Chapter One

Introduction

There are two possible outcomes: If the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery.

— Enrico Fermi (1901–1954)

1.1 Studying Intracellular Dynamics

The past decades have witnessed development of groundbreaking tools and techniques for imaging and studying cellular and intracellular structures and processes. The advent of confocal microscopy in the early sixties accompanied by discovery of fluorescent proteins has triggered the development of new imaging techniques and revolutionized the way biologists study cells and the way they function. Currently, fluorescence microscopy imaging is still the most important and frequently used tool for studying intracellular dynamics with a high spatial and temporal resolution. Proper understanding of cellular and molecular processes is of great interest to academic researches as well as pharmaceutical industries. The possibility to influence those processes in a controlled way is a prerequisite to combat diseases and improve human health care, which will have profound social and economic impact.

In fluorescence microscopy, the studying of the dynamical processes within a cell is usually done by labeling intracellular structures of interest with fluorescent proteins and following them in time using time-lapse imaging (see Fig. 1.1). The observed dynamical processes can be either studied qualitatively or using some quantitative measures that characterize intracellular behavior. Tracking of subcellular structures in time leads to creating of so called life histories, from which motion parameters such as velocity, acceleration and/or intensity changes in time can be easily estimated.

In practice, fluorescence microscopy, which in many laboratories become a universal tool for studying cellular and intracellular life, has some inherent limitations. One of them is autofluorescence. Autofluorescence describes the emission of fluorescence from naturally fluorescent molecules other than the fluorophore of interest. In fluorescence microscopy imaging, it is a significant source of background noise in images, which can be reduced either by special sample preparation or by background subtraction using image processing methods [185]. Another limiting factor is photobleaching. Photobleaching is the photochemical destruction of a fluorophore, which complicates the observation of fluorescent molecules, since they will eventually be destroyed by the

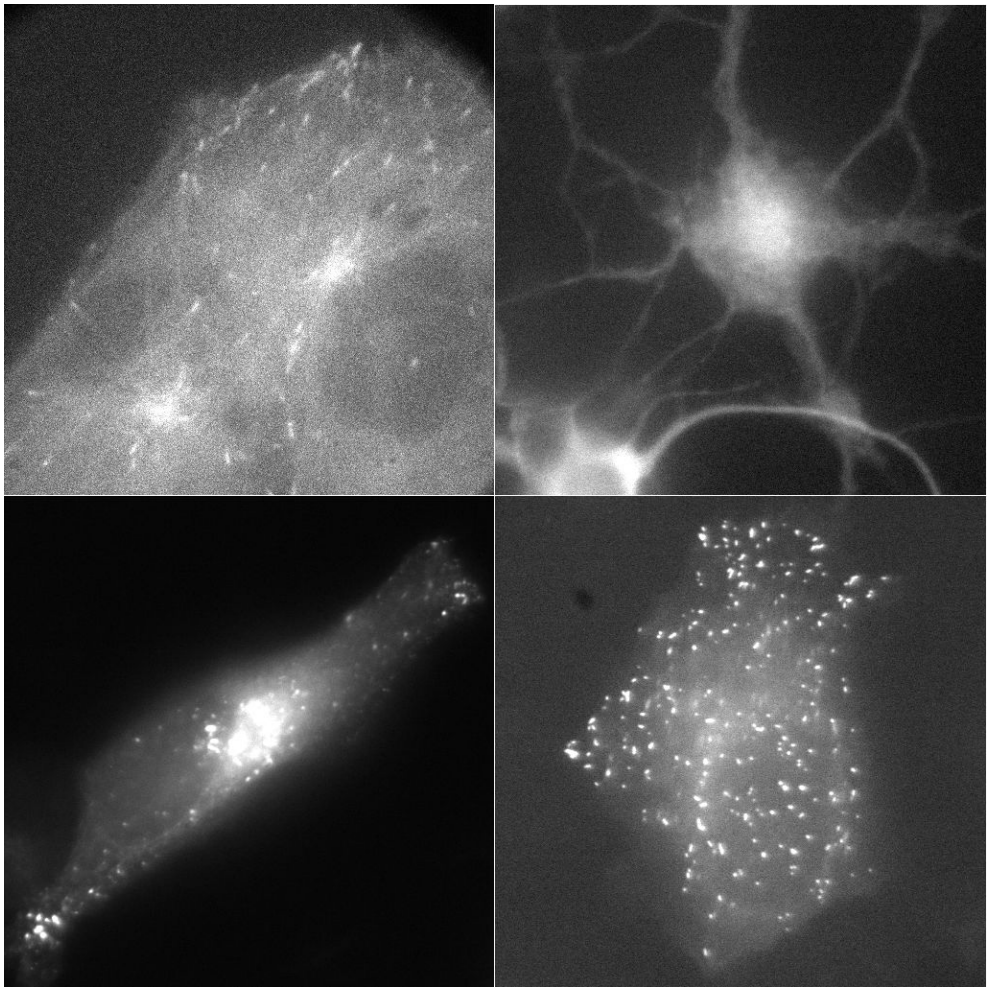

Figure 1.1. Examples of images acquired for different biological studies based on GFP labeling and fluorescence microscopy. The images are single frames from 2D time-lapse studies of activity of microtubule plus-ends (top left), microtubule plus-ends in neurons (top right), Rab5 (bottom left) and peroxisomes (bottom right).

light exposure necessary to stimulate them into fluorescing. This is especially problematic in time-lapse microscopy, where the fluorescence signal is imaged in time, but due to the photobleaching fades permanently lowering the image quality. On the positive side, however, these limiting phenomena serve as a basis for many advanced fluorescence measurement techniques. An example of this is fluorescence recovery after photobleaching (FRAP), which allows to determine diffusion coefficients, binding and dissociation rates [156, 185]. With this technique, the region of interest within a cell is photobleached, and the subsequent recovery in the bleached region as a result of movement of nonbleached fluorescent molecules from the surrounding areas is ob-

served and studied. By measuring the extent and speed to which this recovery occurs, conclusions can be drawn about diffusion of proteins within a membrane or protein turnover in complexes.

