]$ which measures the similarity between detections and ground truth. The assignment between particle detections and ground truth is determined by the Munkres algorithm [122] with a maximal gating distance of 5 pixels. Second, the tracking performance was evaluated using all SNR levels of the microtubule scenario and low density. The image sequences comprise 100 images with 512×512 pixels each. We computed for each method the performance metrics α, β, JSC_θ, JSC, and $RMSE$. The overall matching quality of ground

Table 6.19: Detection performance ($F1$ score) for the first frame of the microtubule image sequences for all SNR levels.

	GF	SEF	DetNet
SNR 1	0.048	0.143	**0.437**
SNR 2	0.865	0.817	**0.889**
SNR 4	0.943	0.892	**0.955**
SNR 7	**1.000**	0.987	**1.000**

truth with computed trajectories is represented by $\alpha \in [0,1]$, whereas $\beta \in [0,\alpha]$ quantifies additionally spurious tracks. The similarity of entire tracks is represented by the Jaccard similarity coefficient $JSC_\theta \in [0,1]$ and the detection performance by $JSC \in [0,1]$. The localization is evaluated by $RMSE$.

In Table 6.19, the detection performance for the first frame of the image sequences for all SNR levels is shown. The best performing method is highlighted in bold. It can be seen that DetNet outperforms GF and SEF. Interestingly, the tracking results in Table 6.20 reveal that directly using the DetNet detections for DetNet-PDAE does hardly improve the results compared to the previous top-performing Method 5 (e.g., for SNR=2 and 4). However, using our method with optimized hyperparameters (DetNet-PDAE (opt)) outperforms the non-optimized DetNet-PDAE for almost all performance measures and all SNR levels, and yields better results than Method 5. For SNR=1, 4, and 7, DetNet-PDAE (opt) is the overall best method. For SNR=2, DetNet-PDAE (opt) yields the best results for α and JSC_θ compared to all methods.

6.3.1.2 Evaluation on HCV Live Cell Microscopy Data

We also evaluated our method using live cell microscopy image sequence displaying fluorescently labeled HCV NS5A proteins. The data consists of 34 live cell images with 512×512 pixels, for which 108 ground truth trajectories were determined manually. The data was acquired with an Ultra-View ERS spinning disk confocal microscopy mounted on a Nikon TE2000-E. DetNet was trained on a different single image from a different dataset using a section of 147×280 pixels corresponding to one cell with only 66 ground truth detections.

We compared DetNet-PDAE with SEF-PDAE, the ParticleTracker (PT) [102], and a Kalman filter approach (KF) [200]. For DetNet-PDAE and SEF-PDAE, a random walk motion model was used for $\mathbf{F}$. We evaluated the detection performance for all methods using the $F1$ score $\in [0,1]$ for the first frame of the HCV live cell data. For the tracking performance, we employed the measure $P_{track} \in [0,1]$, which quantifies the percentage of correctly computed trajectories relative to the number of true trajectories [48].

In Table 6.21, the detection results for the first frame of the HCV live cell data are shown. DetNet achieves the best $F1$ score followed by SEF. In addition, DetNet yields a higher number of true positives than SEF (56 vs. 52). Example detection

Table 6.20: Tracking performance of different methods for the microtubule scenario for all SNR levels.

Metric	α	β	JSC_θ	JSC	$RMSE$
SNR=1					
Method 5	0.058	0.057	0.194	0.092	**2.188**
Method 1	0.043	0.034	0.168	0.056	2.253
Method 2	0.117	0.061	0.213	0.122	2.717
SEF-PDAE	0.079	0.038	0.142	0.069	2.445
DetNet-PDAE	**0.165**	0.071	0.187	0.144	2.724
DetNet-PDAE (opt)	0.137	**0.082**	**0.248**	**0.171**	2.867
SNR=2					
Method 5	0.518	**0.485**	0.732	**0.652**	**1.570**
Method 1	0.319	0.250	0.488	0.397	2.100
Method 2	0.418	0.300	0.503	0.479	2.226
SEF-PDAE	0.391	0.356	0.646	0.530	2.050
DetNet-PDAE	0.479	0.410	0.708	0.562	1.842
DetNet-PDAE (opt)	**0.519**	0.470	**0.742**	0.630	1.754
SNR=4					
Method 5	0.750	0.728	**0.917**	0.874	1.086
Method 1	0.541	0.495	0.874	0.792	1.951
Method 2	0.562	0.259	0.356	0.369	1.911
SEF-PDAE	0.570	0.511	0.790	0.777	1.928
DetNet-PDAE	0.754	0.732	0.903	0.889	1.148
DetNet-PDAE (opt)	**0.807**	**0.782**	0.916	**0.896**	**0.836**
SNR=7					
Method 5	0.803	0.787	0.939	0.894	0.844
Method 1	0.657	0.621	0.902	0.837	1.375
Method 2	0.694	0.686	**0.959**	**0.954**	1.520
SEF-PDAE	0.513	0.492	0.924	0.875	2.256
DetNet-PDAE	0.810	0.784	0.910	0.910	0.908
DetNet-PDAE (opt)	**0.850**	**0.830**	0.930	0.930	**0.705**

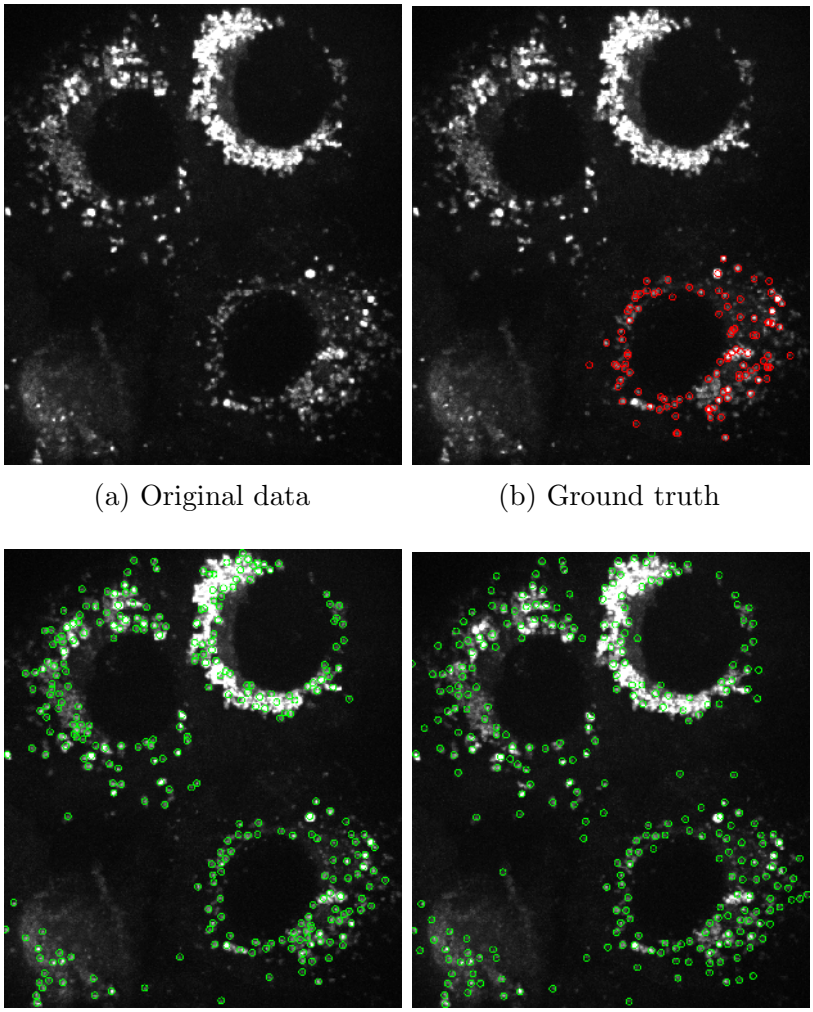
Table 6.21: Detection performance ($F1$ score) for the HCV live cell data.

PT	KF	SEF	DetNet
0.507	0.549	0.635	**0.642**

Table 6.22: Tracking performance P_{track} for the HCV live cell data.

PT	KF	SEF-PDAE	DetNet-PDAE	DetNet-PDAE (opt)
0.309	0.411	0.623	0.579	**0.646**

results for SEF and DetNet are displayed in Fig. 6.14 c, d. In Table 6.22, the tracking results for the HCV live cell data are presented. Analogously to the data from the



(a) Original data (b) Ground truth

(c) SEF (d) DetNet

Figure 6.14: Detection results for the first frame of the HCV live cell microscopy data. a) Original data. b) Ground truth detections indicated by red circles. c) SEF. d) DetNet.

Particle Tracking Challenge, it can be seen that the non-optimized DetNet-PDAE method yields a worse result compared to the previous SEF-PDAE method, while the optimized DetNet-PDAE (opt) outperforms all methods.

Summary DetNet-PDAE was evaluated using data of the Particle Tracking Challenge as well as live cell microscopy images of HCV viral proteins. Data association parameters that depend on the detection result were automatically determined by hyperparameter optimization. It turned out, that the optimized DetNet-PDAE yields better tracking results and generally outperforms previous methods.

6.3.2 Deep Learning for Particle Detection and Tracking

We evaluated the proposed approach denoted as DetNet-DPHT (Section 5.2) using data of the Particle Tracking Challenge [185] with different SNR levels. In addition, we determined the performance for live cell fluorescence microscopy images of hepatitis C virus (HCV) proteins. It turned out that our approach yields state-of-the-art results or improves the results compared to previous methods. We also found that using deep learning for particle detection instead of classical methods improves the result of deep learning-based tracking.

6.3.2.1 Particle Tracking Challenge Data

We evaluated the proposed DetNet-DPHT approach using data from the Particle Tracking Challenge [185], and compared it with the overall top-three methods (Method 5, 1, 2) of the challenge. Method 5 localized particles by determining local intensity maxima followed by Gaussian fitting (GF) and uses probabilistic data association with elliptical sampling (PDAE) for tracking [48]. Method 1 performs iterative centroid calculation for particle detection and uses combinatorial optimization for tracking [102]. Method 2 detects particles by convolution with a disk shaped object model and performs MHT for tracking [131]. In addition, we compared our approach with the previous methods SEF-PDAE [48], DetNet-PDAE [15], and SEF-GF-DPHT [145]. SEF-PDAE employs the spot-enhancing filter (SEF) [106] for particle detection and PDAE for tracking. DetNet-PDAE uses DetNet for particle detection and PDAE for tracking. SEF-GF-DPHT [145] uses SEF and GF for particle detection and DPHT for tracking.

We evaluated the tracking performance using data of the microtubule scenario with all SNR levels and low density. Each of the image sequences comprises 100 images with 512×512 pixels. The tracking performance is quantified by the metrics α, β, JSC_θ, JSC, and $RMSE$. $\alpha \in [0, 1]$ represents the overall matching quality of ground truth with computed trajectories, whereas $\beta \in [0, \alpha]$ takes additionally into account spurious tracks. The similarity of entire tracks is represented by the Jaccard similarity coefficient $JSC_\theta \in [0, 1]$. $JSC \in [0, 1]$ represents the overall detection performance, whereas the localization performance is evaluated by $RMSE$.

In Table 6.23, the tracking results are shown and the best results are highlighted in bold. It turns out, that DetNet-DPHT yields the best results for three out of five metrices for SNR = 1, 4, and 7. Our approach outperforms all methods for JSC_θ for low SNR levels (SNR = 1, 2). For high SNR levels (SNR = 4, 7), our approach outperforms all methods for α, β, and $RMSE$. Further, it performs better than SEF-GF-DPHT for α and β for all SNR levels. Fig. 6.15 shows example tracking results for DetNet-DPHT and SEF-GF-DPHT for SNR = 2. It can be seen that deep learning for particle detection improves the tracking result compared to a classical detection method (see the red trajectory in the middle). Fig. 6.16 compares DetNet-DPHT with DetNet-PDAE. We used SNR = 7 to focus on the task of correspondence finding.

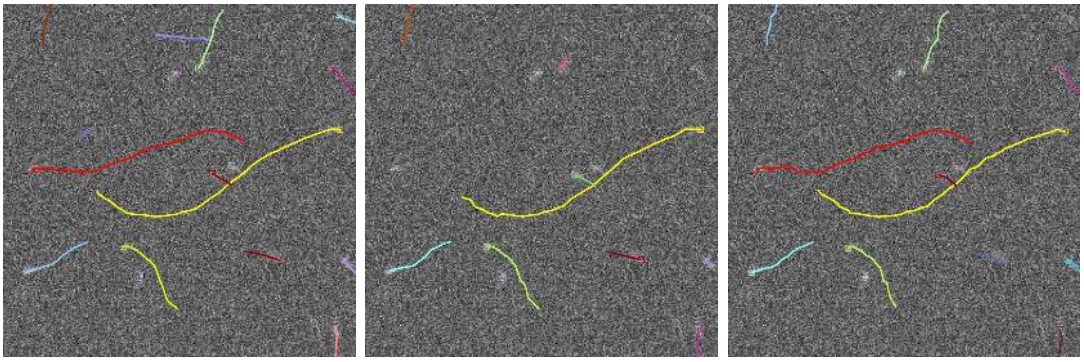
Table 6.23: Tracking performance of different methods for data of the microtubule scenario for all SNR levels.

Metric	α	β	JSC_θ	JSC	$RMSE$
SNR=1					
Method 5	0.058	0.057	0.194	0.092	2.188
Method 1	0.043	0.034	0.168	0.056	2.253
Method 2	0.117	0.061	0.213	0.122	2.717
SEF-PDAE	0.079	0.038	0.142	0.069	2.445
DetNet-PDAE	**0.137**	0.082	0.248	0.171	2.867
SEF-GF-DPHT	0.101	0.080	0.290	0.121	**2.143**
DetNet-DPHT	0.135	**0.101**	**0.342**	**0.202**	2.603
SNR=2					
Method 5	0.518	**0.485**	0.732	**0.652**	1.570
Method 1	0.319	0.250	0.488	0.397	2.100
Method 2	0.418	0.300	0.503	0.479	2.226
SEF-PDAE	0.391	0.356	0.646	0.530	2.050
DetNet-PDAE	**0.519**	0.470	0.742	0.630	1.754
SEF-GF-DPHT	0.477	0.443	0.737	0.569	1.602
DetNet-DPHT	0.506	0.476	**0.747**	0.606	**1.477**
SNR=4					
Method 5	0.750	0.728	0.917	0.874	1.086
Method 1	0.541	0.495	0.874	0.792	1.951
Method 2	0.562	0.259	0.356	0.369	1.911
SEF-PDAE	0.570	0.511	0.790	0.777	1.928
DetNet-PDAE	0.807	0.782	0.916	**0.896**	0.836
SEF-GF-DPHT	0.803	0.776	**0.928**	0.890	0.888
DetNet-DPHT	**0.811**	**0.788**	0.915	0.884	**0.666**
SNR=7					
Method 5	0.803	0.787	0.939	0.894	0.844
Method 1	0.657	0.621	0.902	0.837	1.375
Method 2	0.694	0.686	0.959	**0.954**	1.520
SEF-PDAE	0.513	0.492	0.924	0.875	2.256
DetNet-PDAE	0.850	0.830	0.930	0.930	0.705
SEF-GF-DPHT	0.861	0.848	**0.970**	0.936	0.671
DetNet-DPHT	**0.870**	**0.852**	0.945	0.936	**0.556**

It can be seen that our approach yields a better result.

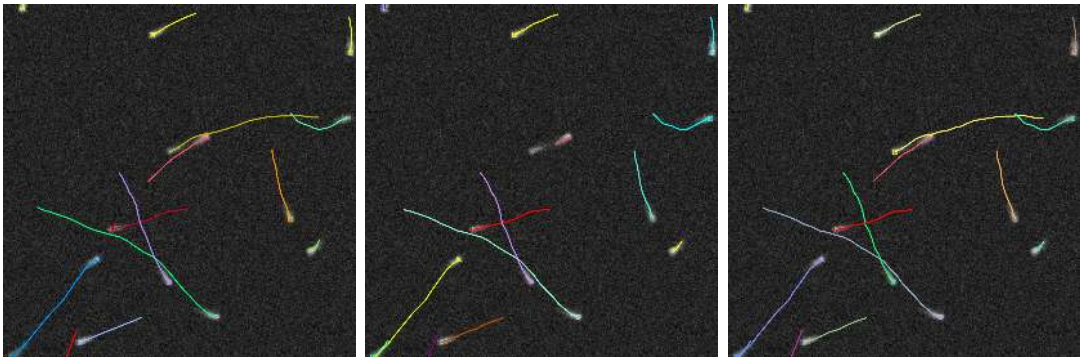
6.3.2.2 HCV Live Cell Microscopy Data

We also evaluated DetNet-DPHT using live cell microscopy data showing fluorescently labeled hepatitis C virus (HCV) proteins NS5A within Huh7/LunetCD81H cells [18]. The image sequence comprises 34 images with 177×249 pixels and we manually determined 28 ground truth trajectories. The considered data was acquired with a



(a) Ground truth (b) SEF-GF-DPHT (c) DetNet-DPHT

Figure 6.15: Ground truth and tracking results for data of the microtubule scenario with SNR = 2 (t = 65).



(a) Ground truth (b) DetNet-PDAE (c) DetNet-DPHT

Figure 6.16: Ground truth and tracking results for data of the microtubule scenario with SNR = 7 (t = 100).

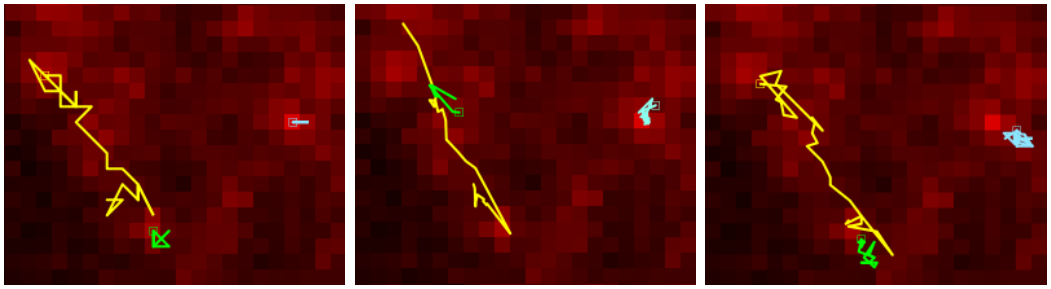
Table 6.24: Tracking performance P_{track} for the HCV live cell data.

PT	KF	SEF-PDAE	DetNet-PDAE	SEF-GF-DPHT	DetNet-DPHT
0.334	0.400	0.605	0.660	0.567	**0.662**

spinning disk confocal microscope. DetNet was trained using only one frame of the image sequence.

We compared our approach with the ParticleTracker (PT) [102], a Kalman filter approach (KF) [200], SEF-PDAE [48], DetNet-PDAE [15], and SEF-GF-DPHT [145]. The tracking performance is evaluated with the measure $P_{track} \in [0, 1]$, which quantifies the percentage of correctly computed trajectories relative to the number of true trajectories [48].

In Table 6.24, the tracking results are shown. It can be seen that DetNet-DPHT performs best. In addition, particle detection using deep learning (DetNet) improves the tracking result compared to a classical detection method (SEF-GF). Fig. 6.17 shows example tracking results for DetNet-DPHT and SEF-GF-DPHT compared to



(a) Ground truth (b) SEF-GF-DPHT (c) DetNet-DPHT

Figure 6.17: Ground truth and tracking results for the HCV live cell data (t = 34).

the ground truth. Our approach yields a better result (unbroken yellow trajectory, green trajectory at the bottom).

Summary We evaluated our DetNet-DPHT approach using data of the Particle Tracking Challenge and live cell microscopy data of HCV proteins. We found that our approach yields state-of-the-art or improved results compared to previous methods.

7 Hyperparameter Optimization and Cell Tracking

In this chapter, methods for hyperparameter optimization and cell tracking are presented. First, a hyperparameter optimization framework for HCV particle detection is described and evaluated. Second, a novel method for double-labeled HIV-1 particle detection in multi-channel fluorescence microscopy images based on hyperparameter optimization is proposed. Third, a novel MHT-based cell tracking approach with integrated cell division detection is proposed. Cell tracking in microscopy image data is crucial to gain insights into cellular processes to further determine virus-host interaction. The work has been published in Ritter *et al.* [197, 203] and Schacherer/Ritter *et al.* [204].

7.1 Hyperparameter Optimization for Particle Detection

Automatic analysis of microscopy data typically requires complex pipelines comprising multiple methods to solve different image analysis tasks (e.g., particle tracking [185], cell tracking [205], track analysis [206, 207, 208]). However, most methods suffer from determining application dependent hyperparameters to obtain the best performance. For complex analysis pipelines, manual optimization of hyperparameters is generally very time-consuming and difficult for a high-dimensional hyperparameter space. Thus, automated optimization is required. However, computation of the gradient of the loss function is analytically or computationally infeasible, which prevents the use of first or higher order optimization methods. This limitation can be overcome by using zero-order optimization also known as black-box optimization [209], which does not require gradient information of the loss function. Black-box optimization uses only a limited number of evaluations (hyperparameter configurations) to determine a (local) optimum of the generally non-convex optimization problem.

We have developed a framework for black-box hyperparameter optimization for biomedical image analysis pipelines called HyperHyper. This framework has several advantages compared to existing hyperparameter optimization frameworks such as Google Vizier [210], Sherpa [211], Auto-WEKA [212], Spearmint [213], and HyperOpt [214]. Existing frameworks lack certain features (e.g., modular optimizer, job wrapper, and integrated scheduler), which are essential to optimize complex

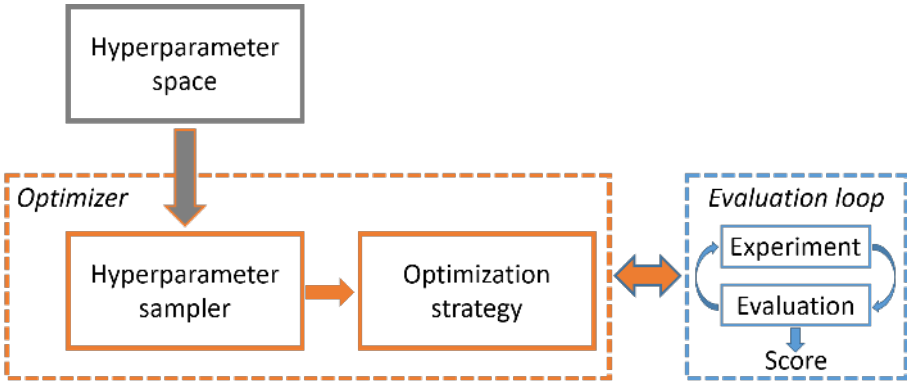


Figure 7.1: Schematic representation of HyperHyper software architecture.

image analysis pipelines using different computing paradigms and environments. To determine optimal solutions, our HyperHyper framework employs more than 40 different optimization methods, while existing frameworks include significantly less methods (e.g., up to five methods as in Sherpa, Spearmint, and HyperOpt). The high number of optimizers in HyperHyper was realised by separation of hyperparameter sampling and optimization strategy. To optimize hyperparameters on different cluster computing infrastructures, we implemented an integrated scheduler which is advantageous when deploying image analysis methods on heterogeneous computing infrastructures.

