Doctoral Thesis

Algorithms for detection and tracking of objects with super-resolution in 3D fluorescence microscopy

Author(s):
Thomann, Dominik Michael

Publication Date:
2003

Permanent Link:
https://doi.org/10.3929/ethz-a-004519113

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Algorithms for Detection and Tracking of Objects with Super-Resolution in 3D Fluorescence Microscopy

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH
For the degree of
Doctor of Sciences

Presented by

Dominik Michael Thomann
Dipl. Phys. ETHZ
Born 06.08.1973
Citizen of Basel and Affeltrangen

Accepted on the recommendation of

Prof. Dr. Gaudenz Danuser, examiner
Prof. Dr. Peter K. Sorger, co-examiner
Prof. Dr. Andreas Stemmer, co-examiner

2003
Seite Leer / Blank leaf
Contents

Zusammenfassung.................................................................................................................. v
Summary................................................................................................................................. vii

1. Introduction....................................................................................................................... 9
   1.1 Resolution extension in microscopy ................................................................. 12
   1.2 Theory of Super-Resolution............................................................................. 14
   1.3 Outline of the thesis.......................................................................................... 16

2. One-step Calibration of Specimen Induced Focal Shifts and Spherical Aberration for Quantitative 3D Microscopy: Approach and First Results ................. 19
   2.1 Introduction ....................................................................................................... 22
   2.2 Parametric Model for the Point Spread Function ........................................ 27
   2.3 Object-constrained parametric blind deconvolution .................................... 33
   2.4 Simulation and preliminary Results ................................................................. 36
      2.4.1 Initial parameter guesses ......................................................................... 36
      2.4.2 Convergence ............................................................................................. 37
      2.4.3 Robustness in the presence of noise ......................................................... 39
   2.5 Discussion ............................................................................................................ 40
   2.6 Acknowledgements ............................................................................................. 41

3. Automatic Fluorescent Tag Detection in 3D with Super-Resolution: Application to the Analysis of Chromosome Movement .................................................... 43
   3.1 Introduction ....................................................................................................... 46
   3.2 Materials and Methods ..................................................................................... 51
   3.3 Algorithm for 3D Spot Detection with Super-Resolution ............................ 53
      3.3.1 A practical 3D model for the point spread function ............................. 55
      3.3.2 Noise reduction by matched filtering ....................................................... 58
      3.3.3 Spot detection ........................................................................................... 59
      3.3.4 Separation of overlapping spots ............................................................... 60
      3.3.5 Classification of spots ............................................................................. 62
      3.3.6 Motion analysis ......................................................................................... 63
   3.4 Results ............................................................................................................... 64
      3.4.1 Performance on biological data ................................................................. 64
      3.4.2 Performance on synthetic data ................................................................. 72
   3.5 Discussion ............................................................................................................ 76

4. Automatic fluorescent tag localization II: Improvement in super-resolution by relative tracking................................................................. 81
   4.1 Introduction ....................................................................................................... 84
   4.2 Algorithm for 3D Spot Tracking with Super-Resolution ............................ 88
      4.2.1 Spot Detection ......................................................................................... 88
      4.2.2 Spot Linking Between Frames ................................................................. 90
      4.2.3 Tracking Strategy .................................................................................... 96
      4.2.4 Tracking Tags Using Multi-Template Matching ..................................... 98
      4.2.5 Testing the Significance of Template Matching Results .................... 104
### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 Results and Discussion</td>
<td>106</td>
</tr>
<tr>
<td>4.3.1 Performance Test on Synthetic Data</td>
<td>106</td>
</tr>
<tr>
<td>4.3.2 Influence of Initial Values</td>
<td>111</td>
</tr>
<tr>
<td>4.3.3 Influence of the Objective Function</td>
<td>114</td>
</tr>
<tr>
<td>4.3.4 Performance on Biological Data</td>
<td>117</td>
</tr>
<tr>
<td>4.4 Conclusion</td>
<td>123</td>
</tr>
<tr>
<td>5. Conclusions</td>
<td>127</td>
</tr>
<tr>
<td>5.1 Summary of the most important contributions</td>
<td>128</td>
</tr>
<tr>
<td>5.2 Outlook</td>
<td>130</td>
</tr>
<tr>
<td>References</td>
<td>131</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>135</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>136</td>
</tr>
<tr>
<td>List of Publications</td>
<td>137</td>
</tr>
</tbody>
</table>
Zusammenfassung


In dieser Doktorarbeit wird ein Weg eingeschlagen, der diese beiden Probleme mit Hilfe von automatischer, rechnergestützter Bildanalyse lösen will. Die Algorithmen beschränken sich dabei auf die Erkennung, Lokalisierung und das Tracking von einer beschränkten Anzahl fluoreszierender Marker in stark verrauschten 3D Daten. Es wird vorausgesetzt, dass die Größe der Marker unterhalb der Auflösung des Mikroskops liegen und damit das Bild der Marker genau der Point-Spread-Function entspricht.

Die erreichbare Genauigkeit einer Messung in den Bilddaten ist begrenzt durch die Qualität der Daten. Deshalb es wichtig, dass Fehler beim Aufnehmen der Daten vermieden werden oder zumindest genau bekannt sind, damit man sie im nachhinein korrigieren kann. In biologischer 3D Mikroskopie treten vor allem bei dickeren Objekten Verzerrungen im Bild auf wenn sich der Brechungsindex ändert. Im ersten Teil der Arbeit wird deshalb ein Verfahren präsentiert, wo mit Hilfe von
Zusammenfassung

Kalibrationskörpern solche Fehler punktuell gemessen und dann auf das Ganze beobachtete Feld interpoliert werden. Damit lassen sich die Abbildungsfehler in den biologischen Daten im nachhinein zurückrechnen.


Im dritten Teil wird ein Tracking Algorithmus präsentiert, der auf dem Detektions Algorithmus aufbaut. Im Gegensatz zur Detektion, wo die Marker im Datensatz zu jedem Zeitpunkt unabhängig von den anderen Zeitpunkten lokalisiert werden, findet im Tracker eine dynamische Analyse von zwei aufeinanderfolgenden Zeitschritten statt. Dies hat zur Folge, dass erstens die Auflösung weiter erhöht werden kann und zweitens die Analyse stabiler wird. Ausserdem erhält man auf Grund der dynamischen Analyse automatisch auch die Trajektorien der Marker über die gesamte Zeitserie, diese benötigt man für weitergehende Untersuchungen der biologischen Prozesse.
Summary

This thesis reports an integrated software system for the analysis of chromosome dynamics during mitosis of yeast cells. Mitotic chromosome segregation is the process by which replicated sister chromatids are divided equally into two physically separated sets around which daughter cells form. To visualize this process, relevant proteins or protein complexes are labeled with fluorescent tags and their dynamics is observed in live cells under the microscope. It turns out that the experimental manipulation and visualization are not the main difficulty. In fact, the problems are rather due to: i) The limited resolution of optical microscopes. At the onset of mitosis the sister chromatids interact on the nanometer scale, which is far below the optical diffraction limit. ii) The second problem is the amount of data generated. The chromosome segregation in yeast cells lasts for about half an hour during which the movement is recorded. Such an assay yields enormous data sets (>100MB) whose analysis is very cumbersome.