To study the localization of the photobleached molecules, sometimes a second fluorophore that remains visible during the imaging is added to the target subcellular structure. This process is called fluorescence localization after photobleaching (FLAP). Another technique, complementary to FRAP, is termed fluorescence loss in photobleaching (FLIP) [156,185]. This procedure involves repeated photobleaching of a cell region, which leads to permanent loss of the fluorescence light signal throughout the whole cell. If the loss is indeed observed, it indicates that free exchange between the molecules occurred between the bleached region and the rest of the cell. Otherwise, if there is no loss in the signal over the whole cell, the molecules in the bleached region are isolated and specifically localized in distinct cellular compartments.

A relatively new technique, which is used to study protein interaction, is fluorescence energy transfer (FRET) [156, 185]. FRET involves the radiationless transfer of energy from a donor fluorophore to an appropriately positioned acceptor fluorophore in a nanometer range. Such colocalization techniques are used to reveal functionally related molecules, and map the potential protein-to-protein interactions with high precision providing better understanding of how the intracellular dynamics is regulated, and thereby establishing its relationship to important disease processes. Other frequently used techniques are fluorescence lifetime imaging (FLIM), fluorescence in situ hybridization (FISH) and fluorescence ratio imaging (RI) [185].

Current biological studies using time-lapse fluorescence microscopy imaging require analysis of huge amounts of image data. A large-scale analysis of the dynamics of subcellular objects such as microtubules or vesicles cannot possibly be done without automatic tracking tools. The possibilities to study new aspects of the intracellular dynamics opened by modern imaging tools in combination with advanced image processing techniques impose high standards on robustness and accuracy of the tracking techniques for quantitative motion analysis. Moreover, there is demand for computationally fast methods that are capable of processing large amounts of data, which are typical for high-throughput experiments.

Tracking of multiple objects in biological image data is a challenging problem largely due to poor imaging conditions and complicated motion scenarios. Existing tracking algorithms for this purpose often do not provide sufficient robustness and/or are computationally expensive. By using such automatic tracking tools, biologists also eliminate the bias and possibly the systematic errors they introduce during manual tracking due to intuitive selection of relatively small subsets of objects of interest that are either nicely imaged or exhibit typical or expected motion patterns. Thus, automatic tracking methods capable of following large number of objects in time and classifying their dynamics, are of major interest.

1.2 Fundamental Limitations in Microscopy

In light microscopy, several factors complicate quantitative data analysis. In practice, careful design of experiments, the imaging system, and selection of appropriate tools

for the analysis can greatly reduce the influence of some of them. Nonetheless, light microscopy also has fundamental limitations that cannot be overcome and, most of the time, in real experiments biologists are inevitably facing those barriers. One of them is the limited spatial resolution of the microscope – there is a fundamental maximum to the resolution of any optical system due to diffraction. The diffraction limit depends on the emission wavelength, the numerical aperture of the objective lens, and defines the microscope point-spread function (PSF), which describes the response of an imaging system to a point light source. The Fraunhofer-diffraction limited PSF (normalized to unit magnitude at the origin) of a wide-field fluorescence microscope (WFFM) with circular aperture is given by [59]

$$\mathrm{PSF}(r,z) = \left| \int_0^1 2J_0(\alpha r \rho) \exp\left(-2i\gamma z \rho^2\right) \rho d\rho \right|^2 ,$$

where

$$\alpha = \frac{2\pi \mathrm{NA}}{\lambda} \qquad \text{and} \qquad \gamma = \frac{\pi \mathrm{NA}^2}{2\lambda n},$$

and $r = \sqrt{x^2 + y^2}$ denotes the radial distance to the optical axis, z is the axial distance to the focal plane, i the imaginary unit number, J_0 the zero-order Bessel function of the first kind, NA the numerical aperture of the objective lens, n the refractive index of the sample medium and λ the wavelength of the light emitted by the specimen. For a laser scanning confocal microscope (LSCM), the PSF is a combination of the excitation and emission intensity distributions. In the case of ideal confocality (infinitely small pinhole size) and assuming that the wavelengths of the emission and excitation light are approximately the same, the LSCM PSF reduces to the product of two WFFM PSFs [190]. In practice, a Gaussian approximation of the PSF is used, which is favored for computational reasons but is nevertheless almost as accurate as more complicated PSF models [55, 190]. The approximation (normalized to unit magnitude at the origin) is given by

$$\mathrm{PSF}_g(r,z) = \exp\left(-\frac{r^2}{2\sigma_r^2} - \frac{z^2}{2\sigma_z^2}\right),$$

where σ_r^2 and σ_z^2 (for a confocal microscope) are given by [190]

$$\sigma_r = 0.16\frac{\lambda}{\mathrm{NA}} \qquad \text{and} \qquad \sigma_z = 0.55\frac{\lambda n}{\mathrm{NA}^2}.$$

In this case, for noise-free images, the lateral and axial distances of resolution, d_{xy}^R and d_{z}^R, for equally bright fluorescent tags is given by the Rayleigh distances [69]

$$d_{\mathrm{xy}}^R = 0.56\frac{\lambda}{\mathrm{NA}} \qquad \text{and} \qquad d_{\mathrm{z}}^R = 1.5\frac{\lambda n}{\mathrm{NA}^2}.$$