The HyperHyper framework subdivides hyperparameter optimization in a hyperparameter space definition, a general optimizer containing a hyperparameter candidate sampler and optimization strategy, and an evaluation loop (Figure 7.1). The candidate sampler and optimization strategy can be selected from a model zoo to design an optimizer for a specific application. In addition, the hyperparameter space definition incorporates prior distributions, bounds, and the sampling resolution. The candidate sampler and optimization strategy can exploit the structure of the hyperparameter space to improve convergence of the optimization. To find the global optimum, Grid Search can be used. Moreover, by design the execution of the evaluation loop can be performed highly distributed and is programming language agnostic. We integrated modules for monitoring and visualization to analyse the optimization problem. These visualizations including an infimum projection can reveal insights into, for example, the performance of the optimization process and the dependencies of the hyperparameters.

Optimizer

To perform optimization, constraints on the hyperparameter space have to be specified. This includes the bounds and hierarchy of each parameter, the sampling resolution, and additional prior distributions (e.g., discrete or continuous uniform, Gaussian, log Gaussian, exponential distributions). In our experiments we used pipelines that

Table 7.1: Results for the HCV protein detection pipeline with different optimizers. The table shows the improvement ΔF1 (mean ± std.) after the warm-up phase and the absolute F1 score (mean ± std.). The best results are highlighted in bold.

Optimizer	ΔF1 (Improvement)	F1
Random	0.043±0.033	0.871±0.002
TPE	0.041±0.023	0.867±0.000
CMA-ES	0.022±**0.008**	0.871±0.001
SMAC-RF	**0.050**±0.037	**0.872±0.000**
SMAC-XGBoost	0.041±0.037	**0.872±0.000**
Grid Search	-	*0.872*

involve non-ordinal parameters. Therefore, we decided to choose optimizers which can handle variables without a natural order.

The most naive optimization strategy is to perform Random Search (Random) by random sampling from the prior distributions. In Sequential Model-based Optimization (SMBO) like SMAC [215], a surrogate model is fitted to the best performing hyperparameters. We investigated SMAC with the original random forest (SMAC-RF), and with XGBoost [216] as surrogate model. We decided to use XGBoost, since it is currently one of the most popular decision tree based models. Moreover, we investigated the Tree of Parzen Estimator (TPE), which performs a nonparametric density approximation of the best performing hyperparameter configurations [217]. Finally, we use Covariance Matrix Adaptation Evolution Strategy (CMA-ES), which is a generic population-based meta-heuristic based optimizer [218]. In CMA-ES feature sets are assumed as "genomes", which undergo evolutionary processes like selection, recombination, or mutation to increase the probability for sampling promising hyperparameter configurations.

Detection of HCV Proteins

We evaluated HyperHyper for live cell fluorescence microscopy data displaying fluorescently labeled HCV NS5A as small round particles. The live cell data was acquired by an Ultra-View ERS spinning disk confocal microscope with an image size of 355 × 447 pixels. To detect HCV proteins, we used the spot-enhancing filter (SEF) [106] which consists of applying a Laplacian-of-Gaussian filter (LoG) with standard deviation σ_{LoG}, followed by thresholding the filtered image. The threshold is based on the mean intensity of the filtered image plus a factor c times the standard deviation of the filtered image intensities [48, 11]. To detect HCV proteins, the hyperparameters σ_{LoG} and c have to be optimized.

We used Random, TPE, CMA-ES, SMAC-RF, and SMAC-XGBoost for hyperparameter optimization and performed 10 runs per optimizer with 3.500 evaluations distributed on 20 compute nodes (total: 175.000 evaluations). To determine the

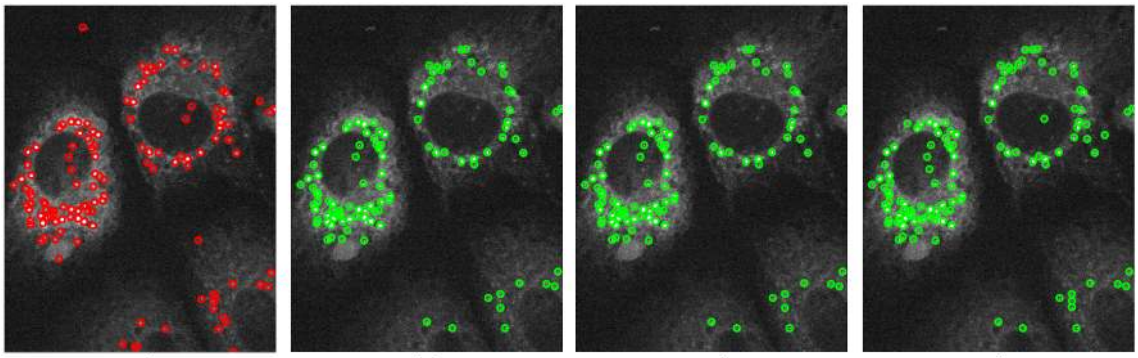


Figure 7.2: Detection results for HCV live cell microscopy data with different hyperparameter optimizations. a) Ground truth annotated by an expert. b) Experiment 2 using Grid Search c) Experiment 2 using SMAC-RF. d) Experiment 3 using Grid Search.

global optimum, we used dense Grid Search with 35.000 evaluations distributed on 20 compute nodes. The performance of SEF detection was optimized and evaluated using the F1 score (balancing precision and sensitivity) and 128 ground truth annotations. The assignment between the ground truth annotations and SEF detections was determined using the Munkres algorithm [122] and a gating distance of 5 pixels. We computed the mean and standard deviation of ΔF1 showing the difference to the F1 score after the warm-up phase.

The results for the different optimizers are shown in Table 7.1. The best performance is obtained by SMAC-RF and SMAC-XGBoost with an F1 score of 0.872, which are the only optimizers reaching the global optimum. The largest improvement ΔF1 is obtained by SMAC-RF with 0.050. In Figure 7.2 c) the result for SMAC-RF (green circles) is shown together with the ground truth (red circles) in Figure 7.2 a) and the global optimum (Grid Search) in Figure 7.2 b). The F1 score as a function of the number of iterations for all optimizers is depicted in Figure 7.3. The fastest convergence is obtained by SMAC-RF and SMAC-XGBoost, whereas TPE is slowest.

Image Pre-Processing for Detection of HCV Proteins

With this experiment we show the importance of an infimum projection as visualization of the loss function to gain further insight on the dependency of the hyperparameters. We study an additional image pre-processing step for HCV protein detection by smoothing the image with a Gaussian filter with standard deviation σ_{Gauss} followed by subtracting the filtered image from the original image to enhance the particles and suppress background noise. We now have a 3D hyperparameter space containing σ_{Gauss}, σ_{LoG}, and c.

We computed the global optimum with Grid Search (total: 175.000 evaluations). The F1 score using pre-processing is 1.6% higher compared to the result without

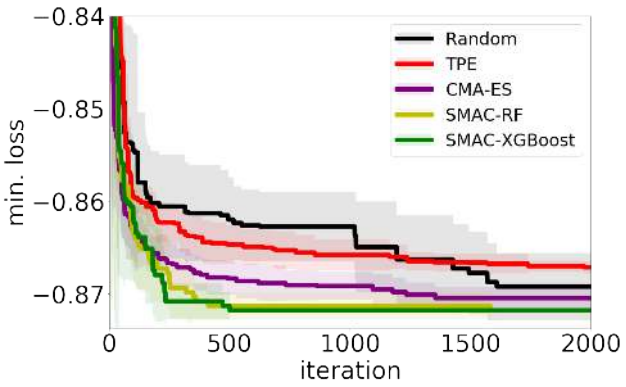


Figure 7.3: Convergence of different optimizers as a function of the number of iterations.

Table 7.2: Results of PCA for the whole loss surface data. The table provides the eigenvectors and eigenvalues of the four principal components (PC) together with the ratio between the cumulative variance and the total variance in [%].

PCA	Variable	PC 1	PC 2	PC 3	PC 4
Eigenvectors	c	0.224	-0.948	$0.001 \cdot 10^{-13}$	0.224
	σ_{LoG}	-0.668	-0.316	-0.081	-0.668
	σ_{Gauss}	-0.054	-0.026	0.997	-0.054
	$loss$	-0.707	$-0.002 \cdot 10^{-12}$	$0.003 \cdot 10^{-14}$	0.707
Eigenvalues		1.692	1.000	1.000	0.308
Cumulative variance ratio		42.3 %	67.3 %	92.3 %	100.0 %

pre-processing. To obtain insights into the optimization process and to quantify the dependency of the hyperparameters, we conducted a principal component analysis (PCA) [40] of the loss function. The results are shown in Table 7.2. The values of the loss function were normalized (zero mean and variance of one), and the eigenvectors with corresponding eigenvalues were computed (principal components, PCs). It can be seen that in order to represent 90% of the variance, the first three PCs need to be taken into account. For the first PC the hyperparameter σ_{Gauss} has a more than ten times smaller influence than c and σ_{LoG}. In addition, the other two PCs have a minor influence on the loss. Therefore, the influence of σ_{Gauss} on the loss is relatively small. However, pre-processing by a Gaussian filter improved the detection performance by 1.6%.

To further investigate the dependencies of the hyperparameters we propose to generate infimum projection visualizations. The infimum projection of a countable finite n-dimensional loss $\mathrm{L} : Q_1 \times ... \times Q_n \rightarrow \mathbb{R}$ into a lower dimensional projection P onto the index set $I \subseteq [\#Q]$ of features $Q = \{Q_1, ..., Q_n\}$ with m elements can be

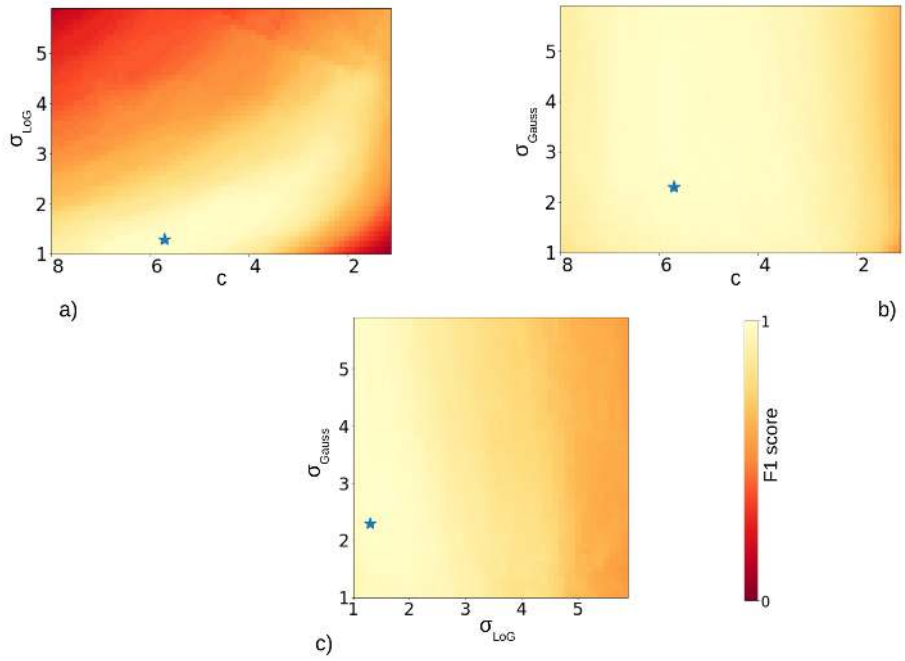
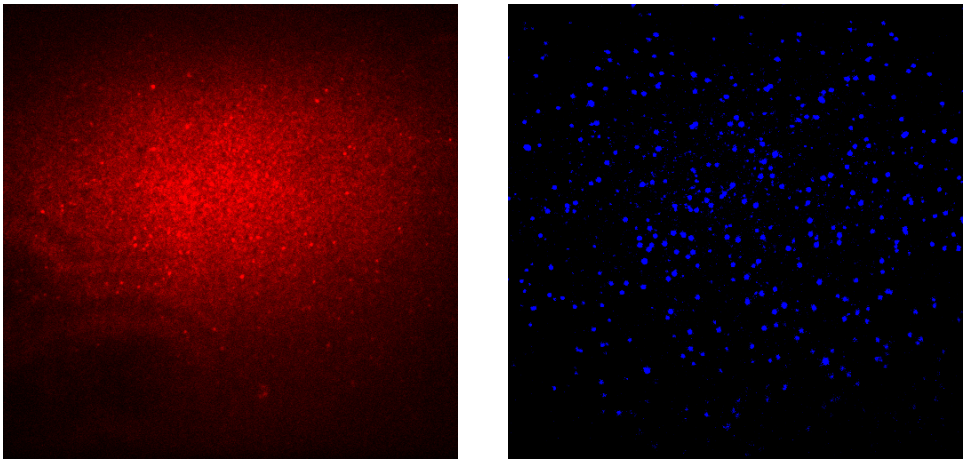


Figure 7.4: Infimum projections of the loss surface from experiment 3 for the 3D hyperparameter space (c, σ_{LoG}, and σ_{Gauss}) sampled with Grid Search. The global optimum is marked with a blue star.

performed by:

$$\mathrm{P}(I; q_1, ..., q_n) = \min_{\substack{q_k \in Q_k \\ k \notin I}} \{\mathrm{L}(q_1, ..., q_n)\} \tag{7.1}$$

In Figure 7.4 a)-c) the infimum projections between the three hyperparameters are shown. From the loss surfaces in Figure 7.4 b) and c) one can see that both the hyperparameters c and σ_{LoG} seem to be independent from the hyperparameter σ_{Gauss} due to the homogeneous structures of the loss surfaces (compare with Figure 7.4 a)). Thus, the infimum projection yields additional information to the PCA analysis and indicate that the optimization problem can be restructured by optimizing separately σ_{Gauss} and c along with σ_{LoG}. This sequential optimization procedure reduces the 3D hyperparameter optimization to a 1D optimization along with a 2D optimization.



(a) Low SNR channel 1 (b) High SNR channel 2

Figure 7.5: Two-channel microscopy image of HIV-1 particles showing different proteins in different channels.

Hyperparameter Optimization for Particle Detection in Multi-Channel Microscopy Images

In previous work, different classical methods (cf. Section 2.4.1) as well as deep neural networks (cf. Section 2.4.3) were introduced for particle detection in fluorescence microscopy data. A disadvantage of these methods is that several parameters need to be tuned and that supervised deep learning methods require manually labeled ground truth. However, manual annotation is tedious, time consuming, and error prone, and often not feasible due to low SNR, background noise, and high object density (see Fig. 7.5 a).

We have developed a novel method to optimize hyperparameters for particle detection in the lower SNR channel of multi-channel microscopy data (see Fig. 7.5). Our method exploits the information from two-channel fluorescence microscopy images, is weakly supervised, and does not require labeled ground truth. Typically, particle detection approaches are optimized using a metric such as the F1 score, which measures the similarity between the detections and the ground truth [108, 110]. In contrast, we propose exploiting colocalization information using a novel measure based on the double labeling efficiency (DLE). The DLE quantifies the labeling efficiency of double-labeled virus proteins [24] in two-channel microscopy images. Often, one channel has a lower SNR and the other a higher SNR (see Fig. 7.5). Our approach exploits the colocalization information in *both* channels by a novel measure based on the DLE to optimize the detections in the challenging low SNR channel, for which ground truth is difficult to determine. For the high SNR channel, ground truth can be relatively easily determined and used for optimization. Due to the labeling procedure, not all particles show a fluorescence signal in both channels. Therefore, for example, the high SNR channel cannot be used as ground truth to optimize the detection result based on the F1 score for the low SNR channel. Instead,

in our method we exploit in the high SNR channel only the information, where colocalization with the low SNR channel exists.

We evaluated our method based on synthetic data and found that our method (without requiring ground truth) yields comparable results to using the F1 score with ground truth. We also applied our method to challenging live cell images of human immunodeficiency virus type 1 (HIV-1) and obtained an improvement compared to the F1 score.

Double Labeling Efficiency for Particle Detection

Our proposed method exploits the colocalization of particles in two-channel fluorescence microscopy data, which is important to study virus-cell interactions [24]. The colocalization of double-labeled particles in channel 1 and channel 2 of a microscopy image can be quantified by the double labeling efficiency (DLE). The DLE is a measure for the fluorescent labeling of virus proteins. For channel k (with $k \in \{1, 2\}$), the measure is defined by

$$\mathrm{DLE}_k = \frac{n_{\mathrm{coloc},k}}{n_{\mathrm{total},k}} \tag{7.2}$$

where $n_{\mathrm{coloc},k}$ is the number of colocalized particles and $n_{\mathrm{total},k}$ the total number of detected particles in channel k. DLE_k yields values in the interval $[0, 1]$. $\mathrm{DLE}_k = 1$ means that all detected points in channel k are colocalized. If no particles are detected in channel k, then $\mathrm{DLE}_k = 0$. Without loss of generality we assume that channel 1 has a lower SNR than channel 2. To determine suitable hyperparameters of a detection method for the low SNR channel, DLE_1 in (7.2) could be optimized. However, doing this yields multiple global optima with DLE_1=1 due to multiple combinations of $n_{\mathrm{coloc},1} = n_{\mathrm{total},1}$. In addition, (7.2) includes only the DLE from one channel. To optimize the detection hyperparameters for the low SNR channel without requiring ground truth, we suggest fusing the information of both channels by a new measure that weights the DLE_k of the two channels by the harmonic mean:

$$\mathrm{DLE}_{harm} = \frac{2 \cdot \mathrm{DLE}_1 \cdot \mathrm{DLE}_2}{\mathrm{DLE}_1 + \mathrm{DLE}_2} \cdot \min\left\{r\,,\, \frac{1}{r}\right\} \tag{7.3}$$

where the ratio $r = \frac{n_{\mathrm{total},2}}{n_{\mathrm{est},2}}$ is defined by the total number of detected particles divided by the estimated number of particles in the high SNR channel. A (rough) estimate of the number of particles $n_{\mathrm{est},2}$ in the high SNR channel can straightforwardly be determined (e.g., by visual assessment or using a standard thresholding scheme). In (7.3), $\min\left\{r\,,\, \frac{1}{r}\right\} \in [0, 1]$ is used to restrict the values of DLE_{harm} in the interval $[0, 1]$. Further, this term also ensures for multiple combinations of $n_{\mathrm{coloc},1} = n_{\mathrm{total},1}$ (i.e. $\mathrm{DLE}_1 = 1$) along with $n_{\mathrm{coloc},2} = n_{\mathrm{total},2}$ (i.e. $\mathrm{DLE}_2 = 1$) that $\mathrm{DLE}_{harm} = 1$ only for $n_{\mathrm{total},2} = n_{\mathrm{est},2}$ ($n_{\mathrm{est},2}$ is a fixed value). Therefore, DLE_{harm} contains only one global optimum with $\mathrm{DLE}_{harm} = 1$. If no particles are detected in any of the two channels,

then $\mathrm{DLE}_{harm} = 0$.

Interestingly, optimizing DLE_{harm} in (7.3) can be interpreted as optimizing the F1 score. The F1 score is often used to assess the performance of detection methods, but requires ground truth. The score quantifies the similarity between two data sets by combining Precision and Recall using the harmonic mean [108, 110]:

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{7.4}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \quad , \quad \mathrm{Recall} = \frac{TP}{TP + FN} \tag{7.5}$$

where TP are true positive, FP false positive, and FN false negative detections. In contrast to the F1 score, DLE_{harm} exploits the information about the position due to utilizing colocalization properties and takes into account information about the number of particles due to the ratio r. If $r = 1$ ($n_{\mathrm{total},2} = n_{\mathrm{est},2}$), then DLE_1 of the lower SNR channel is equal to the Precision in (7.5). Assuming the fluorophores in the higher SNR channel perfectly label all particles, then DLE_2 is equal to the Recall in (7.5), with $n_{\mathrm{total},2}$ being the sum of TP and FN detections in the lower SNR channel.

Experimental Results

We quantitatively compared our proposed method based on DLE_{harm} (without ground truth) to the F1 score (with ground truth) using generated synthetic two-channel images. For both measures in (7.3) and (7.4), we used the spot-enhancing filter (SEF) [106] for particle detection. SEF is based on a Laplacian-of-Gaussian filter with standard deviation σ_{LoG}, followed by thresholding the filtered image with a parameter c. To optimize σ_{LoG} and c for the low SNR channel based on DLE_{harm} and F1, we used the hyperparameter optimization framework HyperHyper [197] with Covariance Matrix Adaption Evolution Strategy. The detection performance for both measures is determined by the F1 score (which favors the method based on the F1 score compared to our method).

In the first experiment, we evaluate the influence of particle density (number of particles) on the detection performance of DLE_{harm} and F1 score for five two-channel images containing 100 up to 500 particles per channel. The data is challenging due to possible particle overlap for increasing particle density. The images of both channels contain the same number of particles located at the same position. The results are shown in Fig. 7.6. It can be seen that DLE_{harm} and F1 yield very similar results. Both measures show a slight performance decrease for increasing particle density due to particle overlap resulting in detection errors.

The second experiment examines the impact of chromatic aberration of the microscope on the detection performance. We used 50 particles in both channels which

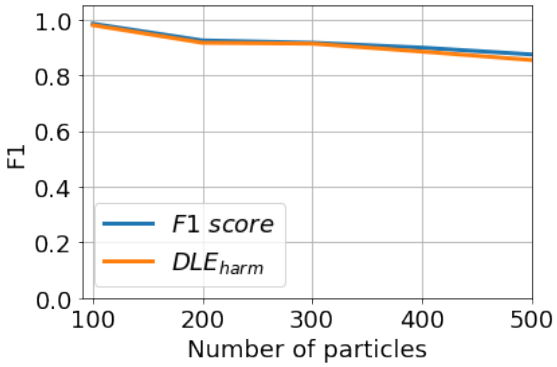
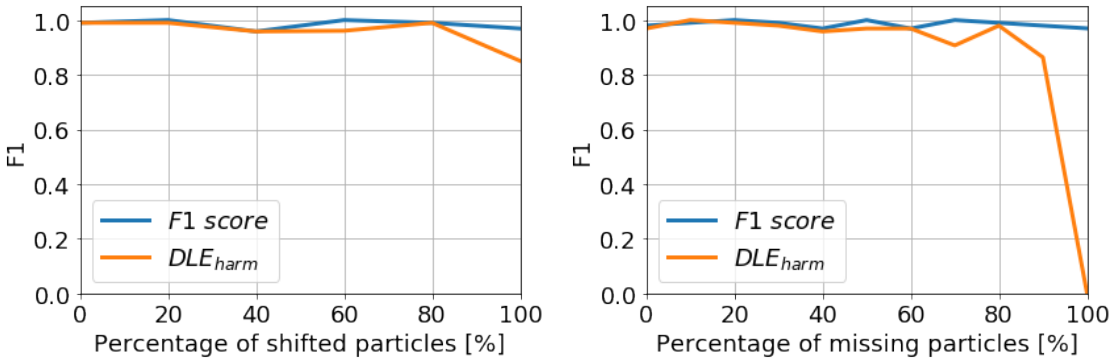


Figure 7.6: Influence of particle density in ch. 1 and ch. 2 on the detection performance for optimized DLE$_{harm}$ and F1 score.



(a) Influence of shifted particles

(b) Influence of missing particles

Figure 7.7: a) Influence of shifted particles in ch. 2 on the detection performance for optimized DLE$_{harm}$ and F1 score. b) Influence of missing particles in ch. 2 on the performance for optimized DLE$_{harm}$ and F1 score.

were initially at the same position. Then, a certain percentage of the particles in the high SNR channel 2 were randomly shifted by a distance of more than the maximum distance used for colocalization (5 pixels). This reduces the number of colocalizations that can be exploited by DLE$_{harm}$ to optimize the parameters. We increased the percentage of randomly shifted particles from 0% to 100%. The results are shown in Fig. 7.7 a). DLE$_{harm}$ yields very similar results as the F1 score even for a high percentage of shifted points.