In this thesis we suggest therefore to address this problems with the help of automated computer analysis. The presented algorithms are designed for the detection, localization and tracking of a finite number of fluorescent markers in noisy 3D data. It is assumed that the tags are sub-resolution features and thus their image corresponds to the point spread function of the microscope.

The accuracy of measurements performed in the acquired images is limited by the quality of the data. It is therefore important that errors in the acquisition process are eliminated or at least that the error is exactly known, such that it can be corrected in post processing. In 3D microscopy of biological samples image distortions occur because of changes in refractive index. This effect is especially prominent in thick samples. In the Chapter 2 of this thesis a framework is presented, that measures imaging errors with the help of calibration objects at random positions and interpolates the distortions to the full depth of field. With this information the acquired biological data can be corrected in mathematical post processing.

In Chapter 3 an algorithm is presented for the detection and localization of fluorescent tags. The two main problems addressed are the low signal-to-noise ratio
(SNR) and the signal interference when two or more tags are at very close distances. Using synthetic data it is shown that the algorithm is capable of localizing tags that are separated by half the classical resolution limit even for SNR-values as low as 5 to 10. The performance of the algorithm is confirmed on yeast data for the detection of fluorescent tags.

In Chapter 4 the detection algorithm is extended to a full relative tracker. In contrast to the detection, where tag segmentation is performed at single time points, the tracker includes dynamic information of subsequent time points. This results in a further increase in resolution and more robust tag localization. An additional advantage of the dynamic information extraction is that the full tag trajectories are obtained, which are required for a subsequent analysis and interpretation of the biological data.
1. Introduction

The development of sophisticated computer vision algorithms for automated image analysis and interpretation has started more than 30 years ago. Traditional fields exploiting computer aided evaluation of image data are, for example, in the classical macroscopic engineering fields, robotics, biomedicine, photogrammetry, satellite image interpretation for meteorological and environmental earth survey, surveillance and automated manufacturing.

With much delay these algorithms are now adopted also in cell and molecular biology for the analysis of microscopy data. The reason why these methods have begun to be used in biology much later than in other fields must be located in the complexity of the encountered scenes and the low signal-to-noise ratios obtainable in light microscopy. Whereas in cell biology the analysis of microscopy data is traditionally performed manually and on a qualitative level, the advent of new imaging and visualization techniques as well as electronic imaging devices has changed this dramatically. For example, with the introduction of Green Fluorescence Proteins (GFP) it became possible to label specific proteins in live cells with fluorescent markers and to observe their dynamics over extended time periods. The development of sensitive, high-resolution digital cameras connected to computer systems rendered the recording of the marker dynamics possible, which is essential for a quantitative analysis and interpretation of these processes. Frequently though, large amounts of data are acquired with very high information content and consequently the bottleneck in these cases is not the experiment, but the extraction and analysis of useful information from the acquired data. An accurate and complete manual analysis is cumbersome or in many cases not feasible at all. With the help of highly automated computational image analysis it will be possible to tackle these challenges.

In this thesis I confined myself to the analysis of objects in 3D fluorescence microscopy. A novel framework is presented for microscopy calibration, automatic
detection and tracking of objects with super-resolution in 3D, where objects consist of a finite number of fluorescent sub-resolution features.

The biological application, for which the framework was developed, is a joint project with the laboratory of Peter Sorger (Department of Biology and Biological Engineering, MIT). The biological objective is to identify and analyze the proteins responsible for the generation of tensile force during chromosome segregation in budding yeast *S. cerevisiae*. Mitotic chromosome segregation is the process by which replicated sister chromatids are divided equally into two physically separated sets around which daughter cells form.

Our hypothesis is that it is possible to perform protein function mapping based on trajectory analysis of chromosomes for wild-type and mutant cells. With genetic mutations one can selectively delete specific proteins in the spindle and observe the impact on the chromosome movement to assess the protein function with a mechanical model. A first step that will allow testing this hypothesis is presented in this thesis, where I report the development of a tool for automated extraction of trajectories with high accuracy. Dynamic measurements in living cells are made possible by tagging chromosomes and spindle pole bodies with green fluorescence proteins (GFP) to generate cells carrying four fluorescent tags, and observe the motion of the tags over time using 3D-fluorescence microscopy. The central problem in tag detection and tracking arises with the partial or complete overlap of tag images when the chromosome and spindle markers are separated by distances below the diffraction limited resolution of the optics. All preparations of labeled yeast strains and image acquisition were performed at the Sorger lab.

To set the stage for the rest of this work I start out with a short introduction and definition of important terms and techniques.

**Precision versus Accuracy**

In measurement devices *precision* refers to the number of significant digits of the measured property. In contrast *accuracy* denotes the "correctness" of the property one seeks to measure. To make this clear, the precision of an instrument does not depend on its calibration, i.e. the number of significant digits does not change with the units chosen but only with instrument setup. On the other hand, the correctness (accuracy) of the value for the measured physical property returned by the device depends strongly on an exact calibration of the device. For an analysis of the device
performance it is sufficient to look at the precision, while for real world applications it is imperative that a calibration is performed. This is especially important when working close to the technical or physical limits of the device, since the calibration inherently sets the limits of accuracy.

The accuracy of an instrument is determined with the aid of calibration objects or by measuring an object with different devices and comparing the results. In Chapter 3 and 4 of this thesis where I describe algorithms for the analysis of live cell data it is not possible to introduce a reference object of known size or to measure the data with a different device. The approach I use to determine the accuracy of the presented algorithms, is to generate well-defined synthetic data where I add noise and perform Monte-Carlo simulations.

Resolution and Super-Resolution in Microscopy

Due to diffraction of light in the optics of a microscope the image of an ideal point light source is not a point but an extended complex intensity distribution called the Point-Spread-Function (PSF). This effect limits the resolution of two point-like objects at close distances.