For typical microscope setups the lateral resolution is on the order of 200 nm, and the axial resolution, which is always worse, is on the order of 600 nm. This resolution barrier is always encountered in experiments where subcellular structures

(microtubules, vesicles, etc.) are studied. Due to subresolution sizes of those objects ($< 10 - 20$ nm), they appear in the images as blurred spots. Attempts to overcome these limits range from engineering of new optical systems, such as multiphoton microscopy, stimulated emission and depletion (STED) microscopy, or 4Pi microscopy, to applying sophisticated post-acquisition computational analysis methods that do not require any modifications of the imaging system [60]. In the latter case, deconvolution algorithms and super-resolution methods are used, which necessarily exploit the prior knowledge about the optical system and/or the image formation process. A number of advanced deconvolution methods are available for image restoration in microscopy imaging [24, 65, 109, 129]. While there are suggestions in the literature to always deconvolve the image data if possible [24], the question whether deconvolution is beneficial in fact depends on the application. Most reports on tracking of subcellular structures do not mention the use of deconvolution, because the localization of such diffraction limited objects can be done with much higher accuracy and precision then the resolution of the imaging system using super-resolution methods [2, 35, 118, 162]. On the other hand, these two (deconvolution and super-resolution) approaches for post-acquisition image enhancement are not completely independent. Some super-resolution algorithms, for instance, are based on fitting (a model) of the PSF – to some degree this is in fact deconvolution, carried out implicitly in the process.

The second factor that complicates the data analysis is noise, which is a stochastic phenomenon that cannot be compensated for, contrary to systematic distortions such as blurring. In light microscopy, the imaging is commonly done using a charge-coupled device (CCD) camera, which is a semiconductor device that converts the incoming light photons first to electrical charges and then to voltages which are read out from the device, quantized and stored as a digital image. Unfortunately, every step of this imaging process is influenced by different noise sources: photon noise, thermal noise (dark current and hot pixels), readout noise (on-chip electronic noise) and quantization noise [179]. Photon noise, which is due to the quantum nature of light, follows the Poisson distribution and is signal/amplitude dependent. Since the Poisson distribution approaches the Gaussian distribution for large numbers, the photon noise in a signal will approach the normal distribution for large numbers of collected photons. Thermal noise is also Poisson distributed but can be greatly reduced by cooling the CCD chip. Readout noise, the influence of which becomes significant only for high readout rates (>1MHz), is caused by the on-chip electronics. This source of noise is Gaussian distributed and independent of the signal. Quantization noise is caused by the conversion of the analog signal (voltage) to digital representation. This noise is additive, uniformly distributed, and with modern analog-to-digital converters is very low and usually ignored. In practice, the influence of all of these noise sources (except for the photon noise) can be made negligible by proper electronic design and careful operation conditions. Thus, photon noise is the main and fundamental limiting factor that defines the signal-to-noise ratio (SNR) of the image data in microscopy imaging. Low SNRs are especially typical in live-cell fluorescence microscopy, since in most experiments the imaged light signal is quite weak – high excitation light rapidly quenches fluorescence and may disturb intracellular processes being studied.

A final complicating factor worth mentioning here is the large variability of biological image data. This especially complicates the development of universal automatic

methods for quantitative image analysis. In molecular biology, which is a highly experimental field, the absence of standardization in the acquisition and data storage protocols leads to image data of strongly varying quality, even within one type of experiments. This heterogeneity negatively influences the validation of the automatic techniques for studying the intracellular processes and lowers the reproducibility of new results and findings. All these factors put high demand on the design of automated image analysis techniques. This is in contrast with medical investigations, where routine clinical studies are based on standardized imaging protocols, leading to more consistent image quality.

1.3 Tracking in Fluorescence Microscopy

The quantitative analysis of time-lapse image sequences that visualize intracellular processes usually requires tracking of multiple objects over time. The majority of the automated tracking techniques described in the literature and available in practice process the data by following a few well established subsequent steps: preprocessing the image data, detecting the objects of interest independently in every image frame, and creating the trajectories by linking the sets of detected objects in subsequent frames. Extracted trajectories are further used for estimation of important parameters that characterize the intracellular dynamics.

1.3.1 Image Preprocessing

The main purpose of preprocessing is to enhance the image quality and, if necessary, compensate for global cell motion. Recent comparative studies demonstrated that the accuracy of commonly used tracking methods is mainly determined by the SNR of the image data [32]. While different SNR measures exist, here we define the noise level as the standard deviation of the intensities within the object, not the background. Correspondingly, the SNR is defined as the difference in intensity between the object and the background, divided by the standard deviation of the object noise [32]. Due to non-Gaussian noise statistics in the images, apart from linear Gaussian filtering [159] or wavelet-based denoising [108, 154], frequently nonlinear methods, such as median filtering [18] or anisotropic diffusion filtering [168] are used.

For studying intracellular dynamics, it may sometimes be necessary to compensate for the global motion of the cell, so as not to over- or underestimate local motion parameters. For this purpose, rigid or nonrigid image registration can be used [123, 150], which came to biological imaging mainly from medical image analysis, where it has been used on a regular basis for years. Another approach is to track the cells over time using existing cell tracking methods [43, 44, 192] and derive the final estimates and conclusions by combining both (cellular and intracellular) sources of information.