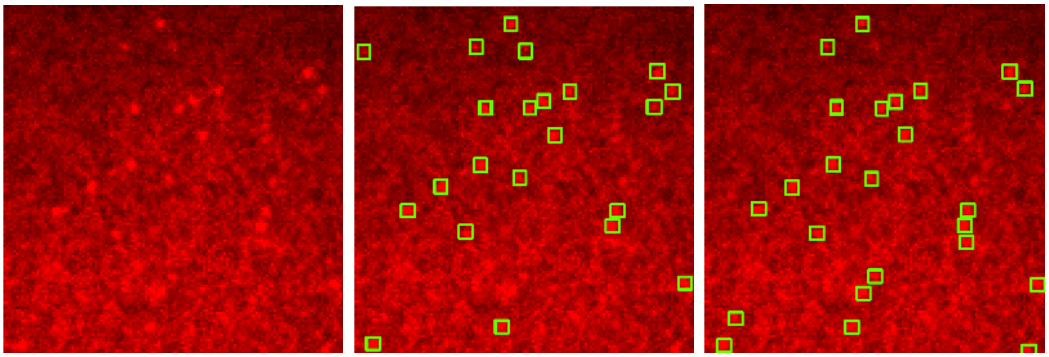
The third experiment evaluates the influence of missing particles in the high SNR channel 2 on DLE$_{harm}$. Missing particles simulate the case when the fluorophores do not perfectly label all particles. We used 50 particles for channel 1 and increased the number of missing particles in channel 2 from 0% to 100%. Particles in channel 1 and 2 are located at the same position. The results are shown in Fig. 7.7 b). The performance of DLE$_{harm}$ is close to the F1 score up to 90% of missing particles yielding a detection performance above 85%.

We also evaluated our method using six two-channel live cell microscopy images of double-labeled HIV-1 particles, which were selected as an example for challenging

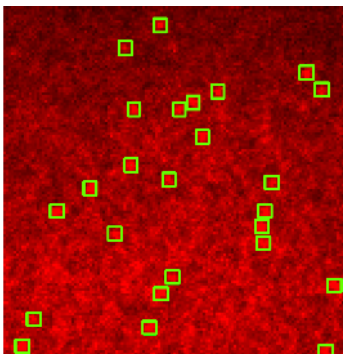
Table 7.3: Detection performance for six two-channel HIV-1 live cell microscopy images.

Image data	Method	F1	RMSE	min. Dist.	max. Dist.
1	F1 score	**0.721**	**0.529 ± 0.254**	**0.001**	**1.003**
	DLE_{harm}	0.694	0.543 ± 0.264	**0.001**	1.515
2	F1 score	**0.630**	0.580 ± **0.298**	0.001	**1.168**
	DLE_{harm}	0.606	**0.562** ± 0.312	**0.0004**	1.182
3	F1 score	**0.602**	**0.902 ± 0.459**	**0.002**	**1.803**
	DLE_{harm}	0.479	1.060 ± 0.648	**0.002**	4.626
4	F1 score	**0.630**	0.992 ± 0.641	**0.001**	**4.677**
	DLE_{harm}	0.570	**0.802 ± 0.626**	**0.001**	4.998
5	F1 score	**0.473**	1.035 ± **0.589**	0.004	**2.857**
	DLE_{harm}	0.389	**0.905** ± 0.630	**0.001**	4.124
6	F1 score	**0.615**	1.123 ± 0.794	0.004	4.727
	DLE_{harm}	0.595	**0.705 ± 0.463**	**0.002**	**2.001**
Mean	F1 score	**0.612 ± 0.080**	0.860 ± 0.257	0.002 ±**0.001**	**2.706 ± 1.342**
± Std. Dev.	DLE_{harm}	0.556 ± 0.107	**0.763 ± 0.201**	**0.001 ± 0.001**	3.074 ± 1.696

data sets. The data is challenging since channel 1 has a very low SNR and includes inhomogeneities due to uneven illumination (see Fig. 7.5). The images have a size of 512×512 pixels. The maximum Euclidean distance for colocalization was set to 5 pixels. For the quantitative evaluation, ground truth was manually annotated for the low SNR channel 1. By visual assessment we determined a rough number of 150 particles in the high SNR channel 2 for all six images. For particle detection, we used the SEF filter. To quantify the detection performance for channel 1, we computed the F1 score, RMSE as well as the minimum and maximum distance to the ground truth. The results are shown in Tab. 7.3. It can be seen that DLE_{harm} yields comparable results as the F1 score, though the number of particles in channel 2 was only coarsely determined and fixed for all images. The localization performance quantified by RMSE and minimal distance is improved by DLE_{harm} compared to F1. Examples of the image data are shown in Fig. 7.8 a. The ground truth of channel 1 is depicted in Fig. 7.8 b. Detection results obtained by optimizing the F1 score and DLE_{harm} are shown in Fig. 7.8 c and d, respectively. It can be seen that the result for DLE_{harm} agrees well with that for the F1 score.



(a) Original image (ch. 1) (b) Ground truth (ch. 1) (c) Opt. F1 score (ch. 1)



(d) Opt. DLE_{harm} (ch. 1)

Figure 7.8: Example detection results for the live cell HIV-1 data.

Summary We introduced a novel method for optimizing the detection performance in two-channel fluorescence microscopy images without requiring ground truth. Our approach exploits colocalization properties and requires only a rough estimate of the number of particles for the channel with higher SNR. We proposed a new measure DLE_{harm} (without using ground truth) and compared it with the F1 score (using ground truth) on synthetic as well as on challenging live cell microscopy images of HIV-1 particles. We found that DLE_{harm} yields comparable detection results as the F1 score and improves particle localization for live cell microscopy images. In our work, we used SEF as detection method, however, our method is also applicable to other particle detection methods.

7.2 Multiple Hypothesis Tracking with Integrated Cell Division Detection

Cell tracking is essential for understanding complex biological processes including immune response, tissue repair, embryonic development, tumor metastasis, and vascular diseases [219]. Besides varying cell densities and low signal-to-noise ratio of microscopy images, a particular challenge encountered in cell tracking is cell division.

Most existing cell tracking approaches rely on the paradigm of detection-based data association [220, 221, 205]. Cells are first detected independently in each image frame and then associated between consecutive frames using a scoring function (e.g., spatially closest cells in two consecutive frames). Cell division is incorporated either by assigning one cell to two cells or by post-processing. Recently, graph-theoretical approaches were introduced that use information from two or more frames based on probability estimates for cellular events [222, 223, 224]. However, the probability estimates often require manually annotated data for training. Multiple Hypothesis Tracking (MHT) [136] uses information from multiple frames and has been applied to biological particle tracking [132] and cell tracking [135, 134]. Yet, [132, 135] focus on non- or weakly interacting objects and do not address division events. [134] used MHT to investigate cell behavior during angiogenesis. However, cell divisions are detected by a deterministic (distance-based) approach.

We have developed a novel MHT-based approach for cell tracking which integrates cell division detection, and is denoted as MHT-CD. Our probabilistic approach exploits information from multiple frames and considers uncertainty information for cell division detection. In addition, MHT-CD does neither require training nor post-processing for detecting cell divisions. The data association problem is translated to a graph-theoretical maximum weighted independent set (MWIS) problem. Cell divisions are integrated by remodelling the MWIS graph to resolve the incompatibility between certain track hypotheses. We evaluated our method using synthetic data as well as data from the Cell Tracking Challenge [221, 205] and obtained state-of-the-art or better results than previous methods.

MHT-CD

The proposed MHT-CD approach is based on Multiple Hypothesis Tracking (MHT) and integrates cell division detection. MHT is a probabilistic tracking approach that uses Kalman filtering for state estimation and exploits information from multiple frames to solve the data association problem [136, 138]. We here use a track-oriented MHT, where the data association problem can be represented as maximum weighted independent set (MWIS) problem, which is not possible for a hypothesis-oriented MHT. Our approach maintains multiple track trees, one for each cell. Nodes in each tree represent cell detections and branches correspond to possible tracks (track hypotheses). For each frame, the trees are extended using all feasible detection-

to-track associations and each track hypothesis is assigned a track score. Two track hypotheses are compatible if they do not share any detection at any frame. The solution to the data association problem (the best global hypothesis) can be determined as the highest scoring set of compatible track hypotheses by solving an MWIS problem [140]. To avoid extending trees with unlikely detection-to-track associations, gating is performed [139]. Each track hypothesis is then extended using all detections within its gating region and a dummy detection which indicates a missing detection. Each detection also initializes a new track tree representing the possibility of an appearing cell. In each frame, each cell can give rise to at most one detection, thus only one-to-one assignments are allowed. The track score $L(T_k^i)$ [139] of the i-th track hypothesis T at time point k is computed recursively as $L(T_k^i) = L(T_{k-1}^i) + \Delta L(T_k^i)$ with

$$\Delta L(T_k^i) = \begin{cases} \ln(1 - p_D) & \text{if no detection assigned} \\ \ln\left(\frac{A}{2\pi}\right) - \frac{1}{2}\ln|\mathbf{\Sigma}_k^i| - \frac{1}{2}d^2 & \text{otherwise} \end{cases} \tag{7.6}$$

where $p_D \in [0, 1]$ is the detection probability, d^2 the gating distance, $\mathbf{\Sigma}_k^i$ the covariance matrix of the estimated position of the track hypothesis, and A the image area.

The data association problem in our MHT-CD approach is translated to an MWIS problem by constructing an *undirected graph* $G = (V, E)$ with vertices $V = \{1, ..., n\}$ and edges $E \subseteq V \times V$. Each vertex i represents a track hypothesis and has a weight w_i equal to the corresponding track score. An edge (i, j) connects vertex i and j if track hypothesis i is incompatible with track hypothesis j. Given the binary decision variable x_i, the MWIS problem is to find an independent set $U \subseteq V$ which maximizes the sum of weights

$$\max_{\mathbf{x}} \sum_{i=1}^{n} w_i\, x_i \qquad \text{s.t.} \quad x_i + x_j \leq 1 \quad \forall\, (i, j) \in E \quad , \; x_i \in \{0, 1\} \tag{7.7}$$

The MWIS problem is solved by the exact Bron-Kerbosch algorithm [225] or the approximate Greedy Randomized Adaptive Search Procedure (GRASP) [226].

A cell division is defined as an event in which a cell divides into at least two daughter cells, implying that more than one track hypothesis within a single track tree is correct. We use three parameters to identify cell divisions in track trees: a quantile threshold q_{div}, a distance threshold d_{div}, and the number $n_{children}$ of cells that may result at most from a cell division. Cell divisions are identified based on the distribution of track scores in each track tree. If there exist two or more track hypotheses associated with a track score larger than q_{div}, the mother cell and the daughter cells are required to be in close distance (lower than d_{div}). In MHT, all track hypotheses of the same tree are incompatible since they share at least one detection and are therefore connected by an edge in the MWIS graph. As shown in Fig. 7.9, MHT-CD identifies cell divisions in the track trees and removes the corresponding

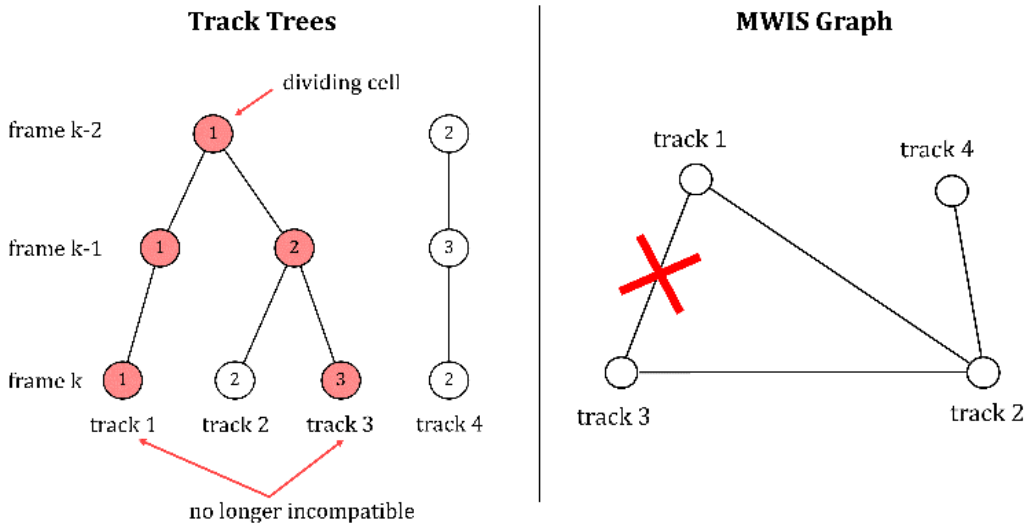


Figure 7.9: Cell division modelled by a removed edge in the MWIS graph.

edges from the MWIS graph. This resolves the incompatibility between the cell division track hypotheses and makes it possible to include them all in the solution of the MWIS problem, i.e. in the best global hypothesis.

Within an image sequence, not only single but also sequential cell divisions may occur if a daughter cell that has emerged from a previous cell division divides again. A specific property of cancer cells is the formation of more than two daughter cells, denoted as multipolar cell division. Sample track trees for a single, sequential, and multipolar cell division are shown in Fig. 7.10 a) to c). By identifying multiple track hypotheses with a high track score within one track tree, MHT-CD is able to remove all corresponding edges from the MWIS graph thereby representing also sequential and multipolar cell division events. We do not allow dummy detections to be involved in a cell division neither as dividing cell nor as a daughter cell. This is shown in Fig. 7.10 d) and e). Further, as shown in Fig. 7.10 f), we require for any time point after a cell division that no detection (except dummy detections) may be shared.

Experimental Results

We evaluated our approach using synthetic image sequences displaying challenging scenarios of cell divisions. To assess the impact of the developed cell division detection scheme, we compared our MHT-based approach with cell division detection (MHT-CD) to MHT without cell division detection (MHT). The tracking performance was quantified by the TRA score $\in [0, 1]$ of the Cell Tracking Challenge [227, 221]. Fig. 7.11 shows the tracking results for MHT-CD and MHT for a single, sequential, and multipolar division event together with the TRA score. It can be seen that all cell divisions are detected and resolved correctly by MHT-CD. In comparison, MHT performed well in tracking but cell divisions are considered as newly appearing cells, yielding a lower TRA score.

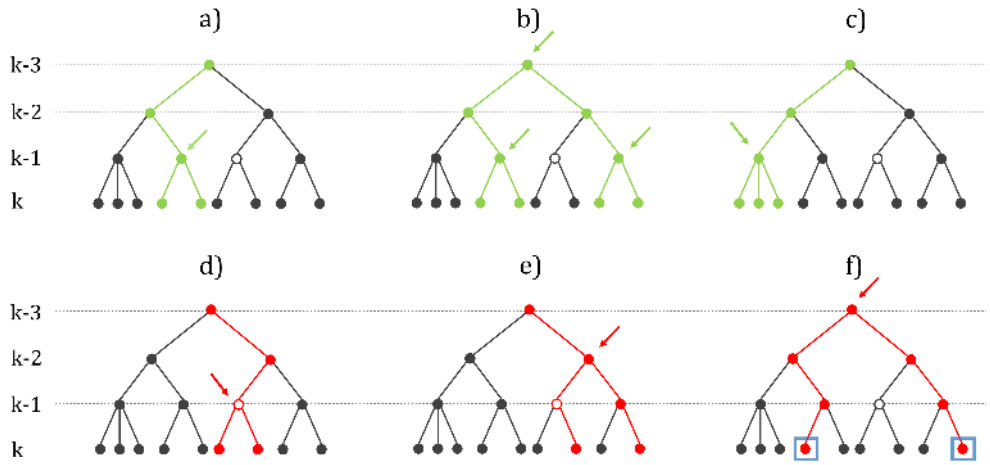
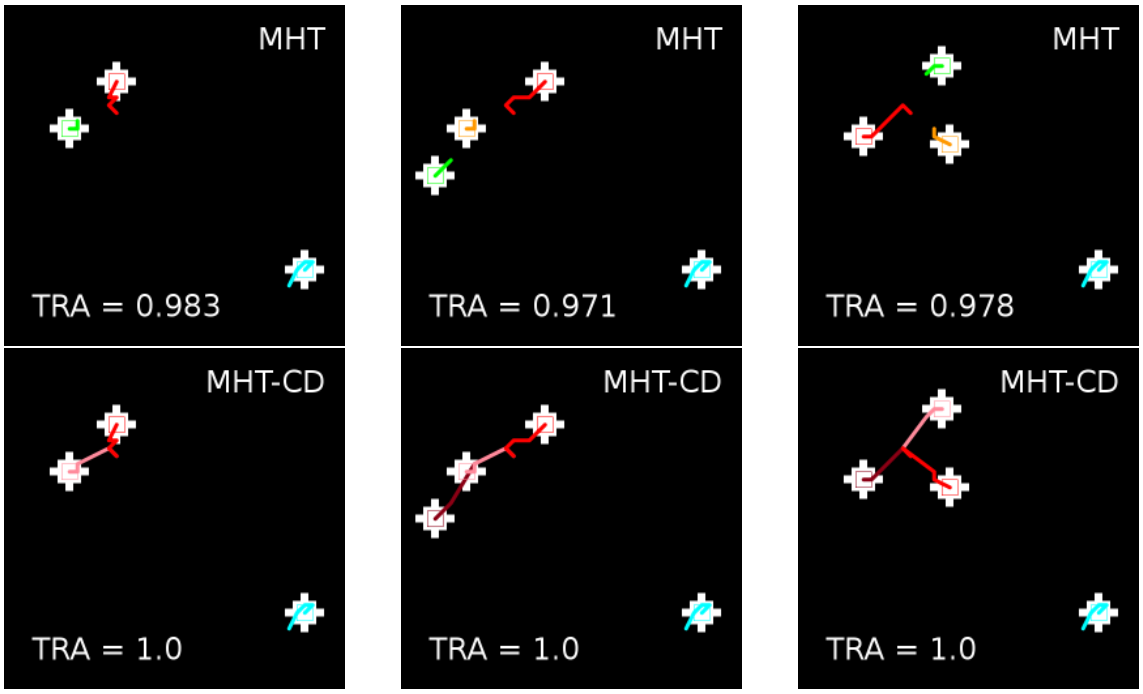


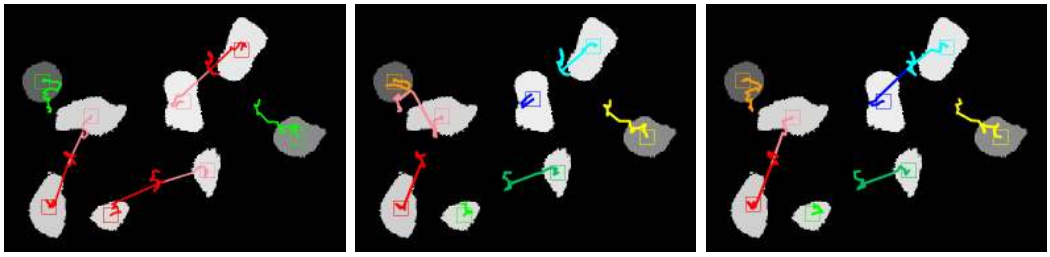
Figure 7.10: Valid and invalid cell division events. Filled nodes are detections, unfilled nodes are dummy detections. Nodes with an arrow correspond to dividing cells. a)-c) Valid cell divisions in green (single, sequential, multipolar). d)-e) Invalid cell divisions in red with dummy detections. f) Invalid cell division in red with identical detections (blue boxes).



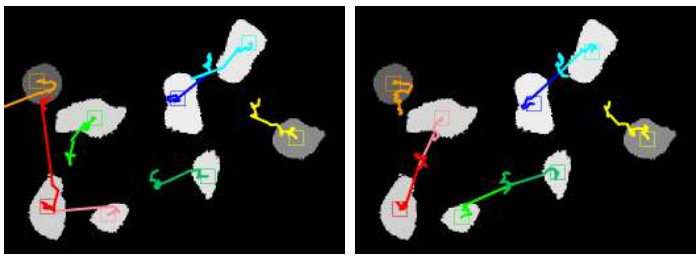
(a) Single (b) Sequential (c) Multipolar

Figure 7.11: Tracking results for different cell division scenarios obtained by MHT (top) and MHT-CD (bottom).

We also evaluated our approach using 8 image sequences with annotations from the Cell Tracking Challenge [221, 205]. We used the Fluo-N2DH-SIM and Fluo-N2DH-SIM+ datasets comprising 50 up to 150 image frames with 480×270 up to 739×773



(a) Ground truth (b) KTH-SE (c) UZH-CH



(d) UP-PT (e) MHT-CD

Figure 7.12: Tracking results for a section of the SIM-02 image sequence. Tracks originating from cell division have the same hue (e.g., light and dark red).

pixels each. We compared MHT-CD with a graph-based approach (KTH-SE [223]) and two distance-based approaches (UP-PT [220], UZH-CH [205]), which were used in the Cell Tracking Challenge [205]. Further, we compared MHT-CD with MHT without cell division detection. KTH-SE uses the Viterbi algorithm to select tracks from a state space diagram using information from multiple frames. Cell divisions are detected with a trained classifier. UP-PT and UZH-CH perform frame-by-frame nearest neighbour association based on Euclidean distances. UZH-CH detects a cell division if there are at least two nearest neighbour cells in the following frame. UP-PT detects cell divisions with a post-processing approach, which merges each track that does not start in the first image with the closest track starting at a previous time point. We quantified the tracking performance by the TRA score $\in [0, 1]$, and the cell division detection performance by the F1 score $\in [0, 1]$ as well as the number of true positives (TP) and false positives (FP). A detected cell division is considered to be correct if both the mother and daughter cells are correctly identified and the division occurs at the correct time point. In Table 7.4, the performance values for tracking and cell division detection are shown. The best method is highlighted in bold. MHT-CD yields state-of-the-art or better tracking results, and best results for four image sequences. Further, MHT-CD yields the best result for cell division detection for 6 out of 8 image sequences and ranks second for the two remaining sequences. Fig. 7.12 shows sample tracking results for a section of the SIM-02 image sequence. It can be seen that only MHT-CD identifies all cell divisions correctly. Fig. 7.13 shows the result of MHT-CD for a challenging section of the SIM-02+ sequence.

Table 7.4: Tracking performance of different methods for the Fluo-N2DH-SIM and Fluo-N2DH-SIM+ data.