The classical Resolution criterion in light microscopy is defined by the diffraction limit of the optics. The minimal distance resolvable between two equally bright point-like objects in the focus plane (lateral direction) is given by the well-known Rayleigh limit

$$d_{xy}^R = 0.61 \frac{\lambda}{NA},$$  

(1)

where $\lambda$ is the emission wavelength and $NA$ the numerical aperture of the optics. For conventional 3D wide-field microscopy the axial direction is measured by shifting the focus plane in discrete steps and acquiring an image at each position. The formula for the limit in axial resolution is given by (Inoue(1995)):

$$d_z^R = 2 \frac{\lambda n}{NA^2}$$  

(2)

Here $n$ is the refractive index of the observed sample.

Alternatively, resolution can be analyzed in the spatial frequency domain. The frequency characteristics of a microscope is described by the optical transfer function (OTF). It turns out that the OTF has, due to diffraction, a distinct frequency
Chapter 1. Introduction

cutoff $\nu_C$ in each spatial direction. Spectral components belonging to frequencies above the cutoff are not transferred by the imaging system. It can be shown that the product of the cutoff and the Rayleigh limit is a constant. For example, this constant is 1 and 1.22 for incoherent imaging systems with a rectangular and circular aperture, respectively (denDekker & vandenBos(1997)).

Sub-Resolution objects are objects that are confined to a region smaller than the resolution limit ($d_{xy}^R \times d_{xy}^R \times d_z^R$). Consequently, the image of a sub-resolution object is the microscope point-spread-function independent of its true geometry.

The general understanding of Super-Resolution refers to methods, which allow measuring objects separated at distances below the limits above. The Super-Resolution factor $f$ is defined as the ratio of the highest recovered frequency over the cutoff of the OTF $f = \frac{\nu_{SR}^*}{\nu_C}$. In this thesis I confine the term "super-resolution" to one specific category of methods as explained in detail in the following section.

1.1 Resolution extension in microscopy

In conventional light microscopy the limit in resolution is given by the diffraction limit of the optics. As evident from equation (1) and (2) the limits are proportional to the wavelength of the emitted light and thus reach in the best case of high NA optics $d_{xy}^R \approx 200\,\text{nm}$ and $d_z^R \approx 700\,\text{nm}$. With regard to cell biology, this precludes the observation of essential parts of the cellular organization. Radically different methods such as electron microscopy have been developed in order to overcome these limits, yet none of these instruments was versatile enough to render the light microscope superfluous. The light microscope remained the only sensor capable of acquiring information in real-time, non-invasively, and from inside transparent specimens. These factors make it particularly suitable for the observation of live samples, where live means both living biological samples as well as non-biological but moving objects. In particular, fluorescence labeling has been perfected to the stage where single proteins can be visualized in live cells, yet diffraction in the excitation and emission light paths prevents the true resolution of such structures.

There are two fundamentally different categories of methods for resolution extension. In the first category the resolution is extended by modification of the
microscope hardware. Changes of the optical characteristics can narrow the point-
spread-function (PSF) or increase the bandpass of the optical transfer function
(OTF). Typical example of PSF shaping methods are the confocal microscope,
where a pinhole is used to narrow the PSF in axial direction, or the multiphoton
microscope and stimulated emission and depletion (STED) microscopy (Klar, et
al.(2000)), which employ non-linear optical effects in the excitation of fluorophores to
narrow the PSF. Other methods, such as 4Pi microscopy or harmonic excitation light
microscopy (HELM) (Frohn, et al.(2000)) can be better understood in the framework
of OTF bandpass extension. For a comprehensive review of these optical
approaches I refer to Gustafsson(1999).

Although these methods have proven to be extremely powerful in certain
applications they usually come at a cost. They either require long acquisition times,
e.g. due to point scanning of the field, or they deposit high energy in a restricted field
of observation. This limits their use for live cell imaging as the specimen are highly
dynamic and very sensitive to phototoxic effects.

The second category of resolution-extending methods does not change the optical
characteristics of the microscope but attempts to recover spatial frequencies beyond
the OTF cutoff using post- acquisition computational analysis. We reserve the term
“super-resolution” exclusively to the second category of methods. The prefix “super”
stimulates the notion of breaking the limit of a given system. This applies to the post-
recovery concept but not to the resolution extension by an optical modification of the
microscope. The latter represents a pushing but not breaking of limits.

The key to achieving super-resolution is the incorporation of prior knowledge during
object reconstruction. Such knowledge constrains the search space during image
analysis and allows the extrapolation of object-space frequencies beyond the cutoff
of the OTF (Pask(1975)). Prior knowledge can be introduced at the level of the raw
signal, for example, in the non-negativity constraint used during iterative
deconvolution of fluorescence image stacks (Carrington, et al.(1995)), or on the
higher, more symbolic level of visual information, where geometric and dynamic
models of the observed specimen are included (Danuser(2001)).

It is important to realize that the second category is conceptually different from the
first and is therefore complementary. A resolution limit reported by one of the
extended resolution microscopes can always be improved further by the application
of super-resolution post-processing.
1.2 Theory of Super-Resolution

The theoretical foundation of super-resolution schemes is based on information theory and has been studied for more than 50 years. The main theorem super-resolution theory is built on, is the invariance of information capacity formulated by Shannon for communication systems (Shannon & Weaver(1949)). Shannon’s information capacity theorem states that the upper bound of a channel capacity of a continuous channel of bandwidth $b$, perturbed noise is given by

$$C = b \cdot \log(1 + \frac{S}{N}),$$  

(3)

where $S$ is the average transmitted signal power and $N$ the average noise power.