1.3.2 Object Localization

After the preprocessing step, object detection methods are applied to the image data in order to locate objects of interest and accurately estimate their positions. The

simplest detection algorithms are based on image intensity thresholding with the underlying assumption that the real objects are brighter than background structures. In the image regions that indicate object presence after thresholding, the object position can be estimated using the centroid method [26, 32]. For a single axis in a 2D image I, the estimate is given by

$$x_c = \sum_x \sum_y xI(x,y) / \sum_x \sum_y I(x,y),$$

where $I(x,y)$ is the image intensity value at position (x,y) and the summation is done over a small image region (mask) that contains the object. Depending on the mask size and image quality, in order to eliminate the bias in the estimated position towards the center of the mask, sometimes only positive differences between the intensity values and the threshold for each pixel within the mask are used in the centroid method [32]. For reasonable performance, such simple methods normally require images with relatively uniform background and high SNR, which makes them unsuitable for most live-cell imaging experiments.

More advanced detectors use additional features, such as object size, shape, volume, etc., for better discrimination from irrelevant background structures. By using additional features, these methods better model the object appearance and try to fit the models to the image data using some similarity measures. The model fitting is usually done by minimizing a predefined error measure (e.g. least squares fitting), or by measuring how good the model correlates with the data. The latter can be done, for example, by computing the normalized covariance for the small intensity template T that describes the object appearance and the original image I. This method is an extension of simple correlation with the template T [32], which originally cannot deal with nonuniform backgrounds. In 2D, the normalized covariance is given by[1]

$$C(x,y) = \sum_{i=-n}^{n} \sum_{j=-m}^{m} \frac{(I(x+i,y+j) - \bar{I}(x,y))(T(i+n,j+m) - \bar{T})}{M_I(x,y) M_T},$$

where $T(i,j)$ is a $(2n+1) \times (2m+1)$ intensity template, $\bar{T}$ the mean value of the template intensity, $\bar{I}(x,y)$ the mean value of the image intensity in the area overlapping with the template, M_T the variance of the template intensity, and $M_I(x,y)$ is the variance of image intensity in the area overlapping with the template [32]. The local maxima in the resulting map $C(x,y)$ indicate the image regions which are highly similar in appearance to the template. By applying a threshold to $C(x,y)$, these regions can be extracted and the object positions can be computed using the centroid method. The normalized covariance can cope with nonuniform background intensity and the only limiting assumption is the fixed and known shape of the searched object.

Another similarity measure that can be used to measure the correspondence between the object appearance template T and the spatial intensity distributions in the image data is a sum of absolute differences (SAD). In this method, the SAD map is

[1] The extension of this and subsequent formulae to 3D is straightforward.

computed for all possible shifts of the template T in the image as

$$\mathrm{SAD}(x,y) = \sum_{i=-n}^{n} \sum_{j=-m}^{m} |I(x+i, y+j) - T(i+n, j+m)|.$$

The minima in the map $\mathrm{SAD}(x,y)$ correspond to the best fits. For multiple object detection the local minima are counted as found objects. Compared to covariance based detection, this method is highly sensitive to intensity scaling of the image and template, which can cause problems in practice since the fluorescent tags are bleaching during acquisition.

For the described correlation based methods, the accuracy of the position estimates is on the order of one pixel, since the shifts of the template are calculated on a discrete pixel grid. The accuracy of the object localization can be substantially increased by using detection methods that fit the object appearance model to the image data. Since the objects under consideration are smaller than the resolution of the imaging devices, the model of the PSF (for example the Gaussian approximation) can be used in order to model object appearance. For multiple object detection in 2D images, the fitting is performed in all the regions of the image where the probability of object existence is high, by minimizing the sum of squared differences

$$\mathrm{MSE}_g(x,y,A,B) = \sum_i \sum_j \left(I(i,j) - A \exp\left(-\frac{(i-x)^2 + (j-y)^2}{2\sigma^2} \right) - B \right)^2 .$$

The parameters that locally minimize the MSE_g, (x,y,A,B), are taken as the features of the found object [32, 161]. This approach is computationally expensive, but it demonstrates the highest accuracy in estimating the object position [32]. The latter conclusion comes from the study [32], where the approaches described above were quantitatively compared under different controlled conditions using artificial 2D time-lapse image sequences and is true only in the case of high SNR image data. For low SNR levels (< 5), which are not uncommon in live-cell fluorescence microscopy imaging, the PSF model fitting breaks down [161].

1.3.3 Solving the Correspondence Problem

Once the objects have been detected in the image sequence, sets of estimated positions $\{\{\mathbf{r}_t^k\}_{k=1}^{M_t}\}_{t=1}^{T_0}$ are available for the next processing step, where $\mathbf{r}_t^k = (x_t^k, y_t^k, z_t^k)^T$ defines the position of object k in frame t, M_t is a time varying number of objects per frame, and T_0 is the number of frames in the image sequence. In order to obtain trajectories, the correspondence between the object positions in different frames needs to be established. Solving the correspondence problem is not a trivial task. In our application, the objects of interest are more or less identical and because of that searching for the corresponding objects in different frames on a basis of appearance information will not produce good results. In practice, the detection procedures are imperfect, which leads to spurious and missing objects that influence the accuracy of the linking procedure. Moreover, real objects can move densely together, be temporarily occluded and/or appear and disappear from the field of view during imaging.

Frequently, such ambiguous scenarios cannot be correctly dealt with using the existing tracking algorithms, and sometimes such situations confuse even expert biologists. This is especially the case when complicated, essentially 3D intracellular processes are studied using 2D confocal slicing.