Seq.	Method	TRA	F1	TP	FP
SIM-01 4 cell divisions	KTH-SE	0.9967	**1.000**	4	0
	UZH-CH	0.9853	0.500	3	5
	UP-PT	0.9782	0.000	0	4
	MHT	0.9947	–	–	–
	MHT-CD	**0.9998**	**1.000**	4	0
SIM-02 5 cell divisions	KTH-SE	**0.9967**	0.000	0	0
	UZH-CH	0.9873	0.889	4	0
	UP-PT	0.9699	0.000	0	4
	MHT	0.9286	–	–	–
	MHT-CD	0.9664	**1.000**	5	0
SIM-03 4 cell divisions	KTH-SE	0.9885	0.750	3	1
	UZH-CH	0.9804	0.400	2	4
	UP-PT	0.9917	0.000	0	4
	MHT	0.9838	–	–	–
	MHT-CD	**0.9998**	**1.000**	4	0
SIM-04 5 cell divisions	KTH-SE	0.9873	**0.889**	4	0
	UZH-CH	0.9821	0.727	4	2
	UP-PT	0.9789	0.000	0	8
	MHT	0.9819	–	–	–
	MHT-CD	**0.9976**	**0.889**	4	0
SIM-05 6 cell divisions	KTH-SE	**0.9840**	0.250	1	1
	UZH-CH	0.9658	0.334	4	14
	UP-PT	0.9405	0.000	0	18
	MHT	0.7824	–	–	–
	MHT-CD	0.8438	**0.667**	4	2
SIM-06 4 cell divisions	KTH-SE	**0.9936**	**1.000**	4	0
	UZH-CH	0.9827	0.545	3	4
	UP-PT	0.9696	0.000	0	8
	MHT	0.8921	–	–	–
	MHT-CD	0.9046	0.571	2	1
SIM-01+ 28 cell divisions	KTH-SE	**0.9860**	0.519	14	12
	UZH-CH	0.9859	**0.912**	26	3
	UP-PT	0.9571	0.103	3	27
	MHT	0.8941	–	–	–
	MHT-CD	0.9392	0.809	19	0
SIM-02+ 44 cell divisions	KTH-SE	0.9052	0.145	6	33
	UZH-CH	0.8541	0.336	22	65
	UP-PT	0.8747	0.132	5	27
	MHT	0.6487	–	–	–
	MHT-CD	**0.9376**	**0.782**	34	9

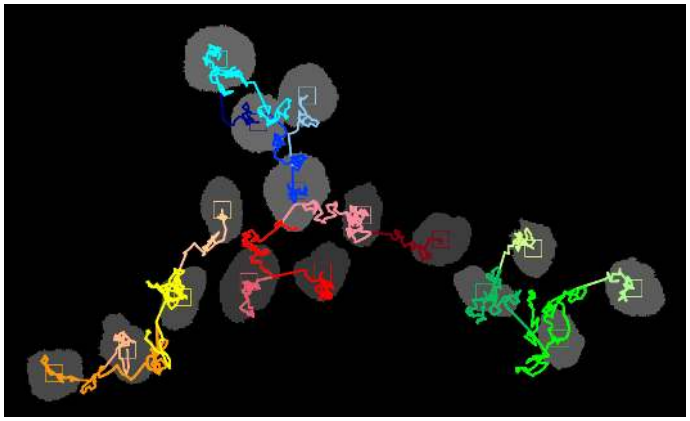


Figure 7.13: MHT-CD tracking results for a section of SIM-02+. Tracks originating from cell division have the same hue.

Summary We introduced a novel MHT-based approach for cell tracking with integrated cell division detection (MHT-CD). Our probabilistic approach exploits information from multiple frames. It remodels the MWIS graph to resolve the incompatibility of track hypotheses and automatically detect cell division. MHT-CD was evaluated using data from the Cell Tracking Challenge and yields state-of-the-art or better results for cell tracking and cell division detection compared to previous methods.

8 Colocalization and Motion Analysis of Subcellular Structures

In this chapter, methods for colocalization and motion analysis of subcellular structures are presented. The methods were applied to fluorescence microscopy image data displaying HIV-1 and HCV particles as well as chromatin structures. To support biologist during live cell experiments in determining spatial relationships of viral particles in multi-channel fluorescence microscopy images, a novel colocalization analysis approach was developed and integrated in a software suite. The multi-channel colocalization approach was evaluated for HCV particles. To determine HIV-1 infection and spread kinetics on a macroscopic level, motion analysis of HIV-1 in 3D collagen structures was performed and diffusion parameters were computed. In addition, to understand spatio-temporal nuclear organization of the genome during active DNA synthesis, motion and colocalization analysis of chromatin structures in confocal microscopy images was performed.

The work was published in Lee/Ritter *et al.* [18] and Imle/Ritter *et al.* [19], and has been submitted (Pham/Ritter *et al.* [20]) or is in preparation for publication (Ritter *et al.* [17]).

8.1 Colocalization Analysis of HCV in Multi-Channel Microscopy Images

HCV replication and assembly sites have been studied in two-channel time-lapse fluorescence microscopy images in conjunction with colocalization analysis of tracked HCV particles [18]. Further, the HCV motility of colocalized HCV proteins (envelope glycoprotein E2 and non-structural protein NS5A) has been analyzed for correlation with virus replication and assembly. It turned out, that a steady increase of the number of E2-NS5A double positive puncta was observed. Further, the velocity of E2 and NS5A was higher if the HCV proteins were not colocalized and reduced for E2-NS5A double positive puncta. The results are in accordance with the assumption that colocalized E2-NS5A corresponds to HCV assembly sites where the structural proteins indicated by E2 and the replicase indicated by NS5A become trapped in order to perform virus particle assembly. However, the characterization of temporal aspects of the HCV protein E2 recruitment to assembly sites was studied only in relation to the HCV protein NS5A and without an relation to virus-host proteins (e.g.,

ApoE [149]). Thus, colocalization analysis was performed for two-channel fluorescence microscopy image data and requires an extension to three-channel colocalization analysis for additionally taking into account virus-host proteins.

We have developed a software suite with a graphical user interface (GUI) for object-based detection, quantification, and visualizations of colocalizations of subcellular structures in multi-channel fluorescence microscopy images. The software enables to detect and colocalize particle-like spots in two-channel and three-channel fluorescence microscopy images. Colocalization is determined by the Python-based software *ColocQuant* which is based on a multi-dimensional graph-based k-d-tree approach to efficiently identify colocalized particles between two channels (double colocalization) or three channels (triple colocalization). In addition, colocalization parameters are quantified such as particle position, channel intensity, and object size for each channel. Visualization of colocalization is possible by the developed ImageJ macro *ColocJ*. This software visualizes the detections of double or triple colocalized objects. In addition, we propose an efficient and intuitive visualization of the color composition of colocalizations by a Maxwell color triangle (triple colocalization) and a color ribbon (double colocalization). Further, the spatial distribution of a triple colocalization is represented by a triangle visualizing the location and distances of particles participating in the triple colocalization. Global statistics of the intensity values, object size, and number of colocalizations over time are provided as box plots, bar plots, or table and can be exported and used in other software. ColocJ can be used for an entire image as well as for a manually selected region-of-interest (ROI) annotated with ImageJ built-in tools.

We illustrate the application of ColocQuant and ColocJ on data of multi-channel live cell microscopy images of hepatitis C virus (HCV) associated proteins. We performed colocalization in two as well as in three channels. Further, we show the visualization of the spatial distribution for triple colocalization and demonstrate color composition analysis for double and triple colocalization. The software suite was also applied in [20] to study intercellular transmission of viral RNA by ApoE associated extracellular vesicles.

Multi-Channel Colocalization with ColocQuant

ColocQuant is a Python-based software which uses a multi-dimensional graph-based approach to determine colocalized objects in two-channel (double colocalization) or three-channel (triple colocalization) fluorescence microscopy images. We use a k-d-tree [228] to perform a nearest neighbor search over two or three channels. In our approach, a triple colocalized particle is represented by a vector $\mathbf{v}_{\text{triple}} = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)^T$ which contains position vectors $\mathbf{p}_i = (x_i, y_i)^T$ for each channel i. $\mathbf{p}_i$ represents the position of a particle (viral structure) in channel i which is detected by the spot-enhancing filter (SEF) [106]. Particle detection by SEF consists of applying a Laplacian-of-Gaussian (LoG) filter with standard deviation $\sigma_{\text{LoG},i}$ (filter size),

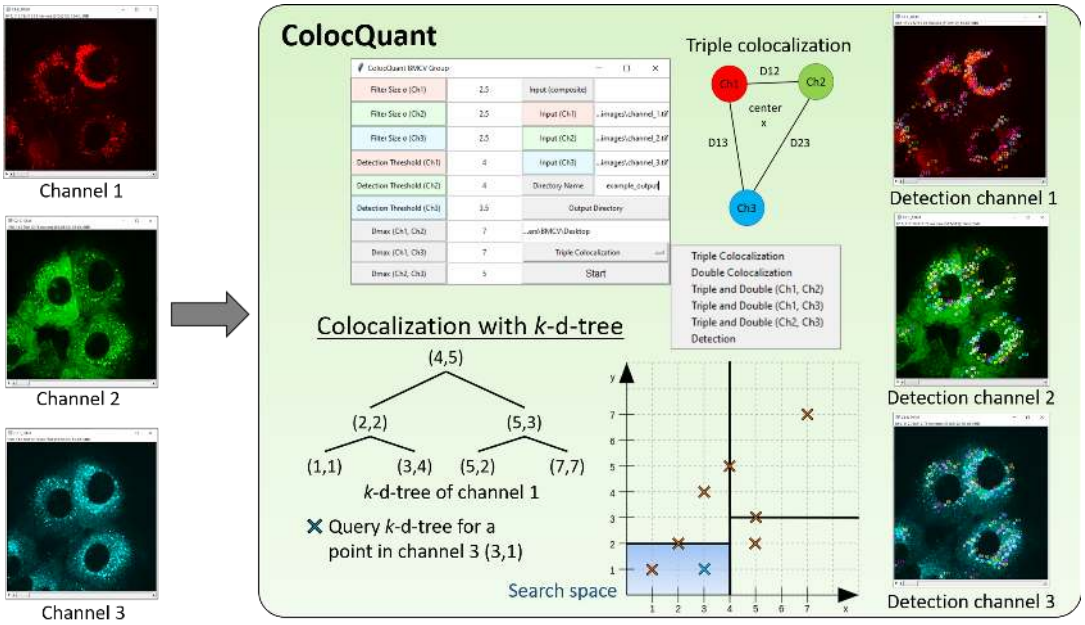


Figure 8.1: Overview of ColocQuant.

followed by thresholding the filtered image using the mean intensity plus a factor c_i (detection threshold) times the standard deviation. After particle detection, a k-d-tree for all particle positions $\mathbf{p}_i \in \mathbb{R}^k$ with $k = 2$ of channel i is constructed which is a binary tree in which each non-leaf node represents a splitting hyperplane that divides the search space into two parts. The root node $\mathbf{p}_{\text{root}}$ of the tree splits all points $\mathbf{p}_i$ into two subspaces according to one of the dimensions (x_i or y_i). The hyperplane is perpendicular to the chosen dimension (x_{root} or y_{root}) and all points $\mathbf{p}_i$ with $\{(x_i, y_i)|x_i < x_{\text{root}}\}$ (or $\{(x_i, y_i)|y_i < y_{\text{root}}\}$ for splitting along the y-dimension) are represented by the left subtree, whereas the remaining points are represented by the right subtree. Afterwards, the subtrees are split in a similar fashion and along alternating dimensions of $\mathbf{p}_i$. Once the k-d-tree is build, a nearest neighbor query between a point $\mathbf{p}_i$ in channels i and points $\mathbf{p}_j$ in channel j is defined as $\{(\mathbf{p}_{j,1} \neq \mathbf{p}_{j,2}) \Rightarrow [D(\mathbf{p}_i, \mathbf{p}_{j,2}) \geq D(\mathbf{p}_i, \mathbf{p}_{j,1})]\}$ where $D(\cdot)$ is the Euclidean distance and $\mathbf{p}_{j,1}$ is the nearest neighbor of $\mathbf{p}_i$ and $\mathbf{p}_{j,2}$ is the second nearest neighbor. We further assume that the nearest neighbor is within a distance $D(\mathbf{p}_i, \mathbf{p}_{j,1}) \leq D_{\text{max},i,j}$, which can be chosen in the GUI of ColocQuant by the user. The construction of a k-d-tree and the representative search space for querying the k-d-tree to determine the nearest neighbor is shown in Fig. 8.1.

To detect double colocalized objects $\mathbf{v}_{\text{double}} = (\mathbf{p}_1, \mathbf{p}_2)^T$, the k-d-tree of channel 1 needs to be queried only once with points from channel 2 using $D_{\text{max},1,2}$. For a triple colocalization, the k-d-tree of channel 1 needs to be queried by channel 2 (with $D_{\text{max},1,2}$) and 3 (with $D_{\text{max},1,3}$) as well as a second k-d-tree needs to be build based on points from channel 2 and queried by channel 3 (with $D_{\text{max},2,3}$). Afterwards, the intersection of all nearest neighbor queries is computed, which determines triple

colocalization within all three channels. Due to the data structure of k-d-trees, the nearest neighbor queries have an average running time of $\mathcal{O}(\log(n))$ where n is the number of particles in a channel [228, 229]. Therefore, this data structure is efficient, easy to construct, and well suited for finding nearest neighbors in order to determine triple colocalization.

Color Composition Analysis with ColocJ

Besides the spatial information of double or triple colocalized objects (e.g., position, size) determined by ColocQuant, intensity information is important to quantify fluorescence signals in each of the channels, or postprocess the identified colocalized objects by intensity thresholding. Therefore, a color analysis approach is required to quantify and visualize the color distribution and composition of double as well as triple colocalization.

In previous work, it was shown that each color c can be obtained by a weighted mixture of the primary colors within the trichromatic RGB space representation (R=red, G=green, B=blue) [230, 231] with $c = e\,R + f\,G + g\,B$. For a better and simpler comparison of the color composition of double and triple colocalized objects, ColocJ visualizes the proportions of the three primary colors within a 2D plot. For double colocalized objects, a color ribbon showing a color gradient between two colors is used to visualize the proportion of both color values. For a triple colocalized object, the Maxwell color triangle is suggested to visualize the proportion of the three primary colors. The Maxwell color triangle consists of three coordinate axes, one for each color, which are the altitudes of the triangle and ranging from zero (foot) to one (vertex of the triangle). The value of each primary color expresses its intensity compared to the intensity of the other two primary colors. To represent an RGB color value in the triangle, the RGB color value needs to be normalized by multiplication with a factor a such that $1 = a\,R + a\,G + a\,B$ is fulfilled. The factor a is based on the sum of the primary color values with $a = 1/(R + G + B)$. As an example, the Maxwell color triangle value of the 8-bit RGB color red with (R=255, G=0, B=0) is ($\mathrm{R_{Maxwell}}$=1, $\mathrm{G_{Maxwell}}$=0, $\mathrm{B_{Maxwell}}$=0) with $a = 1/255$ and located at the red vertex of the triangle. The 8-bit RGB color white (R=255, G=255, B=255) is multiplied by $a = 1/(255 + 255 + 255)$ and located at the center of the Maxwell color triangle. Note that the 8-bit RGB color (R=5, G=5, B=5) is multiplied with $a = 1/(5+5+5)$ and therefore also located at the center of the Maxwell color triangle. This is due to the fact that the Maxwell color triangle quantifies and visualizes the proportion of the three primary colors to each other. Therefore, the Maxwell color triangle provides additional object-based color information compared to the standard box plots of image intensities, which provide single channel global intensity statistics.

The computation of the color composition for double and triple colocalization is illustrated in Fig. 8.2. For a double colocalization, the composite image marks a double colocalized object by a yellow circle. The image intensities are extracted at

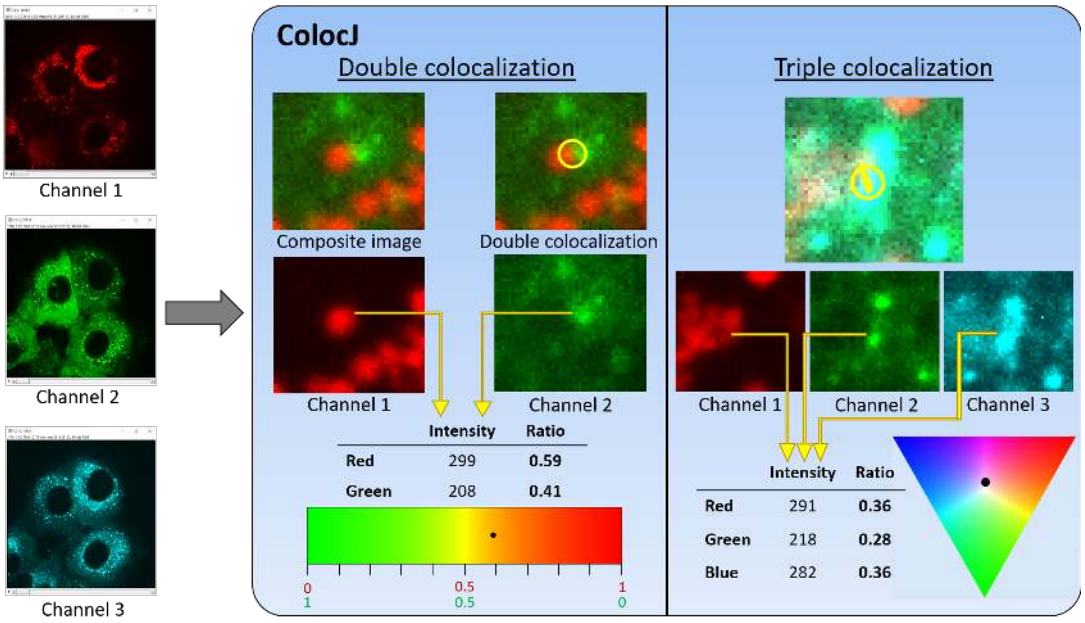


Figure 8.2: Overview of ColocJ.

the center of the double colocalization (center of yellow circle) for each channel and the ratio between the two intensities is computed. The intensity ratio determines the position on the color ribbon and represents the color composition of the double colocalization. In Fig. 8.2 the color composition analysis is shown for channel 1 and 2. Note that ColocJ also provides the possibility to analyse the color composition for colocalization of channel 2 with 3 as well as channel 1 with 3. A triple colocalization is marked with a yellow circle in the composite image whereas the yellow triangle vertices represent the position of detected particles in each of the three channels. The intensity information for each channel can be extracted at the center of the yellow circle representing the overlap region of the colocalization. After extracting the intensity values, the ratio for each primary color is determined and the color composition of a triple colocalization is marked with a black dot in the Maxwell color triangle.

Three-Channel Colocalization Analysis of HCV in Fluorescence Microscopy Images

We evaluated our software *ColocQuant* and *ColocJ* with live cell fluorescence microscopy image data displaying different hepatitis C virus (HCV) associated proteins. The image data consists of a three-channel time-lapse image sequence displaying HCV-infected cells with mCherry-labeled host cell protein apolipoprotein E (ApoE) (channel 1), GFP-labeled HCV envelope glycoprotein E2 (channel 2), and mTurquoise-labeled HCV non-structural protein 5A (NS5A) (channel 3) within Huh7/LunetCD81H cells [18]. The image sequence comprises 10 images with an

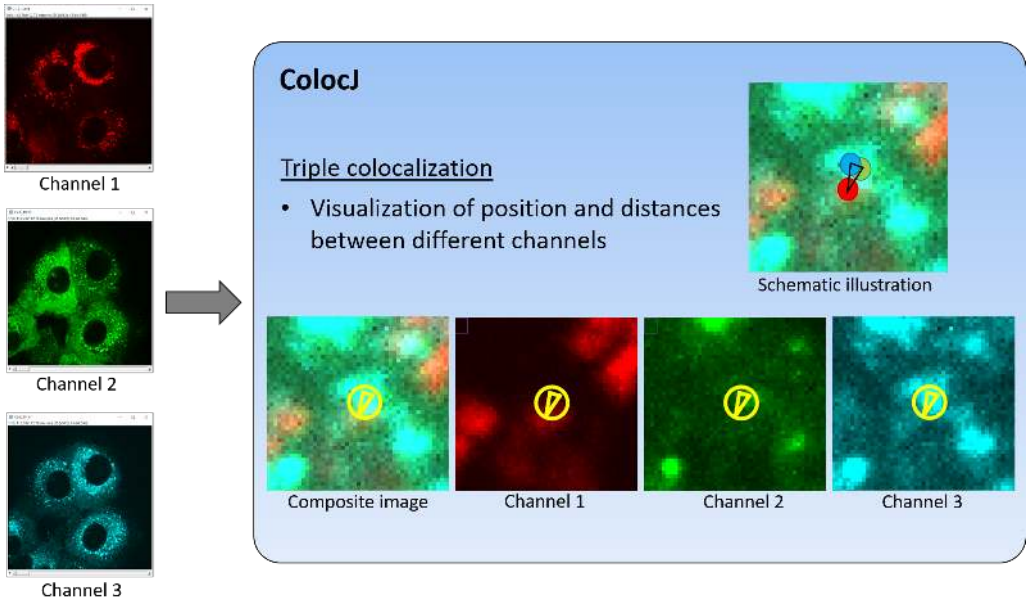


Figure 8.3: Visualization of spatial composition for triple colocalization using ColocJ.

image size of 512×512 pixels and a spatial resolution of 220 nm/pixel. Images were acquired with a PerkinElmer UltraVIEW ERS spinning disk confocal microscope mounted on a Nikon TE2000-E.

With the GUI of ColocQuant (see Fig. 8.1), we parameterized the SEF filter for particle detection with a *Filter Size* (LoG standard deviation $\sigma_{\mathrm{LoG},i}$) of 2.5 for all three channels. The *Detection Threshold* (thresholding the filtered image based on c_i) was set to 4.0 for channel 1 and 2, and to 3.5 for channel 3. For determining a triple colocalization, the maximum Euclidean distance between detections in the different channels was set to 7 for $D_{max}(Ch1, Ch2)$ and $D_{max}(Ch1, Ch3)$, and 5 for $D_{max}(Ch2, Ch3)$.

After calculating and identifying all triple colocalizations for each frame of the image sequence, the determined objects can be visualized by the ImageJ macro *ColocJ*. The visualization of the spatial composition and distribution for a triple colocalization is shown in Fig. 8.3. ColocJ uses a yellow circle to visualize the position where a triple colocalization occurs. The triangle vertices show the location of the detected particles of the triple colocalization in each channel and the edges provide information about the distances. The analysis and the visualization of the spatial composition of triple colocalizations can be done for composite images (overlay of all channels) as well as for each channel separately.