In 1955 Toraldo di Francia was the first to apply the notion of information capacity to determine the resolving power of an optical system (Toraldo di Francia(1955)). In the following years the theory was further extended and analyzed by Lukosz(1966) and Cox & Sheppard(1986). They showed that the mathematical formulation of the capacity of an optical system found in these references is given by

$$C = \frac{1}{2} \left( \prod_{j=1}^{3} (2I_j v_j^C + 1) \right) (2T v_T^C + 1) \log(1 + \text{SNR})$$  

(4)

In the case of a microscope there are four channels present: three spatial channels and one temporal channel. The parameters $v_j^C$ represent the cutoff frequencies of the spatial channels. In most microscopes the cutoffs of the two lateral directions are equal ($v_1^C = v_2^C$), but different from the axial cutoff $v_3^C$. The variable $I_i$ denotes the observation window of channel $i$. The product $I_1 \cdot I_2$ is called the field of view and $I_3$ the depth of field. In a time series of 3D observations the parameter $T$ denotes the total observation time and the bandwidth $v_T^C$ represents the frame rate. The last term in equation (4) addresses the stochastic part of the image. It relates the signal-to-noise ratio (SNR) to the total information capacity. The SNR is defined in equation (3) as the average of transmitted signal power over the average noise power. The capacity
increases with the logarithm of the SNR and vanishes as expected for SNR=0. The
interesting implication of equation (4) is, that even with limited bandwidth the
capacity can increase without bounds. Therefore, it is not the optical bandpass but
the noise that limits the capacity. In Figure 1 the information capacity of an optical
system is displayed as a function of SNR. The SNR range represents typical values
found in fluorescence microscopy. Note that especially at the lower end a small
increase in SNR implies a large increase in information capacity. This implies that in
this region the achievable super-resolution is mainly limited by the SNR, which is in
perfect match with the results shown in theory and experiment in Chapters 3 and 4.

Figure 1
Information capacity of an optical system with fixed bandpass as a function of the SNR.
The means to improve resolution in a multi-channel system is the use of unsaturated channels. The information capacity theorem gives an upper limit to the transferable amount of information of all channels. If the bandwidth of one channel is not fully used, additional information up to the amount of the free capacity can be transmitted. This leads to an increase of the bandwidth in the other dimensions of the system. In a super-resolution scheme the additional information is present in the form of prior knowledge, which can be recovered in image analysis. Unused information capacity arises whenever a priori knowledge of the scene can be used to constrain the image analysis.

There are different ways to introduce prior knowledge. Inverting the image formation process mathematically using the PSF is known as deconvolution. Super-resolution in deconvolution can be achieved by introducing the prior knowledge that the intensities in the reconstructed image are constrained to positive values. This is a computationally very expensive procedure and has only become possible with the development of computer-based image analysis.

Other examples to introduce prior knowledge are: i) Information of the object geometry is taken into account. ii) Information of the object dynamic and deformation is known, e.g. the direction of motion, boundaries of movement steps and velocity.

In this thesis I combine prior knowledge on several levels. In Chapter 2 and 3 we make use of the knowledge of the PSF of the system, while in Chapter 4 information of the object dynamics is included.

1.3 Outline of the thesis

The focus in this thesis is on the following research questions:
- What is the ultimate tracking resolution achievable for three-dimensional low SNR fluorescence microscopy images and model-based image processing?
- How accurately can the signals be localized?

In chapter 2 this is addressed in terms of the measurement of geometric deformation of spherical objects, in chapter 3 in terms of absolute position
measurements of sub-resolution features and in chapter 4 for relative
displacement measurements of sub-resolution features.

- How robust can software for image analysis and understanding of complex
  scenes be programmed with regard to a completely automated and reliable
  extraction of cell biologically relevant data?

Chapter 2 presents a framework for the calibration of focal shifts and spherical
aberrations for 3D microscopy. This is especially important for thick samples of up to
100 microns where shifts of 1 micron or more can occur leading to uncontrollable
errors in quantitative analysis. The algorithm presented, allows in a one-step-
procedure for a simultaneous calibration of specimen induced focal shifts and depth-
dependent spherical aberrations.

In Chapter 3 an algorithm for the detection of fluorescent sub-resolution tags in 3D
microscopy images with super-resolution is presented. The algorithm is described
and its performance analyzed on yeast data for the detection of fluorescently labeled
chromosomes and spindle poles.

Chapter 4 extends the work of Chapter 3 to a novel approach for 3D tracking of
fluorescent objects, consisting of a finite number of fluorescent sub-resolution tags.
The performance improvements of the tracker in comparison with the detector are
analyzed in detail.

Chapter 5 summarizes the contributions of the thesis and concludes with an outlook
to future work.
Chapter 2

2. One-step Calibration of Specimen Induced Focal Shifts and Spherical Aberration for Quantitative 3D Microscopy: Approach and First Results

Dominik Thomann, Edgar Stüssi and Gaudenz Danuser Laboratory for Biomechanics at the Swiss Federal Institute of Technology, Wagistrasse 4 8952 Schlieren, Switzerland

Platform Presentation at BIOS 2001, SPIE; San Jose, CA, January 20-26, 2001;

Relevance and own contribution

The following paper describes a framework for the calibration of focal shifts due to mismatched refractive index of the sample and spherical aberrations in 3D fluorescence microscopy. With regard to high accuracy distance measurements at the limit of optical resolution, which is one of the prime elements of this thesis, it is imperative to eliminate such distortions of the image geometry and quality as much as possible. In microscopy data of thick biological samples (up to 100 μm) relative shifts from top to bottom of more than 1 μm can occur. When measuring distances significant errors are introduced since the length scale varies across the sample. Such distortions are not recognizable by visual inspection and can therefore not be corrected if no additional information about the real object sizes is available. With the scheme presented here, it is possible to measure such shifts and distortions and interpolate a distortion field, which can be used to correct the acquired raw data in post processing.

The core idea is to use a field of 3D distributed beads of which the size but not the absolute positions are known. Graded distortion of the bead size is a measure for the overall shifts throughout the sample. To measure the apparent changes in bead size across the field it was necessary to develop an algorithm that utilizes a 3D model of the bead and to fit it to the measured image data, where an axial scaling factor was one of the unknown parameters to be estimated. We refer to this mathematical process as object-constrained deconvolution.

The paper is co-authored by my two supervisors at that time, Prof. Dr. E. Stüssi and Prof. Dr. G. Danuser.