In the case of almost indistinguishable objects, the linking procedure is mainly based on assumptions about the underlying object motion. The most frequently used motion models are the nearest-neighbor model (NNM) and the smooth motion model (SMM) [33, 171]. The NNM does not incorporate velocity information and is solely based on positional information. For object k in frame t and candidate object s in frame $t+1$, the score $c_t^{\mathrm{NN}}(k,s)$ is defined as $c_t^{\mathrm{NN}}(k,s) = \|\mathbf{r}_t^k - \mathbf{r}_{t+1}^s\|$. The object pair with the lowest score has the highest chance to be linked. If an object stays in one place, the score $c_t^{\mathrm{NN}}(k,s) = 0$. The SMM, on the other hand, assumes that both velocity direction and magnitude change slowly from frame to frame. For this model, the corresponding score is defined as

$$c_t^{\mathrm{SM}}(k,s) = w\left(1 - \frac{\mathbf{v}_t^k \cdot \mathbf{v}_t^{ks}}{\|\mathbf{v}_t^k\|\|\mathbf{v}_t^{ks}\|}\right) + (1-w)\left(1 - \frac{2\sqrt{\|\mathbf{v}_t^k\|\|\mathbf{v}_t^{ks}\|}}{\|\mathbf{v}_t^k\| + \|\mathbf{v}_t^{ks}\|}\right),$$

where $\mathbf{v}_t^k = \mathbf{r}_t^k - \mathbf{r}_{t-1}^k$, $\mathbf{v}_t^{ks} = \mathbf{r}_{t+1}^s - \mathbf{r}_t^k$, and w is a weighting coefficient. The first term in the expression for $c_t^{\mathrm{SM}}(k,s)$ accounts for the angular deviation of the displacement vectors by computing their dot product. The second term accounts for the speed deviation. Using this score, a candidate object in frame $t+1$ is searched that best satisfies the uniform motion assumption. If the object moves uniformly, $\mathbf{r}_t^k - \mathbf{r}_{t-1}^k = \mathbf{r}_{t+1}^s - \mathbf{r}_t^k$, and $c_t^{\mathrm{SM}}(k,s) = 0$. If applicable, appearance similarity measures can be used in addition to the described spatial proximity criteria. Furthermore, the combination of these measures can be used to define some probability of assignment as a score, e.g.

$$c_t^{\mathrm{P}}(k,s) = \exp\left(-(\mathbf{r}_{t+1}^s - \mathbf{r}_t^k)^T \boldsymbol{\Sigma} (\mathbf{r}_{t+1}^s - \mathbf{r}_t^k)\right) \exp\left(-\frac{(I_{t+1}^s - I_t^k)^2}{\sigma_I^2}\right),$$

where I_{t+1}^s and I_t^k are the intensities of objects s and k in the corresponding frames, and $\boldsymbol{\Sigma}$ and σ_I^2 are the parameters that account for small deviation in displacement and variation in intensity, respectively.

In order to link the objects and form the trajectories, the described assignment scores can be used in several ways. First, there are greedy algorithms that make decisions about the best assignment by taking into account the score values only in the current frame. The disadvantage of the greedy search is its tendency to stop in the first local minimum of the searched space. At the same time, if the density of the objects in the image data is relatively low, and the motion is either slow or uniform, so that the NNM or SMM are appropriate, then the greedy approach is a good choice (also because it is computationally quite cheap). In general, the linking procedures can operate either globally in space, where the assignment is performed jointly for all objects in one frame, or globally in time, depending on how many frames are taken into account at the same time for similarity measurement, or both (be global in space and time). Most of the time, a greedy assignment is done first. Then, the iterative

procedures start to make changes in these assignments and check how the global score behaves (if it is lowered).

Many linking techniques solve the correspondence problem in a spatially global manner, by defining a global score that is afterwards minimized [33, 52, 74, 132, 166, 171]. For example, some form of global optimization can be accomplished using graph theory [132]. This approach is implemented in a publicly available tracking software *ParticleTracker*, the performance of which is evaluated in Chapter 3. Here, linking is based on finding the spatially global solution to the correspondence problem in a given number of successive frames. The solution is obtained using graph theory and global energy minimization [132]. The linking also utilizes the zero- and second-order intensity moments of the object intensities, which helps to resolve object intersection problems and improves the linking results.

Another solution to the correspondence problem can be obtained by using dynamic programming [128]. With dynamic programming, the total cost, which is in this case the weighted sum of c_t^{NN} and object intensity I_t, is optimally minimized in a temporally global way. With this approach, tracking of a single object through the entire image sequence is possible [128]. Multiple object tracking can be achieved by tracking the objects one by one, which is not an attractive and workable solution for image data with large numbers of interacting objects.

Recently presented advanced linking techniques use fuzzy-logic and linear assignment problem (LAP) frameworks. In the former approach [74], four cost functions that measure the object similarity in consecutive frames are introduced: two of them are similar to the two summands in c_t^{SM}, and two additional costs are based on the objects appearance, $c_t^I(k,s) = 1 - |I_t^k - I_{t+1}^s|/|I_t^k + I_{t+1}^s|$ and $c_t^S(k,s) = 1 - |A_t^k - A_{t+1}^s|/|A_t^k + A_{t+1}^s|$, where I_t^k is the total intensity and A_t^k is the total area of the spot k in frame t. Further, the fuzzy-logic system is employed to estimate the similarity between the object in frame t (parent object) and a set of candidate objects in frame $t+1$. Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise. A set of if-then rules is introduced, where each rule uses the values of the four similarity measures and outputs a real value between 0 and 1. This gives the possibility to extend the binary concept that a parent object is similar ("1") or not similar ("0") to a candidate object to a broader range: "least-similar", "median-similar", "most-similar", etc. The outputs of all the rules are aggregated and a common score is derived for each candidate object. The parent object is connected with the candidate object that has the highest score. With this approach, fuzzy rule selection plays an important role and it strongly affects the performance of the tracking algorithm. Additionally, in the described algorithm [74], the linking is performed separately for each object, so the whole procedure is global neither in time, nor in space.