The color composition and statistical analysis of triple colocalizations by ColocJ is shown in Fig. 8.4. ColocJ provides a bar plot showing the number of colocalizations in each frame and allows the user to analyse the amount of triple colocalizations over time. To study viral replication and assembly, the amount of colocalization over time is an import factor to describe and model the temporal behavior of assembly processes. In addition, ColocJ visualizes the distribution of the object sizes of

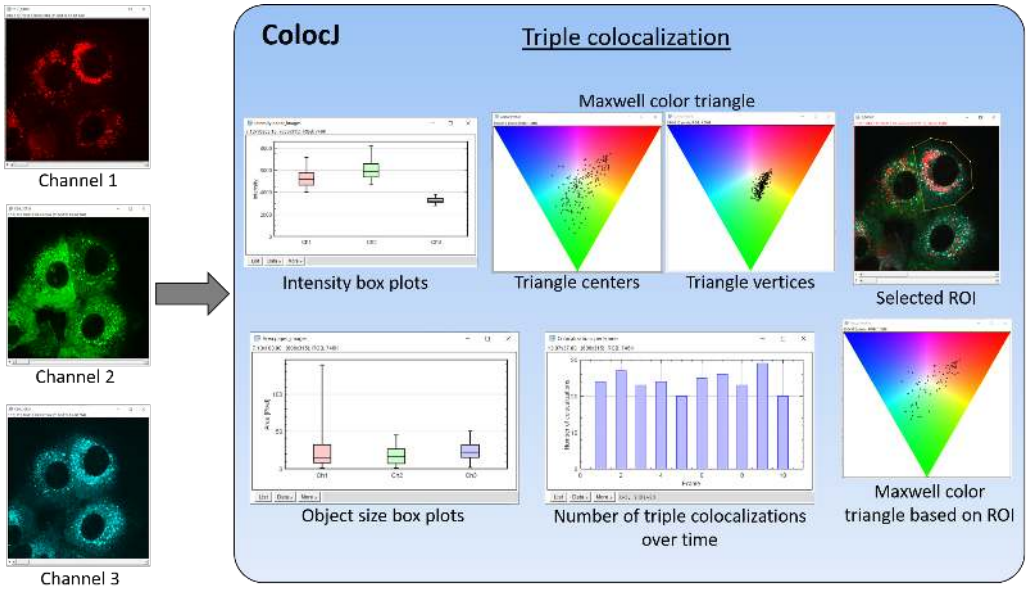


Figure 8.4: Triple colocalization analysis of HCV live cell image sequence using ColocJ.

particles for each channel participating in the triple colocalization. The object size is given in pixels and is determined by counting the pixels which belong to the connected component [232] representing the detected particle by the SEF filter. Based on the object sizes, the chemical process of fluorescence labelling as well as characterization of the viral protein can be evaluated and differences between channels can be determined. For the used HCV image sequence, all three channels show a similar object size for each channel and the number of colocalizations is almost stable over time. To analyze the intensity value distribution of triple colocalizations, ColocJ provides box plots of the intensity values for each channel as well as the Maxwell color triangle. The box plots visualize (global) statistics of the absolute intensity values independently of the remaining two channels. In comparison, the Maxwell color triangle shows the proportion of intensities between the three channels for each triple colocalization and represents an object-based statistical visualization. Therefore, the two visualization approaches display different aspects of the triple colocalization data and complement each other. In addition, ColocJ provides the possibility to extract the intensity information for the Maxwell color triangle at the center of the triple colocalization (center of the yellow circle shown in Fig. 8.3) as well as at the position of the yellow triangle vertices. The center of the yellow circle provides information about the area of overlap, whereas the triangle vertices extract information at the location of the detected particles in each channel contributing to the triple colocalization. Comparing both Maxwell color triangles for the HCV image sequence, the color compositions in the area of overlap (Triangle center) seems more heterogeneous due to the more widespread distribution within the Maxwell color triangle compared to the color composition based on the location of the particles

Results

File Edit Font Results

	number	X_mean	Y_mean	frame	X_ch1	Y_ch1	Intensity_ch1	Area_ch1	X_ch2	Y_ch2	Intensity_ch2	Area_ch2	X_ch3	Y_ch3	Intensity_ch3	Area_ch3
1	1	143.8683	132.2042	1	145.8777	134.6768	4930.7857	11	143.3446	131.1552	6616.0291	34	142.3828	130.7808	3074.8972	16
2	2	161.6578	125.011	1	164.1444	121.9458	4887.3146	29	161.9014	126.9156	4644.402	37	158.9277	126.1715	2784.9698	15
3	3	169.2292	263.3473	1	169.877	264.3349	4664.3629	9	169.4905	262.9389	5959.2051	27	168.3201	262.7682	2758.9361	12
4	4	170.6658	135.7957	1	167.4618	132.8855	5371.8535	49	172.6493	137.3679	6315.8106	26	171.8864	137.1337	3222.2569	8
5	5	173.5404	150.8256	1	174.3915	146.4532	5879.7969	34	173.6524	152.9839	7955.0118	42	172.5773	153.0396	3184.1629	30
6	6	192.4086	154.1666	1	191.4752	149.6876	5358.1788	17	193.584	156.331	6724.1519	26	192.1668	156.4813	3340.6276	10
7	7	241.1697	148.0253	1	240.945	149.462	5079.4983	32	242.0556	146.9313	5365.0705	5	240.5087	147.6826	3237.0314	29

Figure 8.5: Results of triple colocalization analysis of multi-channel live cell HCV image sequence represented as table by ColocJ. The table provides information about the number of colocalizations, position, frame number, intensity, and object size of the particles involved in the triple colocalization.

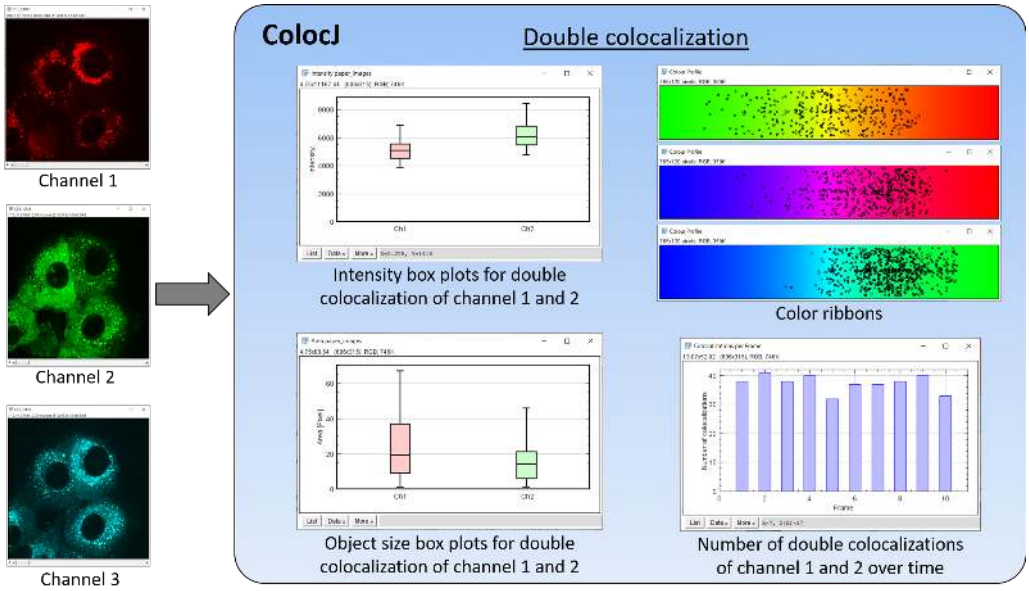


Figure 8.6: Double colocalization analysis of HCV live cell image sequence with ColocJ.

(triangle vertices). In order to analyze the live cell image data at the single-cell level, ColocJ provides the possibility to visualize and analyse triple colocalization for a manually selected ROI (see Fig. 8.4). The ROI can easily be annotated by the ImageJ built-in area selection tool *Polygon* and imported to ColocJ. All visualized data of the colocalization analysis by ColocJ is also provided in a table format shown in Fig. 8.5. The table can be saved as a *CSV* (comma-separated values) file in ImageJ and therefore can easily be imported and exploited by other software.

Besides triple colocalization analysis, ColocQuant and ColocJ are designed to also analyse double colocalizations of two-channel fluorescence microscopy images. Based on the selected channels for colocalization analysis in ColocQuant, ColocJ visualizes the number of double colocalizations as a bar plot shown in Fig. 8.6. As for triple colocalizations, ColocJ also displays statistics about the object size as box plots. The color distribution of double colocalizations can be analysed by intensity box plots showing statistics of the absolute intensity values and the color composition can be

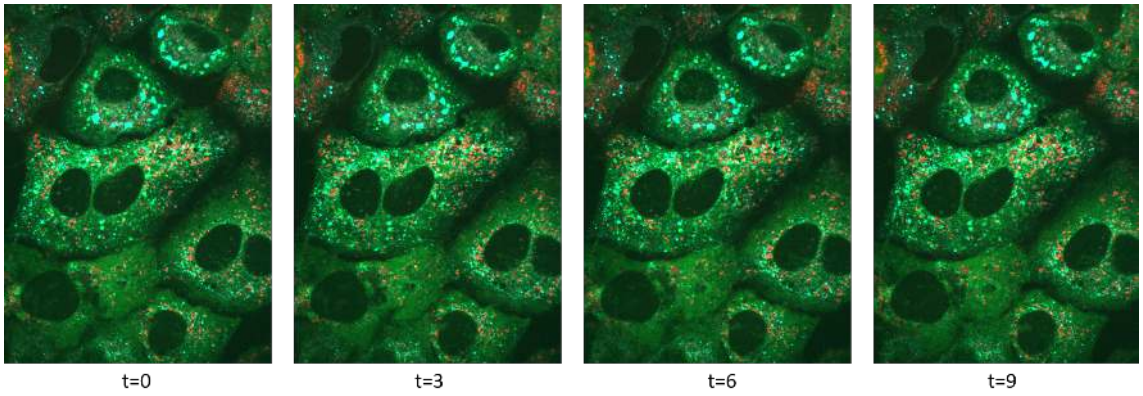
analysed with a color ribbon for object-based analysis. In Fig. 8.6 the three-channel HCV image sequence is analysed for double colocalization between channel 1 and 2 as an example. However, as indicated by the different color ribbons in Fig. 8.6, ColocQuant and ColocJ can analyse and visualize all combinations of channels selected by the user. In addition, our software suite is also able to quantify double colocalizations for a manually selected ROI. As for the triple colocalization, the user can easily import an ROI annotated by the ImageJ built-in area selection tool *Polygon*.

Maxwell Color Triangle and Image Intensity Variations

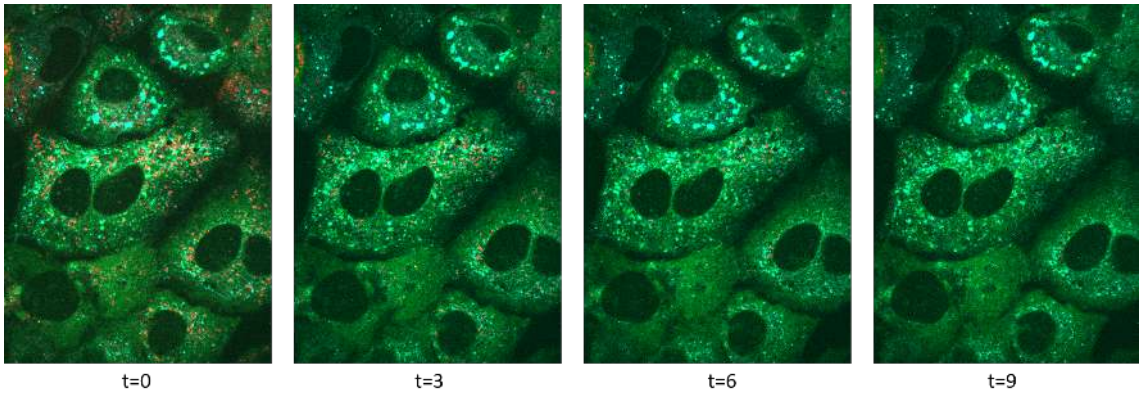
We evaluated the Maxwell color triangle in *ColocJ* for different typical intensity variations occurring in multi-channel fluorescence microscopy. Intensity variations can roughly be separated in *additive* and *multiplicative* intensity changes of the fluorescence channels. An *additive* intensity change over time means that the intensity values linearly increase or decrease with a constant factor. An additive intensity change in the labeled virus associated proteins with decreasing intensity values over time can occur, for example, due to virus-host cell interactions [25, 167, 233]. A *multiplicative* intensity change over time means that the intensity values non-linearly increase or decrease over time. As an example, photobleaching in fluorescence microscopy is a process where the number of fluorescence molecules is reduced due to permanent photochemical destruction and leads to an exponential (non-linear) intensity decrease over time [234, 235].

We assessed the Maxwell color triangle with a three-channel live cell fluorescence microscopy image sequence consisting of 10 frames with an image size of 1114×1496 pixels and a spatial resolution of 89 nm/pixel. The image data displays HCV-infected cells with mCherry-labeled host cell protein ApoE (channel 1), YFP-labeled HCV protein E2 (channel 2), and mTurquoise-labeled HCV protein NS5A (channel 3) [18]. Images were acquired with a PerkinElmer UltraVIEW VoX spinning disk confocal microscope mounted on a Nikon TiE.

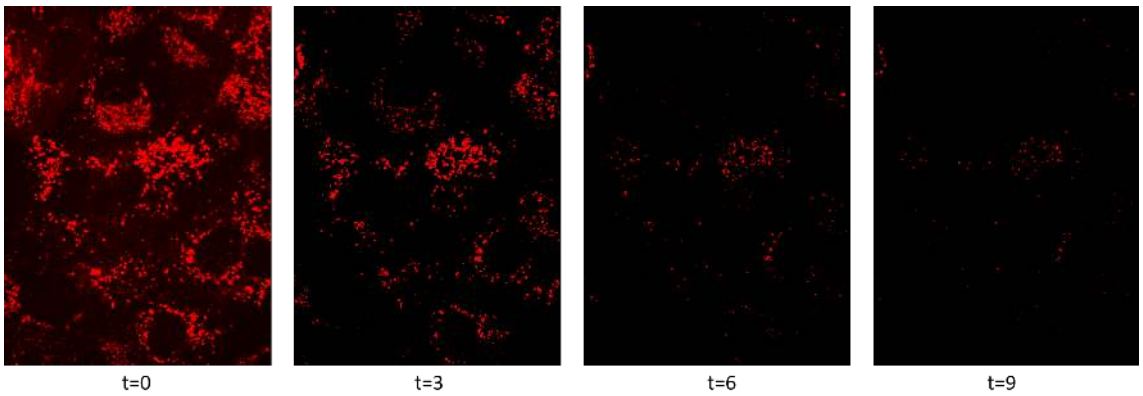
We evaluated the Maxwell color triangle for the case where intensity values of channel 1 (ApoE) decrease over time with a constant factor. For every time point, the original intensity values are decreased by 20 times the time point t (e.g., for $t = 9$ all intensity values of Ch 1 are decreased by 180). The intensity values of channel 2 (E2) and channel 3 (NS5A) remain unchanged. Fig. 8.7 shows the original composite images for different time points as well as the composite images with decreased channel 1 intensity values and images of channel 1. Fig. 8.8 shows the Maxwell color triangle for the original image sequence as well as for the image sequence with decreasing channel 1 intensity values. The black dots in the Maxwell color triangles for both image sequences represent the same triple colocalizations. It can be seen that the black dots in the Maxwell color triangle tend to shift torwards the blue/green color region for progressing time points since the red color is decreased in the color



(a) Original three-channel image sequence at different time points



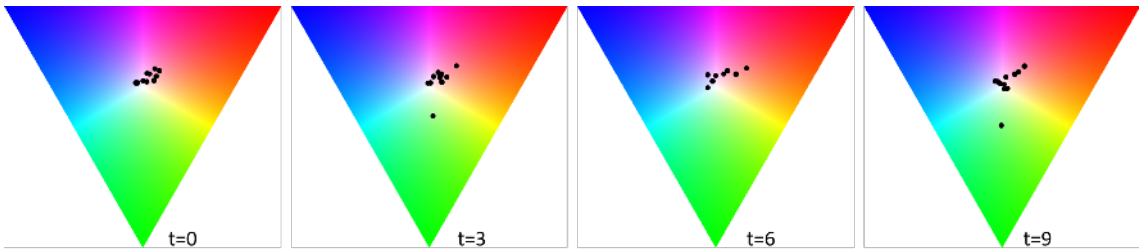
(b) Three-channel image sequence with decreasing intensity values in channel 1 (red)



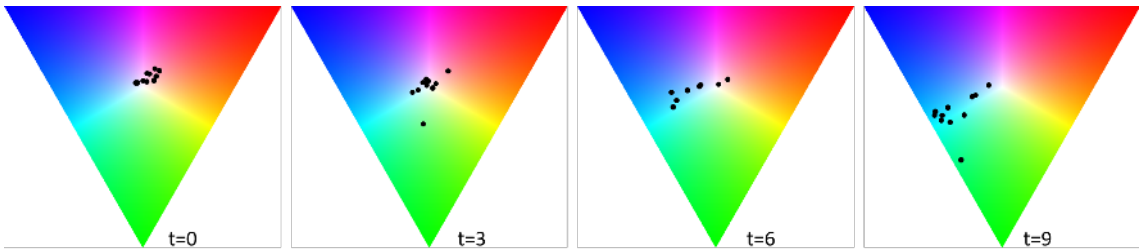
(c) Channel 1 with decreasing intensity values

Figure 8.7: Live cell fluorescence microscopy data of HCV. a) Original composite image sequence. b) Composite image sequence with decreasing channel 1 intensity values over time. c) Decreasing channel 1 intensity values over time.

composition due to decreasing channel 1 (red) intensity values. For time point $t = 0$ both Maxwell color triangles are equal since the intensity values for channel 1 are equal for both image sequences.



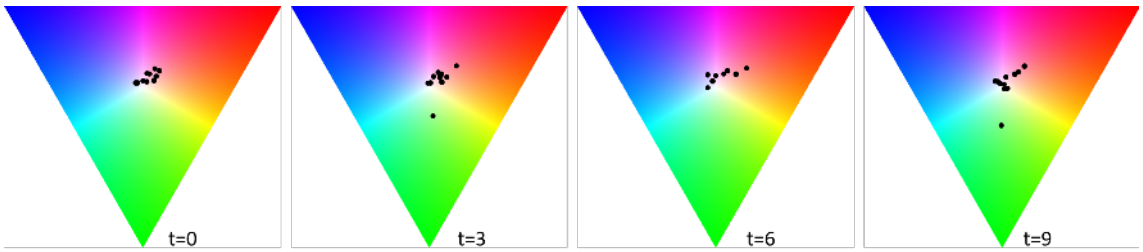
(a) Original image sequence



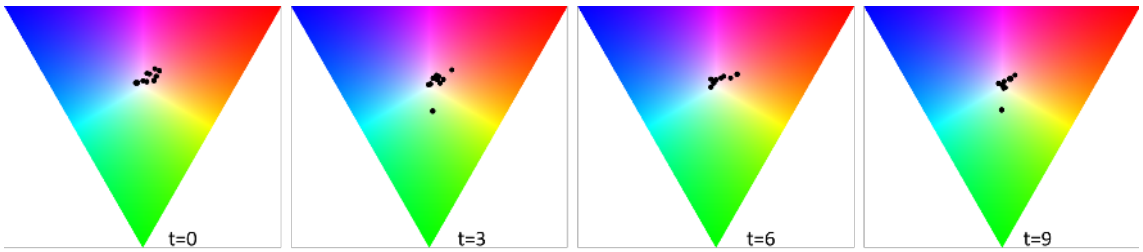
(b) Image sequence with decreasing intensity values in channel 1 (red)

Figure 8.8: Maxwell color triangles for a live cell fluorescence microscopy image sequence of HCV. a) Original image sequence. b) Image sequence with decreasing intensity values for channel 1 over time.

We also evaluated the case when the intensity values of all three channels are increasing over time with a constant factor. For every time point, the original intensity values of all three channels are increased by 20 times the time point t (e.g., for $t = 9$ all intensity values of all three channels are increased by adding 180). Fig. 8.9 shows the Maxwell color triangle for the original image sequence as well as for the image sequence with increasing intensity values in all three channels. The black dots in the Maxwell color triangles for both image sequences represent the same triple colocalizations and therefore can be compared. It can be seen that the general structure of the distributions of the black dots is similar. It can also be observed that for progressing time points the black dots of the image sequence with increased intensity values tend to shift torwards the center of the Maxwell color triangle, although the intensities in all channel are increased by the same value. Below, we mathematically proof that the positions of black dots within the Maxwell color triangle (i) change for additive intensity increase (decrease) in all channels with a constant factor, (ii) generally shift torwards the center of the Maxwell color triangle for additive intensity increase (decrease) in all channels with a constant factor, and (iii) are invariant against an equal multiplicative change of the intensity values of all three channels. To proof (i), it is assumed that $R \neq G \neq B$ and that all intensities of all three channels are changed by an additive value $\beta \in \mathbb{R}$. The RGB



(a) Original image sequence



(b) Image sequence with increasing intensity values in all channels

Figure 8.9: Maxwell color triangles for a live cell fluorescence microscopy image sequence of HCV. a) Original image sequence. b) Image sequence with increasing intensity values in all channels over time.

color values within the Maxwell color triangle fullfill:

$$\frac{R}{R+G+B} + \frac{G}{R+G+B} + \frac{B}{R+G+B} = 1 \tag{8.1}$$

The position within the Maxwell color triangle for an additive intensity increase (decrease for $\beta < 0$) is not changed if the ratio for the red channel fullfills:

$$\frac{R}{R+G+B} = \frac{R+\beta}{R+\beta+G+\beta+B+\beta} \tag{8.2}$$

$$R = \frac{1}{2}(G+B) \tag{8.3}$$

Similar, the ratio for the green channel must fullfill $G = \frac{1}{2}(R+B)$ and for the blue channel applies $B = \frac{1}{2}(R+G)$. However, all three ratios are only simultaneously fullfilled if $R = G = B$, which is in contrast to the assumption $R \neq G \neq B$. Therefore, due to the additive intensity increase (or decrease if $\beta < 0$) the proportion of the three intensity values is changed and the position in the Maxwell color triangle as well. To proof (ii), it is assumed that all intensities of all three channels are changed by an additive value $\beta \in \mathbb{R}$. If β increases torwards infinity, the ratio for each channel goes torwards 1/3 (e.g., red channel):

$$\lim_{\beta->\infty} \left(\frac{R+\beta}{R+G+B+3\beta} \right) = \frac{1}{3} \tag{8.4}$$

A ratio of 1/3 for each of the three channels is represented by the center of the Maxwell color triangle. For proofing (iii), the intensity values in the original image sequence in all three channels are multiplied by a constant factor $\beta \in \mathbb{R}$, the positions of the black dots within the Maxwell color triangle are not changed since

$$\frac{R \cdot \beta}{R \cdot \beta + G \cdot \beta + B \cdot \beta} = \frac{R}{R + G + B} \tag{8.5}$$

Therefore, the Maxwell color triangle is invariant under an equal multiplicative change of the intensity values of all three channels.

Summary We proposed a new software suite consisting of *ColocQuant* and *ColocJ* to analyze the colocalization of viral structures in multi-channel fluorescence microscopy image sequences. ColocQuant and ColocJ perform object-based identification of double or triple colocalization and color composition analysis. ColocQuant uses an k-d-tree approach to determine double or triple colocalizations. ColocJ is an ImageJ macro which provides global statistics of the intensity values, object size, number of colocalizations over time, and a color composition visualization by a Maxwell color triangle (triple colocalization) or a color ribbon (double colocalization). We applied ColocQuant and ColocJ to multi-channel live cell microscopy image data of HCV associated proteins. We provide insights on the color composition of double and triple colocalizations and demonstrated the different data analysis results given by ColocJ.