My contributions consisted in:
- Preparation and imaging of calibration samples using a confocal light microscope
- Design and evaluation of a simple mathematical model for the approximation of a depth dependent point-spread-function
- Design and implementation of the object-constrained deconvolution scheme including a specific, highly efficient convolution algorithm
Abstract

With the advent of laser scanning confocal and multiphoton microscopy, 3D life tissue characterization has been rendered possible. This involves the restoration of thick section images (in the depth range of 100 microns) of biological samples. In contrast to thin samples new effects become important when imaging thick samples: Because of changes of the refractive index across the specimen or the embedding medium, strong (>1 micron) focal shifts and spherical aberration occur and scattering effects get more prominent. For tissue mapping it is essential to correct for such aberrations and distortion effects. In this paper, we propose a calibration framework, which allows us to determine thick section focal shifts and spherical aberrations in a one-step-procedure. Gradients in the focal shift induce a scaling in the Z-direction of the observed sample. A second effect arises with depth-dependent spherical aberration. We model our microscope as a linear-shift-NON-invariant system (LSNI) where multiple depth classes are assigned distinct point spread functions (PSFs). We measure these two effects in a 3D sample of randomly distributed fluorescent focal check beads. The beads are embedded in gelatine, a medium with high resemblance to real biological tissue. The PSF is approximated by a mathematical parametric model. While estimating the parameters of the PSF with object-constrained deconvolution, we track depth dependent changes of the observed bead diameter. This allows us to determine the gradient of the focal shift across a thick section. By numerical integration along the optical axis we obtain the focal shift values as a function of sample depth. In the end, these values will be applied as a correction to compensate depth distortions in the tissue images.
2.1 Introduction

With advent of laser scanning confocal and multiphoton microscopy, 3D life tissue characterization has been rendered possible. This involves the restoration of thick section images (in the depth range of 100 microns) of biological samples. It is well-known that for thick section imaging a mismatch between the nominal refractive index, i.e. the refractive index which the objective lens is designed for, and the actual refractive index of the specimen or embedding causes two types of image degradations (de Grauw, et al.(1999), Hell, et al.(1993), Sheppard & Torok(1997), Torok, et al.(1997)):

1. Focal shifts, i.e. point-like objects lined up at equal spacing along the optical axis (Z-axis) appear in a 3D image stack as a non-equidistant row of points. Notice that this effect does not affect the visual quality or resolution of the image but merely distorts the geometry of the point arrangement.

2. Spherical aberrations, i.e. point-like objects do not appear as diffraction limited spots, but are further blurred. In thick sections with refractive index mismatches the amount of blur is depth dependent. Image formation in such systems can no longer be described by the standard linear shift invariant (LSI) formalism. Instead, an imaging model is required where the point spread function (PSF) varies with the Z-position, resulting in a linear shift non-invariant (LSNI) system.

For accurate geometric measurements in thick sections, focal shifts must be compensated. Our ultimate goal is to implement a quantitative confocal microscope for 3D tissue mapping and monitoring of cortical and internal tissue deformations. The goal of the present study is to establish a calibration procedure which allows us to estimate the focal shifts in a thick section of a test medium that mimics biological tissue. These values will be applied as posterior correction factors to Z-stacks taken from a real tissue. We have chosen gelatine (10%(w/v) SIGMA gelatin type A) as our test medium. Gelatine consists of a dense meshwork of collagen and other filamentous proteins interspersed with an aqueous phase. Therefore, it resembles closely the structure of a typical biological tissue that contains cells and extra-cellular matrix.
In principle, the focal shifts in such a medium could be determined by measuring depth-dependent distance distortions in a point distribution with known reference coordinates. This would correspond to the calibration strategy used in classical photogrammetry. The major difficulty with this approach arises with embedding a 3D calibration standard in a test volume of e.g. 1000×1000×100 μm³. To circumvent this problem we propose a differential calibration technique. The principle of this method is illustrated in Figure 2. Instead of relying on a point set with known absolute coordinates, we employ a volume with randomly distributed focal check beads (Molecular Probes Focalcheck™ triple stained). The beads have a reference radius \( \bar{p} \). Focal check beads are coated with a thin, fluorescent layer generating a perfectly spherical shell as a light source in the object space. We select the bead radius dependent on the numerical aperture of the objective. As a rule of thumb, the radius should be two to three times larger than the diffraction limited, axial resolution

\[
\delta_{\text{axial}} = 2\frac{\lambda n}{NA^2}
\]

where \( n \) is the specimen refractive index, \( \lambda \) is the wavelength of light, and \( NA \) denotes the numerical aperture. Consequently, the microscope optics will resolve the beads as a hollow spheres. As depicted by Figure 2, due to focal shifts the top and bottom poles as well as the (invisible) centers of the spheres are not observed in the associated actual focal planes (AFPs) but in nominal focal planes (NFPs). As a result the center of a sample bead \( i \) is shifted downwards by the amount \( s(i) \). Notice that there is a gradient in focal shift across the bead. For the bead \( i \) the shift of the top pole \( s_t(i) \) differs from the center shift \( s(i) \) as well as from the bottom pole shift \( s_b(i) \). This leads to an apparent deformation of the bead. The nominally spherical shell is perceived as a rotation symmetric ellipsoid stretched (or shortened) in Z-direction. Instead of the radius \( \bar{p} \), the bead image has an extension \( \rho(i) \) in Z-direction. Notice that this effect is not related to the inherent resolution anisotropy of microscopes. What has been mentioned so far is a simple specimen induced Z-scale distortion that is completely independent of the PSF shape. There is an immediate relationship between bead deformation and focal shift which can be formalized as follows:
Chapter 2. One-Step Calibration of Focal Shifts

\[ s_b(z_i) - s_i(z_i) = \rho(i) - \bar{\rho} = \frac{ds(z_i)}{dz} \bar{\rho}. \]

**Equation 1**

In a linear approximation and after introduction of the relative scaling factor \( \zeta(z_i) = \rho(i)/\bar{\rho} \), Eq. 1 is written as:

\[ \zeta(z_i) - 1 = \frac{ds(z_i)}{dz}. \]

**Equation 2**

Eq. 2 suggests that the relative Z-scaling factor \( \zeta(z_i) \) of a bead located at \( z_i \) provides a direct measure for the focal shift gradient at this position. Both \( \zeta(z_i) \) and \( z_i \) are variables that can be measured in a 3D confocal image stack. Given \( \zeta(z) \) values for infinitely many beads we can calculate the focal shifts as a function of the stack depth \( d \)

\[ s(d) = \int_0^d (\zeta(z) - 1) dz, \text{ subject to } s(0) = 0. \]

**Equation 3**

In practice, only a finite number of \( \zeta(z_i) \) measurements will be available. They can be sorted in ascending order \( z_{i-1} \leq z_i \leq z_{i+1} \), with \( z_i = 0 \) and \( s(z_i) = 0 \). In this case, we can compute the focal shift at any position \( z_i \) as

\[ s(z_i) = \sum_{k=1}^{i-1} (\zeta(z_k) - 1)(z_{k+1} - z_k) \]

**Equation 4**

Focal shifts at positions between \( z_i \) and \( z_{i+1} \) must be interpolated. Notice that this method for focal shift calibration relies only on the premise that the shape of the bead is truly spherical.
Chapter 2. One-Step Calibration of Focal Shifts

Figure 2

Relationship between apparent bead scaling and focal shifts induced by mismatches of refractive index. A calibration of focal shifts is possible using focal check beads. We propose a differential method where the focal shifts are derived from measurements of the depth-dependent, vertical bead scaling $\zeta(z) = \rho(z)/\overline{\rho}$. The parameter $\rho$ describes the apparent bead extension in Z-direction, $\overline{\rho}$ the nominal radius of the bead. The later can be deduced from a lateral slice through the equator of the bead image. The figure indicates that the Z-scaling of a bead results from a change of the focal shift across the bead. The absolute focal shift can be restored by solving a differential equation. Measurements of the Z-scaling of beads at various Z-positions are needed as input quantities. See text for further details.