One of the most recent approaches in the literature [72] constructs the set of trajectories from the set of detected objects in two steps. First, the greedy assignment between the consecutive frames is performed using the cost function based on the distance between two objects. This step produces many short and broken tracks. The second step attempts to link the track segments (close the gaps) and deal with track splitting and merging using additionally the object intensity information. For this stage, corresponding closing, splitting, and merging cost functions are defined, which

have to be tailored to the specific application. Although the first step is greedy, solving the subsequent track segment optimization problem is done globally, overcoming the shortcomings of the previously described techniques. The method was shown to perform robust tracking of multiple objects under high-density conditions [72]. The shortcoming of this and the previously described method is again the separation of the detection and tracking procedures.

A somewhat different solution to the correspondence problem is presented in [17]. Rather than adopting the usual frame-by-frame approach, the authors consider the time-lapse 2D+t image sequence as one 3D spatiotemporal volume, where the tracks appear as 3D curves. The correspondence problem is then solved by finding the geodesics in a Riemannian metric computed from the 3D image. Similarly to the method described in [72], the cost optimization procedure, which is global only in time, is split into two steps. After the object detection, the nearby objects are grouped into short trajectories that are not complete due to possibly poor detection results. Then, partial tracks are linked with minimal paths to constitute complete tracks. The construction of minimal paths takes into account information from both image features and tracking constraints (maximum object displacement, etc.). Moreover, each time a minimal path is added to a trajectory, image information is removed along the path in order to avoid trajectories to merge.

The quality of the solution to the correspondence problem highly depends on the nature of the dynamical processes that were studied in the experiments, for example the number of objects, density of objects, type of motion, etc. Many of the described methods perform poorly when applied to biological data because of too simplistic assumptions of object behavior, which cannot cope with the real heterogeneity of subcellular dynamics. Additionally, due to separation of the tracking procedure into detection and linking, for low quality image data, the linking methods have to deal with lots of spurious objects detected in the first stage. Commonly, detectors do not specify any confidence measure for each detected object, that could be used to distinguish real objects from possible false detections. If that would be possible, the results of linking could be improved. Such confidence measure, for instance, can be specified in terms of variance in the object position measurements and is frequently used in probabilistic tracking approaches, which are the focus of this thesis.

1.3.4 Probabilistic Methods for Tracking

Solving the correspondence problem and creating tracks can also be described as a state estimation problem and solved using probabilistic methods [9,126]. Probabilistic tracking is a state estimation problem, where the object hidden state $\mathbf{x}_t$ is estimated in time based on previous states, noisy measurements $\mathbf{z}_t$, and prior knowledge about object properties. Mathematically, it can be formulated as

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}, \mathbf{v}_t), \qquad \mathbf{z}_t = h_t(\mathbf{x}_t, \mathbf{u}_t), \tag{1.1}$$

where f_t and g_t are possibly nonlinear state transition and observation models respectively, and $\mathbf{v}_t$ and $\mathbf{u}_t$ are white noise sources. If the measurement-to-object association is known, (1.1) can be solved either exactly (when f_t and g_t are linear and $\mathbf{v}_t$ and $\mathbf{u}_t$

are Gaussian) using the Kalman filter, or (in the general case) using SMC approximation methods [9]. The solution is the posterior probability distribution function (pdf) $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, where $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$, from which minimum mean square error (MMSE) or maximum a posteriori (MAP) state estimations can be easily computed [9].

In order to obtain the trajectory estimate for one object using the Kalman filter, the state vector $\mathbf{x}_t$, which may include object position, velocity, acceleration, etc., and which cannot be directly measured is estimated on the basis of noisy measurements $\mathbf{z}_{1:t}$, for example extracted positions $\mathbf{r}_{1:t}$ using detection methods described above. It is assumed that the state transition and the observation process are specified as follows,

$$\mathbf{x}_t = F_t\mathbf{x}_{t-1} + \mathbf{v}_t, \qquad \mathbf{z}_t = H_t\mathbf{x}_t + \mathbf{u}_t, \tag{1.2}$$

where F_t and H_t are system matrices defining the linear functions, and the covariances of $\mathbf{v}_t$ and $\mathbf{u}_t$, which are statistically independent random variables with zero mean, are respectively Q_t and R_t. The solution of (1.2), $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, in this case is given by the following recursive relationship:

$$\begin{aligned} p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) &= \mathcal{N}(\mathbf{x}_{t-1}|\mathbf{m}_{t-1|t-1}, P_{t-1|t-1}), \\ p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) &= \mathcal{N}(\mathbf{x}_t|\mathbf{m}_{t|t-1}, P_{t|t-1}), \\ p(\mathbf{x}_t|\mathbf{z}_{1:t}) &= \mathcal{N}(\mathbf{x}_t|\mathbf{m}_{t|t}, P_{t|t}), \end{aligned} \tag{1.3}$$

where

$$\begin{aligned} \mathbf{m}_{t|t-1} &= F_t\mathbf{m}_{t-1|t-1}, \\ P_{t|t-1} &= Q_{t-1} + F_tP_{t-1|t-1}F_t^T, \\ \mathbf{m}_{t|t} &= \mathbf{m}_{t|t-1} + K_t(\mathbf{r}_t - H_t\mathbf{m}_{t|t-1}), \\ P_{t|t} &= P_{t|t-1} - K_tH_tP_{t|t-1}, \end{aligned} \tag{1.4}$$

and where $\mathcal{N}(.|\mathbf{m}, P)$ is a Gaussian distribution with mean $\mathbf{m}$ and covariance P, and

$$\begin{aligned} S_t &= H_tP_{t|t-1}H_t^T + R_t, \\ K_t &= P_{t|t-1}H_t^TS_t^{-1}. \end{aligned}$$