8.2 Motion Analysis of HIV-1 in 3D Collagen Structures

Pathogen spread in the infected host is a complex process where multifaceted interactions depending on key parameters can lead to disease or asymptomatic host control. For intracellular pathogens such as viruses, the intrinsic host cell behavior and the local tissue environment is important and determines the pathogen spread efficiency. The global aim is to get insights into complex physiological processes involving large numbers of heterogeneous cells and tissue. Ex vivo tissue explant based on knowledge derived from in vivo analysis are used as physiological surrogate [236]. However, organotypic cultures are often hard to refactor concerning experimental manipulation of tissue organization and composition, whereas complementary efforts aim at reconstituting complex physiology by assembly of components ex vivo [237]. These novel culture systems have the advantage to study physiological processes at different levels such as the molecular, single-cell, and cell population level, but require a system level understanding of the factors involved in order to understand the contribution of individual process. In conjunction with mathematical and computational models, experimental data that are generally limited in spatial and/or

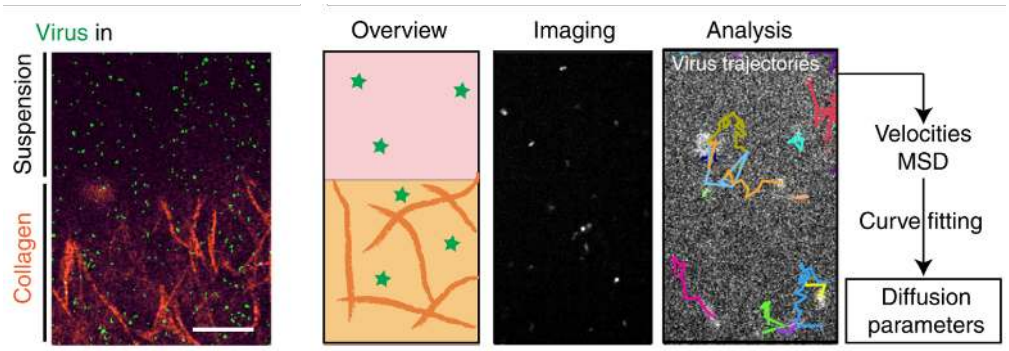


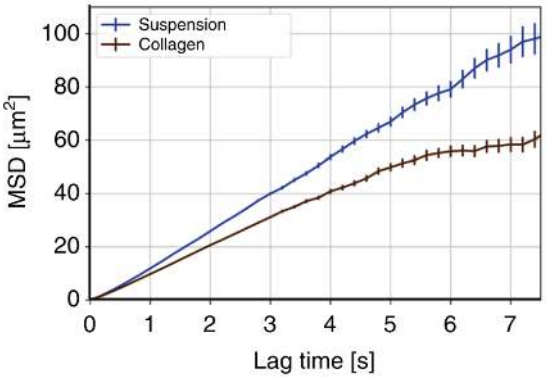
Figure 8.10: INSPECT-3D on the level of single pathogens. Fluorescent viral particles (green) were incorporated into collagen (orange) or resuspended in medium overlaying the 3D collagen matrix (suspension). Viral particles within collagen or in suspension were imaged by spinning disk confocal microscopy and tracked by probabilistic single particle tracking. Viral motility is quantified by diffusion parameters obtained from particle motion analysis (from [19]).

temporal resolution can be connected in order to quantify physiological processes within multicellular systems (e.g., cancer and immunology research [238, 239]).

For the human immunodeficiency virus type 1 (HIV-1), many aspects on the molecular level concerning the replication process in isolated target cells have been clarified including immune mechanisms of the host and viral evasion mechanisms [240]. However, little is known about the influence of the 3D environment on the replication process and how this affects the virus evasion mechanism [5].

In [19] an Integrative method to Study Pathogen spread by Experiment and Computation within Tissue-like 3D cultures (INSPECT-3D) has been proposed. An experimental component of INSPECT-3D is the system of 3D collagen matrices in order to analyze HIV-1 spread ex vivo in primary human CD4 T lymphocytes. A computational component is the quantification of HIV-1 diffusion parameters in order to characterize cell-free virus spread by single particle tracking of HIV-1 in 3D collagen structures followed by motion analysis (see Fig. 8.10).

With INSPECT-3D it was found that the kinetics of the HIV-1 spread on the population level was strongly affected by the 3D collagen environment. In order to understand the HIV-1 spread on the level of single pathogens, we developed computational approaches to find out if 3D collagen structures affect the availability of cell-free virus particles for infection. Therefore, we determined the trajectories of fluorescently labeled HIV-1 particles in medium (suspension) and in 3D collagen structures by single particle tracking of spinning disk confocal microscopy data, followed by motion analysis to obtain diffusion parameters. Automatic tracking of multiple fluorescent HIV-1 particles was performed using a probabilistic particle tracking approach which is based on Bayesian filtering and probabilistic data association [48]. This approach exploits multiple measurements and combines Kalman filtering with particle filtering. For particle detection, a Laplacian-of-Gaussian filter was used. The tracking



(a) MSD

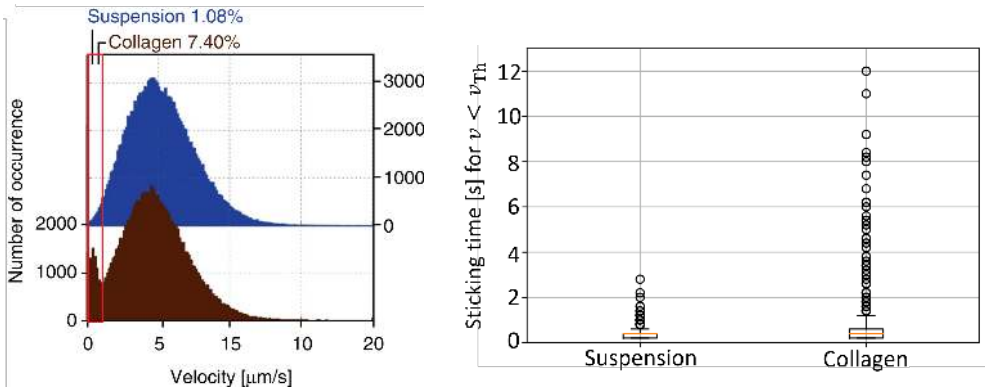
t = 0–7s	MSD = $4\Gamma t^{\alpha}$		For $\alpha = 1$ MSD = $4Dt$
		[µm² s$^{-\alpha}$]	[µm² s^{-1}]
	α	Γ	D
Suspension	1.03	3.18	3.35
Collagen	0.86	2.99	2.36

(b) Diffusion parameters

Figure 8.11: Motion analysis of HIV-1 in suspension and 3D collagen structures. a) The MSD for particles in suspension and 3D collagen structures were computed from the MSD values for at least 20,000 HIV-1 tracks (mean ± SEM). b) Diffusion parameters were computed by fitting a diffusion and an anomalous diffusion model to the MSD values (from [19]).

method yields trajectories of individual HIV-1 particles. Based on the computed trajectories, the motility of HIV-1 particles under different 3D collagen conditions and suspension was analyzed and the motion type was determined. We performed a mean square displacement (MSD) analysis [182]. MSD was computed as a function of the time interval Δt for each trajectory of a tracked HIV-1 particle with a minimum time duration of 0.8 s (corresponding to five time steps). The MSD functions for all trajectories under one condition were averaged. An anomalous diffusion model $\mathrm{MSD} = 4\,\Gamma\Delta t^{\alpha}$ was fitted to the calculated MSD values which yielded the anomalous diffusion exponent α and the transport coefficient [182]. We used MSD values from $0\,\mathrm{s} \leq \Delta t \leq 7\,\mathrm{s}$. The motion was classified into confined diffusion ($\alpha \leq 0.1$), obstructed diffusion ($0.1 < \alpha < 0.9$), and normal diffusion ($\alpha \geq 0.9$) [184]. We also fitted the normal diffusion model $\mathrm{MSD} = 4\,\Gamma\Delta t$ [182] to the MSD values to determine the diffusion coefficient $D = \Gamma\,[\mu\mathrm{m}^2\,\mathrm{s}^{-1}]$. In addition, based on the tracking result, events of particle interaction with the collagen structure were identified (sticking events). For all tracked HIV-1 particles the velocities were computed and represented in a velocity histogram. From the histogram we determined a velocity threshold of $v_{\mathrm{Th}} = 1.0\ [\mu\mathrm{m\,s}^{-1}]$ to distinguish different subpopulations. The interaction time of an HIV-1 particle with the collagen structure was calculated as the time duration for which a particle yielded velocities below v_{Th}. For a consistent comparison of interaction times between different collagen conditions, we considered trajectories with a time duration of up to 15 s (corresponding to 76 time steps).

The results for motion analysis of HIV-1 particles in suspension and 3D collagen structures are shown in Fig. 8.11. For virions in suspension, normal diffusion was observed (MSD values grow linearly with time, $\alpha = 1.03$) with a diffusion coefficient $D = 3.35\ [\mu\mathrm{m}^2\,\mathrm{s}^{-1}]$ which is comparable to the theoretical expectation of $D = 4.38$



(a) Velocity histogram (b) Sticking events

Figure 8.12: Motility analysis of tracked viral particles in suspension and 3D collagen structures. a) Distribution of instantaneous velocities obtained from tracks with durations between 0.8 and 15 s for virus in suspension (blue) or in collagen (brown). The red box indicates the percentage of tracking steps with velocities below $v_{\mathrm{Th}} = 1\,[\mu\mathrm{m\,s^{-1}}]$ (from [19]). b) Events of particle interaction with the collagen structure were identified (sticking events). The sticking time was computed as the time duration for which a particle yielded velocities $v \leq v_{\mathrm{Th}}$.

$[\mu\mathrm{m^2\,s^{-1}}]$ based on the Stokes–Einstein equation for virions of 150 nm diameter at 37 °C in water [19]. However, we observed obstructed diffusion (MSD values grow non-linearly with time, $\alpha = 0.86$) in collagen with a diffusion coefficient $D = 2.36$ $[\mu\mathrm{m^2\,s^{-1}}]$, which is typical for solute transport in porous media [241]. Due to the reduced diffusion coefficient in collagen and the cell spacing in the 3D collagen structures, a virion would take almost 1 day (22.6 h) to reach the 19 nearest neighbours by diffusion starting from a producer cell. Compared to the half-life of HIV-1 particle infectivity of 17.9 h [242], this means 58 % loss of infectivity. In addition, with single particle tracking we identified events of particle interaction with the collagen structure (sticking events) which is shown in Fig. 8.12. Within an observation time of 5 min, 7.40 % of all particles in collagen have a phase of low mobility ($v \leq v_{\mathrm{Th}}$) compared to 1.08 % of all particles in suspension. Further, it is observed in Fig. 8.12 that virions in collagen have no long-lasting collagen interactions (short sticking times) and that the phase of low virion mobility likely represents transient contact with collagen fibres. Based on the detected sticking events and their observed time duration, it can be assumed that all virions will encounter collagen fibres within their infectivity half-life of 17.9 h.

Summary We applied probabilistic particle tracking for HIV-1 in 3D collagen structures followed by motion analysis to quantify HIV-1 diffusion parameters. The results on the level of single pathogens revealed that a 3D collagen environment reduces virion diffusion rates and restricts particle infectivity as well as production. Hence, 3D environments can thus pose a significant barrier to cell-free HIV-1 infection.

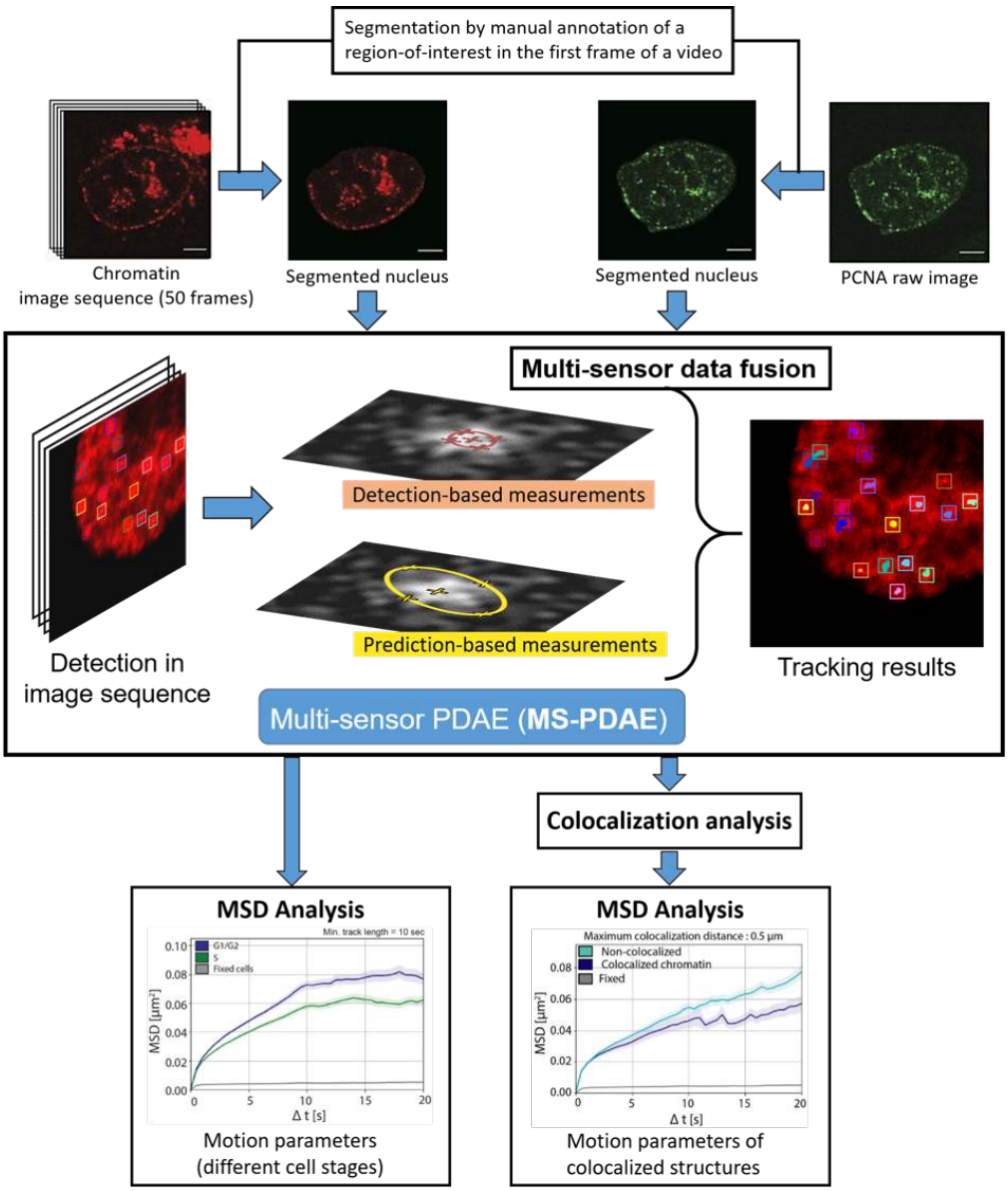


Figure 8.13: Image analysis pipeline for motion and colocalization analysis of chromatin in 2D confocal microscopy image data.

8.3 Motion and Colocalization Analysis of Chromatin in Confocal Microscopy Images

We studied chromatin motility for different cell stages and the relationship between chromatin diffusion and nuclear activity in mammalian cells. Chromatin was fluorescently labelled in 2D time-lapse confocal microscopy image data displaying HeLa cells expressing PCNA labelled with GFP. PCNA is a cell cycle marker and core component of the replication machinery [243]. To understand the organization of the genome during DNA replication in a spatial-temporal context, we used the proposed novel Multi-Sensor PDAE (MS-PDAE) (cf. Section 3.3.1) for tracking

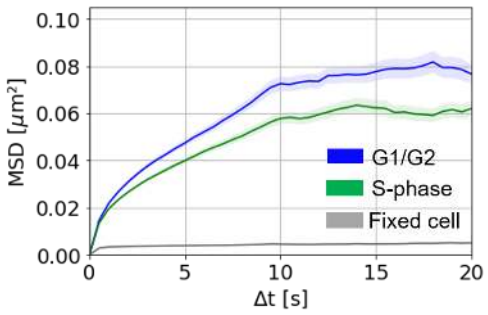
chromatin structures, motion analysis based on MSD, and colocalization analysis . The image analysis pipeline for tracking chromatin in 2D confocal microscopy image data followed by motion and colocalization analysis is depicted in Fig. 8.13.

The motility of fluorescently labelled chromatin structures in live-cell fluorescence microscopy images was quantified within manually segmented single nuclei. The background image intensity was adjusted for each image sequence to the computed mean intensity value over all time points within a manually selected region-of-interest (ROI) of the background. Automatic tracking of multiple fluorescently labelled chromatin structures was performed using a probabilistic particle tracking approach which is based on Bayesian filtering and multi-sensor data fusion [12] (cf. Section 3.3.1). This approach combines Kalman filtering with particle filtering and integrates multiple measurements by separate sensor models and sequential multi-sensor data fusion. Detection-based and prediction-based measurements are obtained by elliptical sampling [48] and the separate sensor models allow taking into account different uncertainties. In addition, motion information based on displacements from past time points is exploited and integrated within the cost function for correspondence finding. Chromatin structures are detected by the spot-enhancing filter (SEF) [106] which consists of a Laplacian-of-Gaussian (LoG) filter followed by thresholding the filtered image and determination of local maxima. The threshold is automatically determined by the mean of the absolute values of the filtered image plus a factor times the standard deviation.

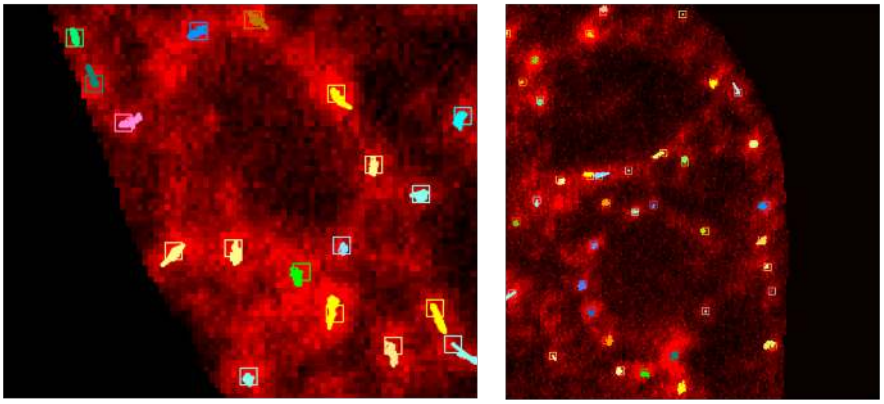
Chromatin motility analysis

Based on the computed trajectories, the motility of chromatin structures was analyzed and the motion type was determined for different experimental conditions corresponding to different cell stages. We performed a mean square displacement (MSD) analysis [182] and computed the MSD as a function of the time interval Δt for each trajectory. The MSD curves for all trajectories with a minimum time duration of 10 s (corresponding to 20 time points) under one condition were averaged. Considering trajectories with a time duration larger than the minimum time duration improved the accuracy of the motility analysis. We fitted the anomalous diffusion model $\mathrm{MSD} = 4\,\Gamma\Delta t^{\alpha}$ to the calculated MSD values to obtain the anomalous diffusion coefficient α and the transport coefficient $\Gamma\,[\mu\mathrm{m}^2\,\mathrm{s}^{-\alpha}]$. The motion was classified into confined diffusion, obstructed diffusion, and normal diffusion [184]. To determine the diffusion coefficient $D\,[\mu\mathrm{m}^2\,\mathrm{s}^{-1}]$, the normal diffusion model with $\alpha = 1$ was fitted to the MSD values.

Example chromatin tracking results for cell cycle stages G1 and S-phase as well as results for motion analysis are shown in Fig. 8.14. We found that chromatin is most motile in G1/G2 and become more constrained during S-phase which is shown by higher MSD values of G1/G2 compared to S-phase. Further, the diffusion coefficient D for chromatin in G1/G2 is higher compared to S-phase. The anomalous diffusion



	D [μm²/s x 10⁻⁵]	α
G1/G2 (1759 trajectories)	133.6	0.41
S-phase (2364 trajectories)	105.6	0.36
Fixed cell (1689 trajectories)	8.5	0.15



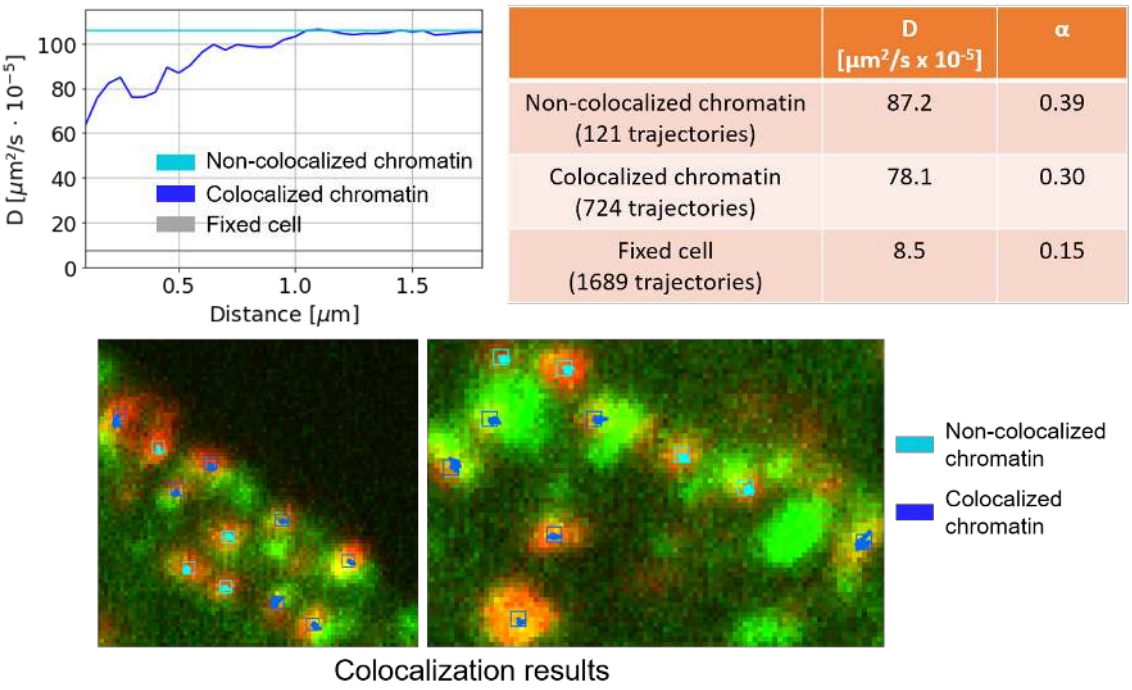
Tracking result, G1 Tracking result, S-phase

Figure 8.14: Tracking and motion analysis results for chromatin in different cell cycle stages. Based on the MSD analysis, diffusion parameters were computed.

coefficient α for chromatin in S-phase is lower compared to G1/G2 which further reflects the more constrained diffusion in S-phase.

Colocalization analysis

Automatic colocalization of chromatin and PCNA was performed using the computed trajectories of chromatin structures and detected sites of active DNA synthesis represented by fluorescently labeled PCNA. Only trajectories of chromatin structures present at the first time point of an image sequence and with a minimum time duration of 10 s (corresponding to 20 time points) were considered. PCNA foci were automatically detected in the fluorescence microscopy images by the SEF filter. For each PCNA image, a single nucleus was manually segmented and the background intensity was adjusted to the computed mean intensity value within a manually selected ROI of the background (see Fig. 8.13). Colocalization was determined for the first time point of the trajectory of a chromatin structure and detected PCNA foci using our proposed graph-based k-d-tree approach (cf. Section 8.1). Due to the k-d-tree structure, this approach allows efficient computation of the nearest neighbor query based on the Euclidean distance between foci in the chromatin and PCNA channel. If a chromatin structure at the first time point of the image sequence



	D [μm²/s x 10⁻⁵]	α
Non-colocalized chromatin (121 trajectories)	87.2	0.39
Colocalized chromatin (724 trajectories)	78.1	0.30
Fixed cell (1689 trajectories)	8.5	0.15

Figure 8.15: Results for tracking, colocalization analysis, and motion analysis. The graph for the diffusion coefficient D represents average diffusion rates of colocalized and non-colocalized chromatin with increasing colocalization distance measured between the centers of PCNA and chromatin foci.

has a nearest PCNA neighbor within a maximum distance, the trajectory of a chromatin structure is considered colocalized. Otherwise, the trajectory is considered non-colocalized.

Example colocalization results of chromatin with PCNA are shown in Fig. 8.15. Further, the diffusion coefficient D and the anomalous diffusion coefficient α for colocalized and non-colocalized chromatin is given. It turned out, that the chromatin motility is locally restricted at sites of DNA replication (colocalization of chromatin and PCNA).