The extremely weak prerequisite towards a calibration standard make the implementation of the procedure straightforward. Particularly, we emphasize that
No prior information about the bead positions is necessary. The calibration can be performed using any random distribution of focal check beads. In some special applications where we can afford to permeate the biological target tissue with a focal check bead solution, we might even think of running the focal shift calibration simultaneous to the actual tissue mapping.

No prior information about the physical bead radius is necessary. The entire procedure relies on the ratio between apparent vertical and lateral bead extension. The procedure can be run on distributions of beads with different sizes. This turned out to be critical as sendings of focal check beads exhibit noticeable variations in bead diameter. Exact knowledge of the absolute bead size is not necessary. Obviously, the key to our focal shift calibration is a precise tracking of depth-dependent changes in the vertical bead extension relative to the lateral bead diameter. As mentioned in the beginning of this section the image of a bead is not only distorted but is degraded by depth-dependent spherical aberration. For thick sections aberrations get very pronounced making the accurate determination of the bead scaling impossible. Recently, very powerful, adaptive optics for real-time reduction of spherical aberration have been reported (Booth & Wilson (2000)). Other schemes pursue posterior compensation of spherical aberration using blind deconvolution. In general, this approach tends to be unstable because of the large number of unknowns involved in the simultaneous estimation of PSF and object fluorescence. The problems get even worse when processing thick image sections where the PSF varies with depth and photon scattering and light adsorption deteriorate the image quality. Over the past years, it has been shown that blind deconvolution can be made robust by the application of parametric models of the PSF (Markham & Conchello (1999)). This reduces the number of unknowns, thus increasing the redundancy of the underlying optimization problem. In our specific case, we can do even better, since we have rigid prior knowledge of our object. We can further stabilize the deconvolution by constraining the object fluorescence distribution to the boundary of a geometrically distorted sphere.

Following the approach of object-constrained parametric blind deconvolution we have investigated ways to estimate the parameters of an aberrated PSF at the same time as the relative Z-scaling factor. In this paper we establish the mathematical foundation of our framework, evaluate its performance in comparison to theoretical models of focal shift and spherical aberration, and present preliminary results of
simultaneous estimation of PSF aberration and bead scaling parameters. In detail, the paper is organized as follows: In Sec. 2 we introduce a parametric model for the PSF including terms that account for depth dependent spherical aberration. In Sec 3 we present a framework for simultaneous numerical estimation of the PSF parameters and the bead scaling. In Sec. 4 preliminary results are shown. We tested the robustness of the framework against noise perturbation and the convergence of the estimator using synthetic data.

2.2 Parametric Model for the Point Spread Function

Any deconvolution procedure requires precise knowledge of the PSF. The PSF can either be determined experimentally, approximated in a theoretical model, or estimated simultaneously with the object fluorescence in a blind deconvolution. As discussed in the Sec. 2.1, blind deconvolution can only be performed reliably if the PSF is described with a parametric model. In the literature, two classes of parametric models have been suggested: The first class relies on a set of equations established from a precise physical description of the optical system. The parameters address specific properties of optical components and therefore allow for a physical interpretation. The second class deals with imaging as a pure signal filtering problem where the PSF takes the role of the impulse response. The PSF parameters have no physical meaning but are determined such that the mathematical model matches the effective image of a point-like object as close as possible.

Over the past years, several physical PSF models have been suggested which can explicitly handle specimen induced aberrations and focal shifts in thick sections with mismatched refractive index (Torok, et al.(1997)). We have examined two approaches. The first one follows the physical model proposed by Török et al. based on Debye approximation (Torok, et al.(1995)), whereas the second approach by Hell et al. (1993) uses the Huygens-Fresnel principle. In a recent paper Egner and Hell (1999) have shown the equivalence of the two approaches. We have implemented both, yet in this paper we refer to the Török model as the only physical PSF model. This model includes parameters for the objective lens, the refractive indices of immersion medium and specimen, and the depth of the plane in focus.
Figure 3 illustrates the depth-dependence of the PSF based on intensity level curves. For points deeper in the section, the PSF gets dramatically stretched in Z-direction. The result indicates that for stacks with a thickness of more than 10 \( \mu m \) the conventional LSI model for a microscope must be replaced by a LSNI formalism. In principle, physical PSF models of the kind shown in Figure 3 would be appropriate to perform blind deconvolution with a parametric approach. Yet, the computational complexity prohibits its use in iterative deconvolution algorithms. For instance, Török’s model requires multiple numerical integration steps (Torok, et al. (1995)), which have to be repeated after every adjustment of one of the many optical parameters. To avoid this difficulty we have sought a numerically less expensive strategy. Based on the intensity distribution from the physical model and actual PSF.
measurements with a confocal microscope we have derived a much simpler mathematical function that suitably generates PSF-like intensity distributions. We lose the connection between the model parameters and the optical components defining the PSF characteristics. Yet, we gain several orders of magnitude in computational speed. Our mathematical approximation is given by the following equation:

\[ h(x,y,z) = \exp \left( -\frac{1}{2} \left( \frac{(x-p_x)^2}{\sigma_x^2} + \frac{(y-p_y)^2}{\sigma_y^2} + \frac{(z-p_z)^2}{\sigma_z^2} \right) \right) + \epsilon(r,z,\xi), \]

Equation 5

with \( r = \sqrt{x^2 + y^2} \). The exponential term describes an ellipsoidal, Gaussian intensity distribution that is rotationally symmetric around the Z-axis. It reflects the typically anisotropic shape of a confocal, diffraction-limited PSF. For the 2D case Santos and Young (Santos & Young 2000) have recently shown that the Gaussian distribution is very accurate in approximating the more complicated Bessel functions involved in PSF modeling. Our choice is strongly motivated by this result. Here we extend it to a 3D PSF representation. The basic exponential term in Eq. 5 defining the Gaussian PSF body is supplemented by a perturbation term \( \epsilon(r,z,\xi) \). The perturbation term accounts for the differences between the real, distorted and aberrated PSF and the idealistic model of a diffraction-limited PSF. Such differences are chiefly associated with spherical aberration and, therefore, become increasingly important for PSF modeling at positions deep inside the specimen.