For multiple object tracking the same framework can be used, but the tracking in this case is complicated by the ambiguous measurement-to-object associations – for every measurement given by the detector at time t it is necessary to know which object it has to be used for to update the predicted state in (1.4). In practice that information is not available. The most efficient tracking approaches that are able to deal with such missing information and still perform tracking, are the multiple hypothesis tracker (MHT) and the joint probabilistic data association (JPDA) filter [15]. The former builds a tree of hypotheses about all possible measurement-to-track associations, and because of that is not suitable for tracking large numbers of objects. The standard JPDA filter is designed for linear Gaussian models in (1.1) and uses all measurements to update each track estimate [15]. For practical reasons, measurement gating is often used, which selects for each object the subset of measurements that most likely originated from the object. Contrary to applications where sensors provide information about the number of objects and their positions, JPDA cannot be

applied directly to our applications, because actual position or velocity measurements are not available, but need to be derived from the image data first.

For analysis of subcellular dynamics a few approaches have been proposed that implement the described probabilistic framework [52, 146]. The first one extends the JPDA filter by using the h-dome detector (see also Chapter 2) and is shown to perform accurate and robust tracking of microtubules, which are growing with almost constant velocity. The second approach [52] implements the idea of interacting multiple model (IMM) filtering [11], which was initially designed only for linear Gaussian models. This type of filtering is useful when it is necessary to track objects that exhibit different types of motion patterns in the same image sequence. Here, several motion models F_t are employed, which predict the object position from frame to frame using the Kalman filter. The method was shown to perform extremely well in comparison with standard Kalman filtering for tracking of endocytosed quantum dots. In Chapter 4, a PF-based method is developed that generalizes the idea of IMM.

With the probabilistic tracking approaches, especially in the case of multiple object tracking, it is also beneficial to specify any prior knowledge about object interactions, additionally to the modeling of the object dynamics. Tracking approaches that assume a one-to-one measurement-to-track assignment (as in most of the deterministic tracking approaches and some of the probabilistic ones), fail to resolve the most ambiguous track interaction scenarios, where two or more objects come in close proximity to each other and produce only one measurement for a few time frames. By incorporating prior knowledge about the objects to be tracked (for example, microtubules are rigid structures that cannot easily bend, and because of that their direction of movement before and after the interaction should be approximately the same), the rate of incorrectly switched tracks can be greatly reduced [141, 146].

1.4 Analyzing Tracking Results

In time-lapse microscopy, the final step of the analysis consists of interpretation of the detection and tracking results in order to confirm or reject hypotheses that were tested during the experiment, or qualitatively or quantitatively look for new findings that would lead to new hypotheses and correspondingly to new experiments. Before applying any quantitative techniques, some qualitative verification of the obtained tracking results might be useful. This is especially true for low SNR image data, where lots of automated techniques either break down or produce nonsensical tracks. For this purpose, modern computer graphics rendering and visualization techniques can be used (see an example in Fig. 1.2) so as to assist in the verification of the tracking and give some initial impression about the possible trends in the data and which quantitative methods for the analysis to choose [54].

Once the results of tracking are verified, a multitude of measures about the geometry of the trajectories and additionally the object appearance can be readily obtained. An example is the total distance traveled by the object or the mean square displacement, which are typically used to study the diffusion characteristics of the motion of individual objects [8, 122, 128, 131, 158]. Other commonly studied parameters are average and instantaneous velocities. Instantaneous object velocity is estimated as