Summary Based on our proposed novel Multi-Sensor PDAE (MS-PDAE) for chromatin tracking in conjunction with motion and colocalization analysis, we obtained novel insights into the spatio-temporal organization of the genome during DNA replication for mammalian systems. We demonstrated that chromatin undergoes different motility for different cell cycle stages. Further, we showed that during active replication the chromatin motion is more restricted in order to facilitate the cellular processes.

9 Summary and Outlook

In this thesis, new probabilistic methods for particle tracking in fluorescence microscopy images were introduced. The proposed methods are based on data fusion and Bayesian smoothing to determine trajectories with higher accuracy. The methods have been quantitatively evaluated with state-of-the-art benchmark data sets as well as real fluorescence microscopy images. Improved results for particle detection, localization, and tracking compared to existing methods were achieved. Besides particle detection and tracking, we also performed colocalization and motion analysis of HCV proteins, HIV-1 particles, and chromatin structures in real fluorescence microscopy image data to support biologists in image analysis for infectious diseases and genome organization research projects.

9.1 Summary

The main contributions can be summarized as follows,

Data Fusion for Particle Detection and Tracking

The main challenges for particle detection and tracking in fluorescence microscopy image data are low signal-to-noise ratio (SNR), small particle size, heterogeneity in particle size, high object density, complex motion, and clutter. To cope with these different challenges, we developed novel particle detection and tracking methods which are based on data fusion approaches to exploit a priori knowledge. We developed a multi-detector data fusion approach based on a novel intensity-based covariance intersection algorithm which exploits information about image intensities, positions, and uncertainties obtained by multiple particle detectors to improve detection and localization. An advantage of the intensity-based covariance intersection algorithm is that it ensures a consistent estimate of multiple fused particle detections and does not require an optimization step. In addition, for each particle a time-varying estimate of the measurement noise covariance is computed to improve update estimation. We performed data fusion of detections from classical and deep learning detectors as well as single-scale and multi-scale particle detectors. Further, we proposed a novel multi-sensor data fusion approach with probabilistic data association for particle tracking which fuses multiple measurements from separate measurement processes and thus takes into account different uncertainties from particle detection and particle prediction to improve trajectory estimation.

Bayesian Smoothing for Particle Tracking

We developed particle tracking methods which exploit temporal information based on Bayesian smoothing approaches. First, we proposed a two-filter Bayesian smoothing approach with probabilistic data association which fuses information from past and future time points, integrates multiple measurements, and combines Kalman filtering and particle filtering. Building on this approach, we developed a particle tracking method based on Bayesian smoothing which uses covariance intersection for fusing predictions obtained from past and future time points. An advantage is that this data fusion approach ensures a consistent estimate of the fused prediction. Smoothing and covariance intersection yield improved state prediction and track initialization. Further, motion information is exploited by fusing displacements from past and future time points and integrated in the cost function for improved correspondence finding. The method is further extended by the proposed novel multi-detector data fusion approach based on our intensity-based covariance intersection algorithm to improve particle detection.

Colocalization and Motion Analysis

To support biologist in image analysis, we developed a colocalization analysis approach which determines spatial relationships of viral particles in multi-channel fluorescence microscopy images by a multi-dimensional graph-based approach. The colocalization analysis approach is integrated in a software suite and visualizes the color composition for colocalized objects by a Maxwell color triangle. To study pathogen spread in tissue-like 3D cultures, we applied probabilistic tracking in conjunction with motion analysis to real fluorescence microscopy image data of HIV-1 particles. Diffusion parameters were quantified and revealed that HIV-1 has a reduced diffusion rate due to the 3D environment and restricted infectivity. We also studied chromatin motility and investigated the relationship between chromatin diffusion and nuclear activity in mammalian cells. We applied colocalization and motion analysis to quantify diffusion parameters and correlated chromatin motility to sites of DNA replication. We demonstrated that chromatin motion is reduced at active DNA replication sites.

Other Contributions We developed a novel approach which combines a convolutional neural network for particle detection with probabilistic data association for tracking. Furthermore, we developed a deep learning-based approach for both particle detection and tracking by combining a convolutional neural network for detection with an advanced recurrent neural network for tracking. To automatically optimize complex image analysis pipelines containing several tunable hyperparameters, we described a hyperparameter optimization framework with a modular architecture and evaluated the framework on live cell fluorescence microscopy data for HCV particle detection. Further, we developed a novel method for hyperparameter optimization for particle detection in multi-channel microscopy images which exploits colocalization in different channels as a surrogate for ground truth. The approach was evaluated on real multi-channel fluorescence microscopy images displaying double-labeled HIV-1 particles. We also developed a novel probabilistic cell tracking approach based on multiple hypothesis tracking with integrated cell division detection. Cell division is integrated by remodelling the graph of the maximum weighted independent set problem to resolve incompatibility between certain track hypotheses.

9.2 Outlook

This section provides research questions which could be addressed in future work.

- The proposed Bayesian smoothing approaches for particle tracking as well as the multi-detector fusion approach for particle detection are based on data fusion using covariance intersection. To obtain less conservative fusion results, other data fusion techniques could be investigated. However, consistency of the fusion results might not be guaranteed as in our methods.

- The particle tracking approaches based on data fusion and Bayesian smoothing exploit motion information for particle prediction as well as correspondence finding. The approaches have shown improved results compared to previous methods. To further increase the tracking performance multiple motion models could be incorporated in the tracking approaches.

- For double-labeled virus particles, colocalization information is exploited to perform weakly supervised particle detection. Incorporating information from more than two microscopy channels using multi-labeled viruses, particle detection might be improved which could be beneficial for a subsequent particle tracking in multi-channel fluorescence microscopy image sequences.

- Motion analysis of HIV-1 particles in real fluorescence microscopy image sequences was performed. Diffusion parameters have been quantified to characterize virus spread on the level of single pathogens. Learning-based approaches for motion analysis could be investigated.

Bibliography

[1] World Health Organization, "HIV-AIDS fact sheet." `https://www.who.int/news-room/fact-sheets/detail/hiv-aids`.

[2] V. Zila, E. Margiotta, B. Turoňová, T. G. Müller, C. E. Zimmerli, S. Mattei, M. Allegretti, K. Börner, J. Rada, B. Müller, *et al.*, "Cone-shaped HIV-1 capsids are transported through intact nuclear pores," *Cell*, vol. 184, no. 4, pp. 1032–1046.e18, 2021.

[3] A. N. Engelman and P. K. Singh, "Cellular and molecular mechanisms of HIV-1 integration targeting," *Cell. Mol. Life Sci.*, vol. 75, pp. 2491–2507, 2018.

[4] M. Stevenson, "HIV-1 pathogenesis," *Nat. Med.*, vol. 9, no. 7, pp. 853–860, 2003.

[5] O. T. Fackler, T. T. Murooka, A. Imle, and T. R. Mempel, "Adding new dimensions: towards an integrative understanding of HIV-1 spread," *Nat. Rev. Microbiol.*, vol. 12, no. 8, pp. 563–574, 2014.

[6] World Health Organization, "HIV-AIDS fact sheet." `https://www.who.int/news-room/fact-sheets/detail/hepatitis-c`.

[7] S. M. Feinstone, A. Z. Kapikian, R. H. Purcell, H. J. Alter, and P. V. Holland, "Transfusion-associated hepatitis not due to viral hepatitis type A or B," *New Eng. J. Med.*, vol. 292, pp. 767–770, 1975.

[8] J. Dubuisson and F.-L. Cosset, "Virology and cell biology of the hepatitis C virus life cycle - An update," *J. Hepatol.*, vol. 61, no. 1, pp. S3–S13, 2014.

[9] B. D. Lindenbach and C. M. Rice, "Unravelling hepatitis C virus replication from genome to function," *Nature*, vol. 436, no. 7053, pp. 933–938, 2005.

[10] T. G. Müller, V. Zila, K. Peters, S. Schifferdecker, M. Stanic, B. Lucic, V. Laketa, M. Lusic, B. Müller, and H.-G. Kräusslich, "HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells," *eLife*, vol. 10, e64776, 2021.

[11] C. Ritter, A. Imle, J.-Y. Lee, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "Two-filter probabilistic data association for tracking of virus particles in fluorescence microscopy images," in *Proc. ISBI*, pp. 957–960, IEEE, 2018.

[12] C. Ritter, T. Wollmann, J.-Y. Lee, A. Imle, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "Data fusion and smoothing for probabilistic tracking of viral structures in fluorescence microscopy images," *Med. Image Anal.*, vol. 73, 102168, 2021.

[13] C. Ritter, J.-Y. Lee, M.-T. Pham, M. K. Pabba, M. C. Cardoso, R. Bartenschlager, and K. Rohr, "Multi-detector fusion and Bayesian smoothing for tracking viral and chromatin structures," *Under review*, 2022.

[14] T. Wollmann, C. Ritter, J.-N. Dohrke, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "DetNet: Deep neural network for particle detection in fluorescence microscopy images," in *Proc. ISBI*, pp. 517–520, IEEE, 2019.

[15] C. Ritter, T. Wollmann, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "Deep learning particle detection for probabilistic tracking in fluorescence microscopy images," in *Proc. ISBI*, pp. 977–980, IEEE, 2020.

[16] C. Ritter, R. Spilger, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "Deep learning for particle detection and tracking in fluorescence microscopy images," in *Proc. ISBI*, pp. 873–876, IEEE, 2021.

[17] C. Ritter, R. Thielemann, J.-Y. Lee, M.-T. Pham, R. Bartenschlager, and K. Rohr, "ColocQuant and ColocJ: Multi-channel colocalization analysis of viral proteins in fluorescence microscopy images," *In preparation*, 2022.

[18] J.-Y. Lee, M. Cortese, U. Haselmann, K. Tabata, I. Romero-Brey, C. Funaya, N. L. Schieber, Y. Qiang, M. Bartenschlager, S. Kallis, C. Ritter, K. Rohr, Y. Schwab, A. Ruggieri, and R. Bartenschlager, "Spatiotemporal coupling of the hepatitis C virus replication cycle by creating a lipid droplet-proximal membranous replication compartment," *Cell Rep.*, vol. 27, no. 12, pp. 3602–3617.e5, 2019.

[19] A. Imle, P. Kumberger, N. D. Schnellbächer, J. Fehr, P. Carrillo-Bustamante, J. Ales, P. Schmidt, C. Ritter, W. J. Godinez, B. Müller, K. Rohr, F. A. Hamprecht, U. S. Schwarz, F. Graw, and O. T. Fackler, "Experimental and computational analyses reveal that environmental restrictions shape HIV-1 spread in 3D cultures," *Nat. Commun.*, vol. 10, no. 2144, pp. 1–18, 2019.

[20] M.-T. Pham, J.-Y. Lee, C. Ritter, R. Thielemann, U. Haselmann, C. Funaya, V. Laketa, K. Rohr, and R. Bartenschlager, "Intercellular transmission of viral RNA by Apolipoprotein E associated extracellular vesicles," *Under review*, 2022.

[21] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nat. Methods*, vol. 2, no. 12, pp. 910–919, 2005.

[22] M. Mulisch and U. Welsch, *Romeis-Mikroskopische Technik*. Springer-Verlag, 2015.

[23] S.-L. Liu, Z.-G. Wang, H.-Y. Xie, A.-A. Liu, D. C. Lamb, and D.-W. Pang, "Single-virus tracking: From imaging methodologies to virological applications," *Chem. Rev.*, vol. 120, no. 3, pp. 1936–1979, 2020.

[24] M. Lampe, J. A. 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.

[25] 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, pp. 1–12, 2009.

[26] R. Gräf, J. Rietdorf, and T. Zimmermann, "Live cell spinning disk microscopy," in *Microscopy Techniques*, pp. 57–75, Springer-Verlag, 2005.

[27] I. K. Stoeck, J.-Y. Lee, K. Tabata, I. Romero-Brey, D. Paul, P. Schult, V. Lohmann, L. Kaderali, and R. Bartenschlager, "Hepatitis C virus replication depends on endosomal cholesterol homeostasis," *J. Virol.*, vol. 92, no. 1, pp. 1–26, 2018.

[28] X. Michalet, O. Siegmund, J. Vallerga, P. Jelinsky, J. Millaud, and S. Weiss, "Detectors for single-molecule fluorescence imaging and spectroscopy," *J. Mod. Opt.*, vol. 54, no. 2-3, pp. 239–281, 2007.

[29] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes, "Single particle tracking: From theory to biophysical applications," *Chem. Rev.*, vol. 117, no. 11, pp. 7331–7376, 2017.

[30] W. Burger and M. J. Burge, *Digital image processing: An algorithmic introduction using Java*. Springer-Verlag, 2016.

[31] H. T. Beier and B. L. Ibey, "Experimental comparison of the high-speed imaging performance of an EM-CCD and sCMOS camera in a dynamic live-cell imaging test case," *PLoS ONE*, vol. 9, no. 1, e84614, 2014.

[32] M. Schelhaas, H. Ewers, M.-L. Rajamäki, P. M. Day, J. T. Schiller, and A. Helenius, "Human papillomavirus type 16 entry: Retrograde cell surface transport along actin-rich protrusions," *PLoS Pathogens*, vol. 4, no. 9, e1000148, 2008.

[33] D. McDonald, M. A. Vodicka, G. Lucero, T. M. Svitkina, G. G. Borisy, M. Emerman, and T. J. Hope, "Visualization of the intracellular behavior of HIV in living cells," *J. Cell Biol.*, vol. 159, no. 3, pp. 441–452, 2002.

[34] 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.

[35] K. E. Coller, K. L. Berger, N. S. Heaton, J. D. Cooper, R. Yoon, and G. Randall, "RNA interference and single particle tracking analysis of hepatitis C virus endocytosis," *PLoS Pathogens*, vol. 5, no. 12, e1000702, 2009.

[36] N. Jouvenet, P. D. Bieniasz, and S. M. Simon, "Imaging the biogenesis of individual HIV-1 virions in live cells," *Nature*, vol. 454, pp. 236–240, 2008.

[37] M. Monici, "Cell and tissue autofluorescence research and diagnostic applications," *Biotechnol. Annu. Rev.*, vol. 11, pp. 227–256, 2005.

[38] L. Wald, "Some terms of reference in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, 1999.

[39] M. E. Liggins, D. L. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice.* CRC Press, 2009.

[40] H. B. Mitchell, *Data fusion: Concepts and Ideas.* Springer-Verlag, 2012.

[41] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, pp. 1–19, 2013.

[42] K. R. Fowler and J. L. Schmalzel, "Sensors: The first stage in the measurement chain," *IEEE Instrum. Meas. Mag.*, vol. 7, no. 3, pp. 60–65, 2004.

[43] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, "Decentralized data fusion with inverse covariance intersection," *Automatica*, vol. 79, pp. 35–41, 2017.

[44] C.-Y. Chong, S. Mori, F. Govaers, and W. Koch, "Comparison of tracklet fusion and distributed Kalman filter for track fusion," in *Proc. Int. Conf. Inf. Fusion*, pp. 1–8, IEEE, 2014.

[45] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques.* YBS Publishing, 1995.

[46] 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. Imag.*, vol. 27, no. 6, pp. 789–804, 2008.

[47] 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.

[48] W. J. Godinez and K. Rohr, "Tracking multiple particles in fluorescence time-lapse microscopy images via probabilistic data association," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 415–432, 2015.

[49] P. Roudot, L. Ding, K. Jaqaman, C. Kervrann, and G. Danuser, "Piecewise-stationary motion modeling and iterative smoothing to track heterogeneous particle motions in dense environments," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5395–5410, 2017.

[50] J. Ajgl, M. Šimandl, and J. Dunik, "Millman's formula in data fusion," in *Proc. Int. Workshop Syst. Contr.*, pp. 1–6, 2009.

[51] M. Reinhardt, B. Noack, P. O. Arambel, and U. D. Hanebeck, "Minimum covariance bounds for the fusion under unknown correlations," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1210–1214, 2015.

[52] L. Chen, P. O. Arambel, and R. K. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," *IEEE Trans. Automat. Contr.*, vol. 47, no. 11, pp. 1879–1882, 2002.

[53] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 803–805, 1986.

[54] B. Noack, J. Sijs, and U. D. Hanebeck, "Inverse covariance intersection: New insights and properties," in *Proc. Int. Conf. Inf. Fusion*, pp. 1–8, IEEE, 2017.

[55] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. Am. Control Conf.*, pp. 2369–2373, IEEE, 1997.

[56] W. Kahan, "Circumscribing an ellipsoid about the intersection of two ellipsoids," *Can. Math. Bull.*, vol. 11, no. 3, pp. 437–441, 1968.

[57] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *Proc. Int. Conf. Inf. Fusion*, pp. 901–904, IEEE, 2002.

[58] D. Franken and A. Hupper, "Improved fast covariance intersection for distributed data fusion," in *Proc. Int. Conf. Inf. Fusion*, vol. 1, pp. 154–160, IEEE, 2005.

[59] Y. Wang and X. R. Li, "Distributed estimation fusion under unknown cross-correlation: An analytic center approach," in *Proc. Int. Conf. Inf. Fusion*, pp. 1–8, IEEE, 2010.

[60] M. Reinhardt, B. Noack, and U. D. Hanebeck, "Closed-form optimization of covariance intersection for low-dimensional matrices," in *Proc. Int. Conf. Inf. Fusion*, pp. 1891–1896, IEEE, 2012.

[61] J. Sijs, M. Lazar, and P. Bosch, "State fusion with unknown correlation: Ellipsoidal intersection," in *Proc. Am. Control Conf.*, pp. 3992–3997, IEEE, 2010.

[62] S. Grime and H. F. Durrant-Whyte, "Data fusion in decentralized sensor networks," *Control Eng. Pract.*, vol. 2, no. 5, pp. 849–863, 1994.

[63] B. Noack, J. Sijs, and U. D. Hanebeck, "Algebraic analysis of data fusion with ellipsoidal intersection," in *Proc. MFI*, pp. 365–370, IEEE, 2016.

[64] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[65] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[66] M. S. Grewal and A. P. Andrews, *Kalman filtering: Theory and practice using MATLAB*. John Wiley & Sons, 2008.

[67] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[68] S. Haykin, *Kalman filtering and neural networks*, vol. 47. John Wiley & Sons, 2004.

[69] G. Kitagawa, "Non-gaussian state—space modeling of nonstationary time series," *J. Am. Stat. Assoc.*, vol. 82, no. 400, pp. 1032–1041, 1987.

[70] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[71] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Automat. Contr.*, vol. 14, no. 4, pp. 387–390, 1969.

[72] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[73] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[74] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[75] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, and Q. Liao, "Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review," *Int. J. Autom. Comput.*, vol. 14, no. 5, pp. 503–519, 2017.

[76] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, pp. 234–241, Springer-Verlag, 2015.

[77] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. ICCV*, pp. 1520–1528, IEEE, 2015.

[78] J. Yi, P. Wu, D. J. Hoeppner, and D. Metaxas, "Pixel-wise neural cell instance segmentation," in *Proc. ISBI*, pp. 373–377, IEEE, 2018.

[79] T. Wollmann, M. Gunkel, I. Chung, H. Erfle, K. Rippe, and K. Rohr, "GRUU-Net: Integrated convolutional and gated recurrent neural network for cell segmentation," *Med. Image Anal.*, vol. 56, pp. 68–79, 2019.

[80] R. Girshick, "Fast R-CNN," in *Proc. CVPR*, pp. 1440–1448, IEEE, 2015.

[81] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NeurIPS*, pp. 91–99, 2015.

[82] M. A. Mabaso, D. J. Withey, and B. Twala, "Spot detection in microscopy images using convolutional neural network with sliding-window approach," in *Proc. Bioimaging*, pp. 67–74, SciTePress, 2018.

[83] J. M. Newby, A. M. Schaefer, P. T. Lee, M. G. Forest, and S. K. Lai, "Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D," *PNAS*, vol. 115, no. 36, pp. 9026–9031, 2018.

[84] P. R. Gudla, K. Nakayama, G. Pegoraro, and T. Misteli, "SpotLearn: Convolutional neural network for detection of fluorescence in situ hybridization (FISH) signals in high-throughput imaging approaches," in *Proc. CSH Symposia on Quant. Biol. 2017*, pp. 57–70, CSH Laboratory Press, 2017.

[85] T. Wollmann and K. Rohr, "Deep Consensus Network: Aggregating predictions to improve object detection in microscopy images," *Med. Image Anal.*, vol. 70, 102019, 2021.

[86] W. J. Godinez, I. Hossain, S. E. Lazic, J. W. Davies, and X. Zhang, "A multi-scale convolutional neural network for phenotyping high-content cellular images," *Bioinformatics*, vol. 33, no. 13, pp. 2010–2019, 2017.

[87] T. Pärnamaa and L. Parts, "Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning," *G3*, vol. 7, no. 5, pp. 1385–1392, 2017.

[88] B. Midtvedt, S. Helgadottir, A. Argun, J. Pineda, D. Midtvedt, and G. Volpe, "Quantitative digital microscopy with deep learning," *Appl. Phys. Rev.*, vol. 8, pp. 1–22, 2021.

[89] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016.

[90] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. CVPR*, pp. 3431–3440, IEEE, 2015.

[91] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, e3, 2016.

[92] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 3DV*, pp. 565–571, IEEE, 2016.

[93] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Stat. Plan. Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[94] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, pp. 448–456, PMLR, 2015.

[95] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," in *arXiv:1607.08022*, 2016.

[96] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen netzen," *Diploma thesis, Technical University of Munich*, 1991.

[97] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, pp. 770–778, IEEE, 2016.

[98] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, 1994.

[99] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[100] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *arXiv:1406.1078*, 2014.

[101] I. J. Cox, "A review of statistical data association techniques for motion correspondence," *Int. J. Comput. Vis.*, vol. 10, no. 1, pp. 53–66, 1993.

[102] 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.

[103] 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.

[104] 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.

[105] 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, no. 9, pp. 1384–1395, 2005.

[106] 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. 1372–1383, 2005.

[107] 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. ISBI*, pp. 1158–1161, IEEE, 2009.

[108] 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, no. 248, pp. 1–17, 2010.

[109] I. Smal, M. Loog, W. Niessen, and E. Meijering, "Quantitative comparison of spot detection methods in fluorescence microscopy," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 282–301, 2010.

[110] K. Štěpka, P. Matula, P. Matula, S. Wörz, K. Rohr, and M. Kozubek, "Performance and sensitivity evaluation of 3D spot detection methods in confocal microscopy," *Cytometry Part A*, vol. 87, no. 8, pp. 759–772, 2015.

[111] 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.

[112] E. Breen, G. Joss, and K. Williams, "Locating objects of interest within biological images: The top hat box filter," *J. Comput. Assist. Microsc.*, vol. 3, pp. 97–102, 1991.

[113] 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, no. 1, pp. 49–64, 2002.

[114] 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. ISBI*, pp. 264–267, IEEE, 2008.

[115] K. Pan, A. Kokaram, J. Hillebrand, and M. Ramaswami, "Gaussian mixture models for spots in microscopy using a new split/merge EM algorithm," in *Proc. ISBI*, pp. 3645–3648, IEEE, 2010.

[116] S. H. Rezatofighi, R. Hartley, and W. E. Hughes, "A new approach for spot detection in total internal reflection fluorescence microscopy," in *Proc. ISBI*, pp. 860–863, IEEE, 2012.

[117] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, 1993.

[118] J.-C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognit.*, vol. 35, no. 9, pp. 1989–1996, 2002.