In a first step we applied a simple polynomial in \( r \) and \( z \) with the coefficients \( \xi = [a_0, a_1, \cdots, a_N, b_0, b_1, \cdots, b_N, c_0, \cdots, c_N] \):

\[ \epsilon(r,z,\xi) = \sum_{n=0}^{N} (a_n r^n + b_n z^n + c_n (rz)^n) \]

Equation 6

We evaluated the quality of our PSF representation relative to the physical model as well as relative to real PSF measurements. A theoretical PSF was calculated for an oil immersion objective (NA = 1.1, refractive index of immersion oil \( n_i = 1.518 \)) at a position 10\( \mu \text{m} \) inside the medium (refractive index of specimen medium \( n_s = 1.33 \)).
Figure 4

Approximation of a PSF using the mathematical model in Equation 5. The top row shows the approximation of a synthetic PSF generated with the formalism by Török. a) input PSF, scale bar 1 μm; b) fitted PSF; c) residuals displayed for XZ cross-sections at positions y = 1, 2, 3, 4 pixel; d) histogram of the residuals. Bottom row e) through h): The same panels derived from a fit to a real confocal PSF measurement. The data has been taken with a 100X/1.4 oil immersion objective on a Zeiss LSM510 confocal microscope. Scale bar 500 nm.

Figure 4a displays the model PSF sampled in a $7 \times 7 \times 21$ volume with a voxel of $0.2 \times 0.2 \times 0.2 \mu m^3$. For visualization in this paper, we stretch PSF values to an 8-bit gray-value scale, i.e. to values between 0 and 255. This model PSF was fitted by our parametric description denoted in Eqs 5 and 6. The data fitting was accomplished by non-linear least squares (LS) optimization, similar to the approach used subsequently for blind deconvolution (see Sec. 3).

We examined several formulations of perturbation terms. In the unconstrained case, all coefficients $a_n, b_n, c_n, \forall n = 0 \ldots N$ were estimated. Generally, this formulation does not achieve stable parameter convergence. The coefficients have to be constrained
to certain limits. In the simplest case, we set \( c_n = 0, \forall n = 0 \ldots N \), i.e. mixed terms in \( r \) and \( z \) were oppressed. This step was sufficient in order to guarantee robust solutions for all the remaining coefficients. The mean residual error of the PSF fit amounts on average to 3.6% of the PSF value range. Figure 4b displays the mathematical approximation of the physical PSF using a perturbation function of degree \( N = 2 \). The residuals of the fit are shown in Figure 4c. A series of XZ cross-section panels is plotted for the positions \( y = 1, 2, 3, 4 \) pixel. The histogram in Figure 4d illustrates the residual distribution over the entire PSF volume. It exhibits a zero mean distribution confirming a data fit of reasonably good quality. The residual panels, however, reveal that some details in the PSF shape are not adequately represented by a low order mathematical function. Therefore, we increased the degree of the perturbation term up to \( N = 10 \). As expected, the mean square residual drops with a higher degree of freedom. Yet, the improvement evolved rather slow, suggesting that many of the additional parameters were non-significant. We concluded, that a relatively low order perturbation function is sufficient to represent aberrated PSFs. The application of higher perturbation functions must be supported by a scheme for automatic elimination of non-significant and non-determinable coefficients (cf. Discussion in Sec. 5. Such parameters suffer from high mutual correlations weakening the stability in estimating the whole system.

The same low order perturbation term was applied to fitting a real PSF (Figure 4 bottom row). We measured the fluorescence intensity distribution of sub-resolution beads (Molecular Probes, 90 nm diameter, 543 nm excitation wavelength) embedded in gelatine. We used a 100X/1.4 oil immersion objective on a Zeiss LSM510 confocal microscope. The mean residual error amounted to 3.5% of the image intensity range. This number is comparable to the one obtained from fitting synthetic models without noise. This indicates the high quality of both the Török physical model and our mathematical approximation.

To further improve the perturbation model we have imposed boundary conditions to the polynomial in Eq. 6. A first condition is motivated by the fact that even aberrated PSFs fade towards the border of the PSF volume. Consequently, we force the perturbation term to zero for all the boundary voxels. The second condition arises from the fact that the Gaussian body represents a real PSF very accurately in the
region. This implies that the perturbation term should also vanish in the center voxel of the volume. We can readily fulfill this condition by setting \( a_0 = b_0 = c_0 = 0 \).

The new, constrained perturbation function is formulated in normalized coordinates \( \bar{r} \) and \( \bar{z} \), with \( 0 \leq \bar{r} \leq 1 \) and \( -1 \leq \bar{z} \leq 1 \):

\[
e(\bar{r}, \bar{z}, a_n) = (\bar{r} \pm 1)(\bar{z} \pm 1)(\bar{z}^2 \bar{z}^3 \ldots) \begin{pmatrix} 0 & \ldots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \ldots & a_{NN} \end{pmatrix} \begin{pmatrix} \bar{r} \\ \bar{r}^2 \\ \bar{r}^3 \ldots \end{pmatrix}^T
\]

Equation 7

The function contains \( N^2 \)-1 coefficients. The proposed constraints tend to strengthen the fitting of noisy data. In this case, the boundary conditions prevent estimation instabilities induced by noisy voxels at the border of the PSF volume. For fitting synthetic, noise-free data, the difference between the constrained and unconstrained PSF models is insignificant.

The Z-elongation of the PSF which is particularly pronounced in data with specimen-induced spherical aberration distorts the microscope image enormously. Figure 5a displays the aberrated image of a focal check bead embedded in gelatine. The diameter of the bead is 15 \( \mu \text{m} \). The confocal stack fails to visualize the top and bottom poles of the bead. The vertical extension of the PSF blurs the signal to a degree where the structure is no more perceivable by eye. The same effect is reflected by convolving a synthetic fluorescence distribution of a focal check bead with our mathematical model of the PSF (Figure 5b). Comparing the two pictures, we observe a significant amount of additional fluorescence inside the real image of the spherical shell. This effect is due to scattering which is an additional problem of thick section imaging that can heavily deteriorate the image data quality. Scattering is not directly addressed by our PSF model. Yet, in particular the low order perturbation coefficients will tend to pick up some scattering components.
Chapter 2. One-Step Calibration of Focal Shifts

Figure 5
Aberrated images of a focal check bead. a) Real image of a 15 μm-diameter bead embedded in gelatin. The stack was imaged with a 50x/0.8 objective on a Zeiss LSM510 microscope. Scale bar 10 μm. The elongation in z-direction and the missing top and bottom of the bead are distortion and aberration effects due to mismatched refractive index. b) The same effect is observed in a synthetic bead image where the PSF is modeled in according to Eq. 5.