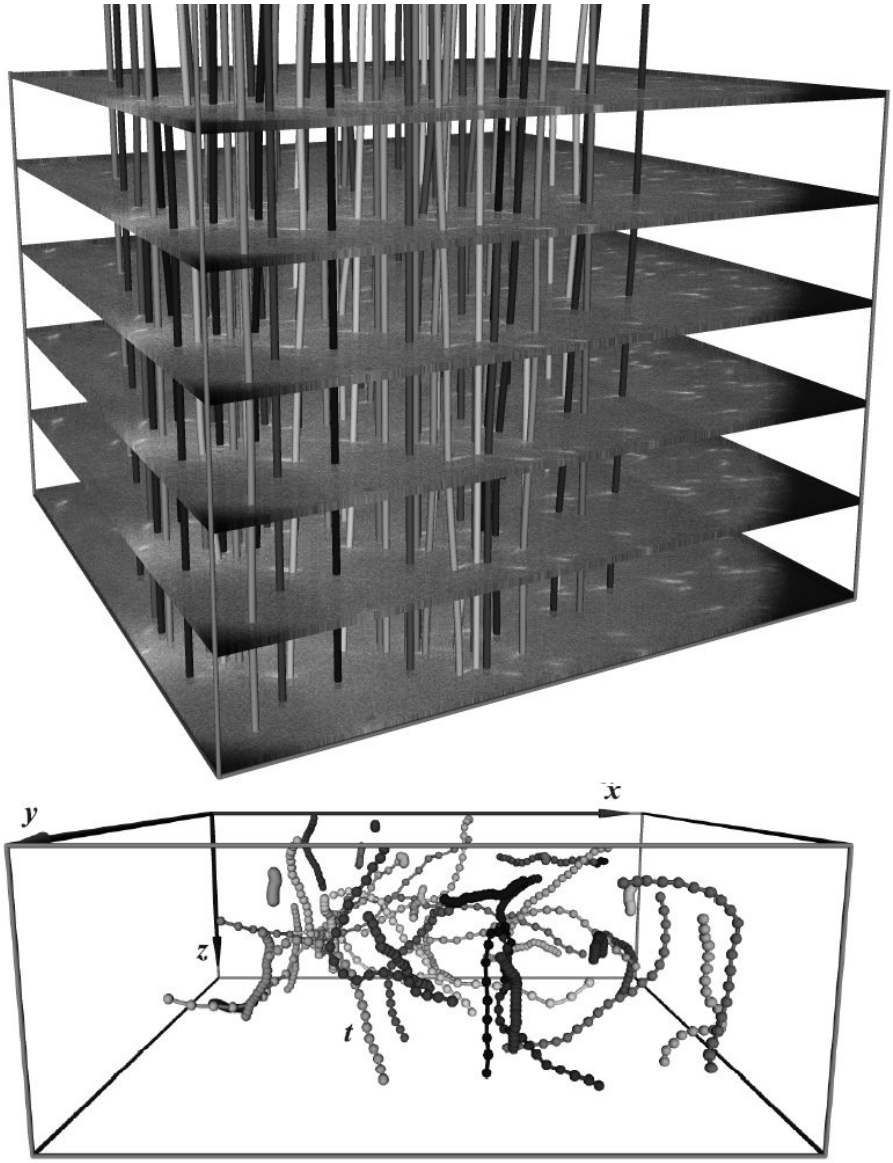

Figure 1.2. Different visualizations of time-lapse image data: combined visualizations of image frames and tracks giving a qualitative impression of the accuracy and consistency of the tracking results (top), and spatiotemporal view of tracks from an artificial 3D time-lapse image sequence (data not shown), with the time coordinate indicated along the trajectories by small spheres (bottom).

the distance traveled by the object between two consecutive frames divided by the corresponding time interval. Average velocity is computed as the sum of the frame-

to-frame distances traveled, divided by the total time elapsed. In many experiments it is enough to estimate the average velocity either per track or a number of tracks (and the variance of those estimates) in order to derive some conclusions about the hypotheses being tested. On the other hand, with current advanced imaging techniques, biologists are eager to inspect and analyze the intracellular motion in more detail. This can be achieved by studying the instantaneous velocities and their distributions [92, 167–169]. Contrary to the average velocity estimate, histograms of instantaneous velocities provide insight into the possible heterogeneity of the intracellular motion and reveal the most dominant modes of motion. Additionally, object acceleration can be easily estimated, but is rarely studied.

Ideally, automated tracking techniques in molecular biology should facilitate the study of behavioral heterogeneity, to find and classify distinctive motion patterns (or confirm absence of such) depending on the experimental conditions. Knowing the typical behavior patterns of "healthy" molecular processes, it will be much easier to understand abnormal behavior that leads to disease and to define strategies that return the deviated system to its normal state. Therefore, comprehensive and automated analysis of large scale experimental data is especially important.

1.5 Thesis Outline

The subject of this thesis is tracking of multiple subcellular objects using time-lapse microscopy imaging. The main focus is on the development of robust and accurate automatic tracking algorithms, built within a probabilistic framework. The Bayesian tracking framework, which recently has become popular in other research fields and was shown to outperform deterministic methods, is capable of solving complex estimation problems by combining available noisy measurements (images, extracted object positions, etc.) with prior knowledge about the underlying object dynamics and the measurement formation process. Nevertheless, it is still only a framework, which gives the solution in a very general form, independent of applications. In order to apply the Bayesian approach in practice, the "ingredients" of the framework must be made application specific. In our case, these are the image formation process, defining the object appearance in the images, the noise sources that influence the image quality, and prior knowledge about the object behavior. The more accurate these aspects are specified and modeled, the closer the estimation to optimal. Nevertheless, even with all these ingredients in place, the optimal Bayesian solution is analytically tractable only in a restrictive set of cases – for example, the Kalman filter provides an optimal solution in the case of linear dynamic systems with Gaussian noise. For most practical cases, approximation techniques must be used. One of the most powerful and especially suitable for this purpose are sequential Monte Carlo methods, also known as particle filtering[2] (PF). With that in mind, this thesis describes a set of PF-based methods, that have been developed and evaluated for tracking of multiple objects in a variety of time-lapse biological studies.

[2]In this thesis (except Chapter 2), the word "particle" does not refer to any real subcellular structures and, because of possible confusion, everywhere in the thesis the word "object" is used for those structures. The word "particle" is reserved for the PF methods.

The thesis organized as follows. First, in Chapter 2, a quantitative comparison of frequently used object detection methods applied to fluorescence microscopy images is described. Even in the case of probabilistic tracking methods, object detection methods are useful, for example for detection of appearing and disappearing objects during tracking. In the chapter, six supervised and two unsupervised (machine learning) techniques are quantitatively evaluated and compared, using both synthetic image data and images from real biological studies. A comparison of this sort has not been carried out before in the literature. Next, in Chapter 3, a new PF-based tracking technique is proposed for tracking of subcellular structures moving with nearly constant velocity, such as microtubules. Experiments on synthetic as well as real fluorescence microscopy image sequences demonstrate the superior performance of the new method compared to popular frame-by-frame tracking methods. Chapter 4 presents an extension of the method developed in Chapter 3, which is able to track multiple types of intracellular objects (microtubules, vesicles, and androgen receptors) and can deal with different types of motion patterns. For that, several improvements over the previous PF are developed. Finally, Chapter 5 describes another biological application, where microtubule dynamics is studied *in vitro*. It presents a novel PF-based approach for analysis of the image data and estimation of important microtubule dynamics parameters. For this application, a special type of particle filters is designed, for the tracking of spatiotemporal structures. The results presented in the various chapters lead to the general conclusion that PF-based methods are very suitable for subcellular object tracking in biological microscopy and are superior to existing deterministic approaches.