[119] A. Basset, J. Boulanger, J. Salamero, P. Bouthemy, and C. Kervrann, "Adaptive spot detection with optimal scale selection in fluorescence microscopy images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4512–4527, 2015.

[120] A. Jaiswal, W. J. Godinez, R. Eils, M. J. Lehmann, and K. Rohr, "Tracking virus particles in fluorescence microscopy images using multi-scale detection and multi-frame association," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4122–4136, 2015.

[121] I. Smal and E. Meijering, "Quantitative comparison of multiframe data association techniques for particle tracking in time-lapse fluorescence microscopy," *Med. Image Anal.*, vol. 24, no. 1, pp. 163–189, 2015.

[122] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logist. Q.*, vol. 2, pp. 83–97, 1955.

[123] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.

[124] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.

[125] 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. on Image Process.*, vol. 15, no. 5, pp. 1062–1070, 2006.

[126] M. Dmitrieva, H. L. Zenner, J. Richens, D. St Johnston, and J. Rittscher, "Protein tracking by CNN-based candidate pruning and two-step linking with Bayesian network," in *Proc. MLSP*, pp. 1–6, IEEE, 2019.

[127] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Syst. Mag.*, vol. 29, no. 6, pp. 82–100, 2009.

[128] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation.* John Wiley & Sons, Inc., 2001.

[129] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffé, "Multi-target tracking using joint probabilistic data association," in *Proc. Conf. Dec. Contr.*, pp. 807–812, IEEE, 1980.

[130] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint probabilistic data association revisited," in *Proc. ICCV*, pp. 3047–3055, 2015.

[131] S. Coraluppi and C. Carthel, "Multi-stage multiple-hypothesis tracking.," *J. Adv. Inf. Fusion.*, vol. 6, no. 1, pp. 57–67, 2011.

[132] N. Chenouard, I. Bloch, and J.-C. Olivo-Marin, "Multiple hypothesis tracking for cluttered biological image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2736–2750, 2013.

[133] L. Liang, H. Shen, P. De Camilli, and J. S. Duncan, "A novel multiple hypothesis based particle tracking method for clathrin mediated endocytosis analysis using fluorescence microscopy," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1844–1857, 2014.

[134] L.-L. S. Ong, M. H. Ang, and H. H. Asada, "Tracking of cell population from time lapse and end point confocal microscopy images with multiple hypothesis Kalman smoothing filters," in *Proc. CVPR*, pp. 71–78, IEEE, 2010.

[135] S. Coraluppi, C. Carthel, S. J. Dickerson, D. Chiarulli, and S. Levitan, "Feature-aided multiple-hypothesis tracking and classification of biological cells," in *Proc. Int. Conf. Inf. Fusion*, pp. 1–8, IEEE, 2014.

[136] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Automat. Contr.*, vol. 24, no. 6, pp. 843–854, 1979.

[137] C. Morefield, "Application of 0-1 integer programming to a track assembly problem," in *Proc. Conf. Dec. Contr.*, pp. 428–433, IEEE, 1975.

[138] Y. Bar-Shalom, ed., *Multitarget-Multisensor Tracking: Advanced Applications.* Artech House, 1990.

[139] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *Proc. ICCV*, pp. 4696–4704, IEEE, 2015.

[140] D. J. Papageorgiou and M. R. Salpukas, "The maximum weight independent set problem for data association in multiple hypothesis tracking," in *Opt. Coop. Contr. Strat.*, pp. 235–255, Springer-Verlag, 2009.

[141] Y. Yao, I. Smal, and E. Meijering, "Deep neural networks for data association in particle tracking," in *Proc. ISBI*, pp. 458–461, IEEE, 2018.

[142] R. Spilger, T. Wollmann, Y. Qiang, A. Imle, J.-Y. Lee, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "Deep particle tracker: Automatic tracking of particles in fluorescence microscopy images using deep learning," in *Proc. MICCAI Workshop DLMIA*, pp. 128–136, Springer-Verlag, 2018.

[143] I. Smal, Y. Yao, N. Galjart, and E. Meijering, "Facilitating data association in particle tracking using autoencoding and score matching," in *Proc. ISBI*, pp. 1523–1526, IEEE, 2019.

[144] Y. Yao, I. Smal, I. Grigoriev, A. Akhmanova, and E. Meijering, "Deep learning method for data association in particle tracking," *Bioinformatics*, vol. btaa597, pp. 1–7, 2020.

[145] R. Spilger, A. Imle, J.-Y. Lee, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr, "A recurrent neural network for particle tracking in microscopy images using future information, track hypotheses, and multiple detections," *IEEE Trans. Image Process.*, vol. 29, pp. 3681–3694, 2020.

[146] R. Spilger, J.-Y. Lee, V. O. Chagin, L. Schermelleh, M. C. Cardoso, R. Bartenschlager, and K. Rohr, "Deep probabilistic tracking of particles in fluorescence microscopy images," *Med. Image Anal.*, vol. 72, pp. 1–18, 2021.

[147] A. Milan, S. H. Rezatofighi, A. R. Dick, I. D. Reid, and K. Schindler, "Online multi-target tracking using recurrent neural networks," in *Proc. AAAI*, pp. 4225–4232, 2017.

[148] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies," in *Proc. ICCV*, pp. 300–311, IEEE, 2017.

[149] K.-S. Chang, J. Jiang, Z. Cai, and G. Luo, "Human Apolipoprotein E is required for infectivity and production of hepatitis C virus in cell culture," *J. Virol.*, vol. 81, no. 24, pp. 13783–13793, 2007.

[150] L. P. Nguyen, T. T. Nguyen, H. C. Nguyen, H. T. Pham, K. M. Han, D. H. Choi, E.-M. Park, S. M. Kang, D. Tark, Y.-S. Lim, *et al.*, "Cortactin interacts with hepatitis C virus core and NS5A proteins: Implications for virion assembly," *J. Virol.*, vol. 94, no. 19, pp. 1–18, 2020.

[151] S. Bolte and F. P. Cordelières, "A guided tour into subcellular colocalization analysis in light microscopy," *J. Microsc.*, vol. 224, no. 3, pp. 213–232, 2006.

[152] K. W. Dunn, M. M. Kamocka, and J. H. McDonald, "A practical guide to evaluating colocalization in biological microscopy," *Am. J. Physiol. Cell Physiol.*, vol. 300, no. 4, pp. C723–C742, 2011.

[153] T. Lagache, N. Sauvonnet, L. Danglot, and J.-C. Olivo-Marin, "Statistical analysis of molecule colocalization in bioimaging," *Cytometry Part A*, vol. 87, no. 6, pp. 568–579, 2015.

[154] E. Manders, J. Stap, G. Brakenhoff, R. Van Driel, and J. Aten, "Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of dna and confocal microscopy," *J. Cell Sci.*, vol. 103, no. 3, pp. 857–862, 1992.

[155] E. Manders, F. Verbeek, and J. Aten, "Measurement of co-localization of objects in dual-colour confocal images," *J. Microsc*, vol. 169, no. 3, pp. 375–382, 1993.

[156] B. van Steensel, E. P. van Binnendijk, C. D. Hornsby, H. Van der Voort, Z. S. Krozowski, E. R. de Kloet, and R. van Driel, "Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons," *J. Cell Sci.*, vol. 109, no. 4, pp. 787–792, 1996.

[157] S. V. Costes, D. Daelemans, E. H. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett, "Automatic and quantitative measurement of protein-protein colocalization in live cells," *Biophys. J.*, vol. 86, no. 6, pp. 3993–4003, 2004.

[158] Q. Li, A. Lau, T. J. Morris, L. Guo, C. B. Fordyce, and E. F. Stanley, "A syntaxin 1, Gα_0, and N-type calcium channel complex at a presynaptic nerve terminal: Analysis by quantitative immunocolocalization," *J. Neurosci.*, vol. 24, no. 16, pp. 4070–4081, 2004.

[159] H. Sheng, W. Stauffer, and H. N. Lim, "Systematic and general method for quantifying localization in microscopy images," *Biol. Open*, vol. 5, no. 12, pp. 1882–1893, 2016.

[160] E. Lachmanovich, D. Shvartsman, Y. Malka, C. Botvin, Y. Henis, and A. Weiss, "Co-localization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: application to immunofluorescence co-patching studies," *J. Microsc.*, vol. 212, no. 2, pp. 122–131, 2003.

[161] F. Jaskolski, C. Mulle, and O. J. Manzoni, "An automated method to quantify and visualize colocalized fluorescent signals," *J. Neurosci.*, vol. 146, no. 1, pp. 42–49, 2005.

[162] Y. Boutté, M.-T. Crosnier, N. Carraro, J. Traas, and B. Satiat-Jeunemaitre, "The plasma membrane recycling pathway and cell polarity in plants: studies on PIN proteins," *J. Cell Sci.*, vol. 119, no. 7, pp. 1255–1265, 2006.

[163] B. Zhang, N. Chenouard, J.-C. Olivo-Marin, and V. Meas-Yedid, "Statistical colocalization in biological imaging with false discovery control," in *Proc. ISBI*, pp. 1327–1330, IEEE, 2008.

[164] J. A. Helmuth, G. Paul, and I. F. Sbalzarini, "Beyond co-localization: Inferring spatial interactions between sub-cellular structures from microscopy images," *BMC Bioinf.*, vol. 11, no. 372, pp. 1–12, 2010.

[165] S. Wörz, P. Sander, M. Pfannmoller, R. J. 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.

[166] T. Lagache, V. Meas-Yedid, and J.-C. Olivo-Marin, "A statistical analysis of spatial colocalization using Ripley's K function," in *Proc. ISBI*, pp. 896–901, IEEE, 2013.

[167] A. Dupont, K. Stirnnagel, D. Lindemann, and D. Lamb, "Tracking image correlation: Combining single-particle tracking and image correlation," *Biophys. J.*, vol. 104, no. 11, pp. 2373–2382, 2013.

[168] D. Vercauteren, H. Deschout, K. Remaut, J. F. Engbersen, A. T. Jones, J. Demeester, S. C. De Smedt, and K. Braeckmans, "Dynamic colocalization microscopy to characterize intracellular trafficking of nanomedicines," *ACS Nano*, vol. 5, no. 10, pp. 7874–7884, 2011.

[169] H. Deschout, T. Martens, D. Vercauteren, K. Remaut, J. Demeester, S. C. De Smedt, K. Neyts, and K. Braeckmans, "Correlation of dual colour single particle trajectories for improved detection and analysis of interactions in living cells," *Int. J. Mol. Sci.*, vol. 14, no. 8, pp. 16485–16514, 2013.

[170] Y. Qiang, J.-Y. Lee, R. Bartenschlager, and K. Rohr, "Colocalization analysis and particle tracking in multi-channel fluorescence microscopy images," in *Proc. ISBI*, pp. 646–649, IEEE, 2017.

[171] A. Rizk, G. Paul, P. Incardona, M. Bugarski, M. Mansouri, A. Niemann, U. Ziegler, P. Berger, and I. F. Sbalzarini, "Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh," *Nat. Protoc.*, vol. 9, no. 3, pp. 586–596, 2014.

[172] J.-F. Gilles, M. Dos Santos, T. Boudier, S. Bolte, and N. Heck, "DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis," *Methods*, vol. 115, pp. 55–64, 2017.

[173] W. Stauffer, H. Sheng, and H. N. Lim, "EzColocalization: An ImageJ plugin for visualizing and measuring colocalization in cells and organisms," *Sci. Rep.*, vol. 8, pp. 1–13, 2018.

[174] A. Gorlewicz, K. Krawczyk, A. A. Szczepankiewicz, P. Trzaskoma, C. Mulle, and G. M. Wilczynski, "Colocalization Colormap -an ImageJ plugin for the quantification and visualization of colocalized signals," *Neuroinformatics*, vol. 18, pp. 661–664, 2020.

[175] T. J. Collins, "ImageJ for microscopy," *BioTechniques*, vol. 43, no. 1S, pp. 25–30, 2007.

[176] H. Ewers, A. E. Smith, I. F. Sbalzarini, H. Lilie, P. Koumoutsakos, and A. Helenius, "Single-particle tracking of murine polyoma virus-like particles on live cells and artificial membranes," *Proceedings of the National Academy of Sciences*, vol. 102, no. 42, pp. 15110–15115, 2005.

[177] N. Parveen, D. Borrenberghs, S. Rocha, and J. Hendrix, "Single viruses on the fluorescence microscope: Imaging molecular mobility, interactions and structure sheds new light on viral replication," *Viruses*, vol. 10, no. 5, pp. 1–21, 2018.

[178] K. Inamdar, C. Floderer, C. Favard, and D. Muriaux, "Monitoring hiv-1 assembly in living cells: Insights from dynamic and single molecule microscopy," *Viruses*, vol. 11, no. 1, pp. 2–13, 2019.

[179] G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, T. Bland, S. Bo, J. A. Conejero, N. Firbas, Ò. Garibo i Orts, A. Gentili, Z. Huang, J.-H. Jeon, H. Kabbech, Y. Kim, P. Kowalek, D. Krapf, H. Loch-Olszewska, M. A. Lomholt, J.-B. Masson, P. G. Meyer, S. Park, B. Requena, I. Smal, T. Song, J. Szwabiński, S. Thapa, H. Verdier, G. Volpe, A. Widera, M. Lewenstein, R. Metzler, and C. Manzo, "Objective comparison of methods to decode anomalous diffusion," *Nat. Commun.*, vol. 12, pp. 1–16, 2021.

[180] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," *Ann. Phys.*, vol. 322, no. 8, pp. 549–560, 1905.

[181] R. Durrett, *Probability: Theory and examples*, vol. 49. Cambridge University Press, 2019.

[182] M. J. Saxton, "Single-particle tracking: The distribution of diffusion coefficients," *Biophys. J.*, vol. 72, no. 4, pp. 1744–1753, 1997.

[183] X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," *Phys. Rev. E*, vol. 82, no. 4, pp. 1–13, 2010.

[184] C. P. 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. 45, pp. 1–14, 2004.

[185] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. Ortiz de Solórzano, J.-C. Olivo-Marin, and E. Meijering, "Objective comparison of particle tracking methods," *Nat. Methods*, vol. 11, no. 3, pp. 281–290, 2014.

[186] K. E. Magnusson and J. Jaldén, "Tracking of non-Brownian particles using the Viterbi algorithm," in *Proc. ISBI*, pp. 380–384, IEEE, 2015.

[187] 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.

[188] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113, IET, 1993.

[189] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," *IEEE Sens. J.*, vol. 2, no. 2, pp. 107–119, 2002.

[190] W. Wu, J. Jiang, W. Liu, X. Qin, and W. Zhang, "A sequential converted measurement Kalman filter in the ECEF coordinate system for airborne Doppler radar," *Aerosp. Sci. Technol.*, vol. 51, pp. 11–17, 2016.

[191] Z. Deng, P. Zhang, W. Qi, J. Liu, and Y. Gao, "Sequential covariance intersection fusion Kalman filter," *Inf. Sci.*, vol. 189, pp. 293–309, 2012.

[192] H. Li, F. Nashashibi, and M. Yang, "Split covariance intersection filter: Theory and its application to vehicle localization," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1860–1871, 2013.

[193] Q. Guo, S. Chen, H. Leung, and S. Liu, "Covariance intersection based image fusion technique with application to pansharpening in remote sensing," *Inf. Sci.*, vol. 180, no. 18, pp. 3434–3443, 2010.

[194] Y. Bar-Shalom, "On the track-to-track correlation problem," *IEEE Trans. Automat. Contr.*, vol. 26, no. 2, pp. 571–572, 1981.

[195] V. Briane, M. Vimond, and C. Kervrann, "An overview of diffusion models for intracellular dynamics analysis," *Brief. Bioinformatics*, vol. bbz052, pp. 1–15, 2019.

[196] S. J. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection," in *Handbook of multisensor data fusion: Theory and practice*, CRC Press, 2009.

[197] C. Ritter, T. Wollmann, P. Bernhard, M. Gunkel, D. M. Braun, J.-Y. Lee, J. Meiners, R. Simon, G. Sauter, H. Erfle, K. Rippe, R. Bartenschlager, and K. Rohr, "Hyperparameter optimization for image analysis: Application to prostate tissue images and live cell data of virus-infected cells," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, pp. 1847–1857, 2019.

[198] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in *Proc. ICLR*, 2018.

[199] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.

[200] J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri, "Trackmate: An open and extensible platform for single-particle tracking," *Methods*, vol. 115, pp. 80–90, 2017.

[201] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, and J.-C. Olivo-Marin, "Icy: An open bioimage informatics platform for extended reproducible research," *Nat. Methods*, vol. 9, no. 7, pp. 690–696, 2012.

[202] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, 2012.

[203] C. Ritter, A. Newrly, S. Schifferdecker, I. Roggenbach, B. Müller, and K. Rohr, "Optimizing particle detection by colocalization analysis in multi-channel fluorescence microscopy images," in *Proc. ISBI*, pp. 882–885, IEEE, 2020.

[204] D. Schacherer, C. Ritter, and K. Rohr, "Multiple hypothesis tracking with integrated cell division detection," in *Proc. ISBI*, pp. 165–168, IEEE, 2021.

[205] V. Ulman, M. Maška, K. E. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, *et al.*, "An

objective comparison of cell-tracking algorithms," *Nat. Methods*, vol. 14, no. 12, pp. 1141–1152, 2017.

[206] 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. Imag.*, vol. 31, no. 9, pp. 1786–1808, 2012.

[207] S. A. Rahman, P. Koch, J. Weichsel, W. J. Godinez, U. Schwarz, K. Rohr, D. C. Lamb, H.-G. Kräusslich, and B. Müller, "Investigating the role of F-actin in human immunodeficiency virus assembly by live-cell microscopy," *J. Virol.*, vol. 88, no. 14, pp. 7904–7914, 2014.

[208] C.-M. Svensson, A. Medyukhina, I. Belyaev, N. Al-Zaben, and M. T. Figge, "Untangling Cell Tracks: Quantifying cell migration by time lapse image data analysis," *Cytometry Part A*, vol. 93, no. 3, pp. 357–370, 2018.

[209] Y. Wang, S. Du, S. Balakrishnan, and A. Singh, *Stochastic zeroth-order optimization in high dimensions.* arXiv:1710.10551, 2017.

[210] D. Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley, "Google vizier: A service for black-box optimization," in *Proc. SIGKDD*, pp. 1487–1495, ACM, 2017.

[211] L. Hertel, J. Collado, P. Sadowski, and P. Baldi, "Sherpa: Hyperparameter optimization for machine learning models," in *Proc. NeurIPS*, 2018.

[212] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms," in *Proc. SIGKDD*, pp. 847–855, ACM, 2013.

[213] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 2951–2959, 2012.

[214] B. Komer, J. Bergstra, and C. Eliasmith, "Hyperopt-sklearn: Automatic hyperparameter configuration for scikit-learn," in *Proc. ICML*, pp. 2825–2830, PMLR, 2014.

[215] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. LION*, pp. 507–523, Springer-Verlag, 2011.

[216] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. SIGKDD*, pp. 785–794, ACM, 2016.

[217] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, 1962.

[218] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley, 1989.

[219] X. Trepat, Z. Chen, and K. Jacobson, "Cell migration," *Compr. Physiol.*, vol. 2, no. 4, pp. 2369–2392, 2012.

[220] T. Esteves, M. J. Oliveira, and P. Quelhas, "Cancer cell detection and tracking based on local interest point detectors," in *Proc. ICIAR*, pp. 434–441, 2013.

[221] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak, *et al.*, "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.

[222] D. Padfield, J. Rittscher, and B. Roysam, "Coupled minimum-cost flow cell tracking for high-throughput quantitative analysis," *Med. Image Anal.*, vol. 15, no. 4, pp. 650–668, 2011.

[223] K. E. Magnusson, J. Jaldén, P. M. Gilbert, and H. M. Blau, "Global linking of cell tracks using the Viterbi algorithm," *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 911–929, 2015.

[224] E. Türetken, X. Wang, C. J. Becker, C. Haubold, and P. Fua, "Network flow integer programming to track elliptical cells in time-lapse sequences," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 942–951, 2017.

[225] C. Bron and J. Kerbosch, "Algorithm 457: Finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[226] X. Ren, Z. Huang, S. Sun, D. Liu, and J. Wu, "An efficient MHT implementation using GRASP," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 86–101, 2014.

[227] P. Matula, M. Maška, D. V. Sorokin, P. Matula, C. Ortiz-de Solórzano, and M. Kozubek, "Cell tracking accuracy measurement based on comparison of acyclic oriented graphs," *PloS ONE*, vol. 10, no. 12, e0144959, 2015.

[228] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[229] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 209–226, 1977.

[230] J. C. Maxwell, "Experiments on colour, as perceived by the eye, with remarks on colour-blindness," *Trans. R. Soc. Edinb.*, vol. 21, pp. 299–301, 1855.

[231] G. Peruzzi and V. Roberti, "The color top and the distinction between additive and subtractive color mixing [historical corner]," *IEEE Antennas Propag. Mag.*, vol. 61, no. 5, pp. 138–145, 2019.

[232] R. Haralick, "Some neighborhood operators," in *Real-Time Parallel Computing*, pp. 11–35, 1981.

[233] S. Padilla-Parra, M. Marin, N. Kondo, and G. B. Melikyan, "Pinpointing retrovirus entry sites in cells expressing alternatively spliced receptor isoforms by single virus imaging," *Retrovirology*, vol. 11, no. 1, pp. 1–14, 2014.

[234] L. Song, E. Hennink, I. T. Young, and H. J. Tanke, "Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy," *Biophys. J.*, vol. 68, no. 6, pp. 2588–2600, 1995.

[235] E. Bakker and P. S. Swain, "Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, 2019.

[236] J.-C. Grivel and L. Margolis, "Use of human tissue explants to study human infectious agents," *Nat. Protoc.*, vol. 4, no. 2, pp. 256–269, 2009.

[237] B. M. Baker and C. S. Chen, "Deconstructing the third dimension: How 3D culture microenvironments alter cellular cues," *J. Cell Sci.*, vol. 125, no. 13, pp. 3015–3024, 2012.

[238] J. B. Beltman, A. F. Marée, J. N. Lynch, M. J. Miller, and R. J. De Boer, "Lymph node topology dictates T cell migration behavior," *J. Exp. Med.*, vol. 204, no. 4, pp. 771–780, 2007.

[239] N. Jagiella, D. Rickert, F. J. Theis, and J. Hasenauer, "Parallelization and high-performance computing enables automated statistical inference of multi-scale models," *Cell Syst.*, vol. 4, no. 2, pp. 194–206.e9, 2017.

[240] M. Altfeld and M. Gale Jr, "Innate immunity against HIV-1 infection," *Nat. Immunol.*, vol. 16, pp. 554–562, 2015.

[241] G. Drazer and D. H. Zanette, "Experimental evidence of power-law trapping-time distributions in porous media," *Phys. Rev. E*, vol. 60, no. 5, pp. 5858–5864, 1999.

[242] S. Iwami, J. S. Takeuchi, S. Nakaoka, F. Mammano, F. Clavel, H. Inaba, T. Kobayashi, N. Misawa, K. Aihara, Y. Koyanagi, and K. Sato, "Cell-to-cell infection by HIV contributes over half of virus infection," *eLife*, vol. 4, p. e08150, 2015.

[243] A. Sporbert, A. Gahl, R. Ankerhold, H. Leonhardt, and M. C. Cardoso, "DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters," *Mol. Cell.*, vol. 10, no. 6, pp. 1355–1365, 2002.