2.3 Object-constrained parametric blind deconvolution

Image formation in a LSNI-type microscope is formulated by a depth-dependent convolution:

\[ g(x) = \int_{\mathbb{R}^3} h_z(x - y) f(y) dy = h_z \otimes f, \]

Equation 8

where \( x \) and \( y \) denote 3D vectors. The object fluorescence distribution is referred to by the function \( f() \). The notation \( h_z \) indicates that the PSF consists of a set of functions associated with a specific depth class \( Z_i \) with \( i = 1 \ldots M \). The objective in blind deconvolution is to find \( h() \) and \( f() \) simultaneously, given the fluorescence image \( g() \). This problem is inherently ill-posed. Some remedy is provided by the non-negativity constraint limiting both \( h() \) and \( f() \) to \( \mathbb{R}^+ \). However, in general this will not suffice to solve the problem.

As discussed in Sec. 1, we have two pieces of information which can support the deconvolution:

1. We estimate a parametric function \( h_z(p_1, p_2, p_3, \sigma_x, \sigma_z, \xi) \) with only a few parameters instead of a generic distribution \( h_z() \).
2. We have exact prior knowledge of the specimen geometry. Therefore, we can also constrain the object fluorescence distribution \( f() \) to only a few parameters. For a focal check bead, the object geometry is determined by the
radius $\rho$ and the relative scaling factor $\zeta$. The latter accounts for the apparent Z-scaling induced by a focal shift gradient across the bead diameter. As a matter of fact, $\zeta$ is the only parameter of interest in our differential focal shift calibration procedure (cf. Sec. 1). All the other parameters are introduced in our formalism in order to ensure unbiased estimation of the relative Z-scaling.

We apply least square (LS) optimization to estimate the parameters of PSF and object fluorescence distribution. To simplify the notation for the rest of the paper we concatenate all parameters to one vector $\omega$. The number of parameters contained by $\omega$ depends on the degree of the perturbation term in the PSF model. Given a degree $N+2$, this term relies on $N^2-1$ perturbation function coefficients. The Gaussian ellipsoid requires additional 5 parameters; plus the 2 parameters used to define the focal check bead geometry yields a total number of $N^2+6$ parameters to be inferred. The goal of our object-constrained parametric blind deconvolution is the estimation of the parameter vector $\omega$ subject to

$$\hat{\omega} = \min_{\omega} \sum_{x} (I(x) - I_m(x, \omega))^2$$

Equation 9

where $I_m(x, \omega)$ denotes a model image of the focal check bead given the parameter vector $\omega$. We can compute $I_m = h_{z_1} \otimes b$, i.e. by convolving the model surface of a Z-stretched bead $b(\rho, \zeta)$ with the parametric model of the PSF $h_{z_1}(p_1, p_2, p_3, \sigma_y, \sigma_z, \zeta)$. $I(x)$ represents the observed confocal image of a focal check bead located at the Z-position of the depth class $Z_i$.

In our current implementation we search the minimum in Eq. 9 by a Gauss-Newton gradient-based approach. This technique improves the parameter vector step-wise starting with an initial guess $\omega^0$. In the iteration step $k$ we obtain an adjusted parameter vector $\omega^k = \omega^{k-1} + \Delta \omega$ by solving the equation system

$$A \Delta \omega = \delta.$$  

Equation 10
The vector \( \delta = I(x) - I_m(x, \omega^{k-1}) \) contains the voxel-by-voxel intensity difference between the real image and the best guess of a model image at that stage. The matrix \( A \) represents the Jacobian \( \nabla_{\omega} I_m \).

Computationally, the Jacobian \( A \) is the most expensive step of this procedure. It requires for each iteration step \( N^2 + 8 \) convolutions. Recall that the column \( A_j = \nabla_{\omega_j} I_m \) for any of the PSF parameters is calculated by

\[
A_j = \frac{dh_z}{d\omega_j} \otimes b.
\]

Equation 11

The derivative of the PSF with respect to the parameter \( \omega_j \) can be calculated analytically. For the PSF we have \( N^2 + 4 \) parameters, i.e. \( N^2 + 4 \) convolutions of the type Eq. 11. For the two bead parameters the associated columns of the Jacobian have to be computed as

\[
A_j = h_z \otimes \frac{db}{d\omega_j}.
\]

The term \( \frac{db}{d\omega_j} \) describes the variation of the object fluorescence distribution with a change in either the bead radius or relative scaling factor. As there are no straightforward analytical expressions for this term we approximate \( A_j \) by

\[
A_j \approx \frac{h_z \otimes b(\omega_j + \Delta \omega_j) - h_z \otimes b(\omega_j)}{\Delta \omega_j}.
\]

Equation 12

Therefore, we need to run two more convolutions for each of the two bead parameters summing up to \( N^2 + 4 + 4 \) convolutions in total. Note that both expressions Eq. 11 and Eq. 12 boil down to the problem of convolving a general function with the infinitely thin surface of a sphere. Instead of using a function for general 3D matrix convolution, which is extremely time consuming, we have developed a very efficient, specialized convolution method that takes all the object symmetries into account. The use of this function speeds up the estimation process by several orders of magnitude.

The iterative LS estimation according to Gauss-Newton has turned out to be unstable, if we start with coarse guesses for the bead radius and scaling. We had to
modify the algorithm in the following ways: First, we keep the bead radius as well as all the PSF parameters fixed and adjust only the relative bead scaling until the changes in this parameter get smaller than a certain threshold. Second, we sequentially add the other parameters. The iteration is continued until all the parameter updates $\Delta \omega$ fall below the user specified numerical resolution. Despite we achieve reasonable convergence and estimation precision for both synthetic and real data, we still experience difficulties if the initial parameter guesses are far off the final solution. For the moment we cannot obtain stable solutions for all the PSF parameters at the same time as the bead radius. Since we have strong prior knowledge of the latter, we have excluded this parameter from our current deconvolution framework. We will test more global search strategies like Levenberg-Marquart optimization to strengthen the method.

2.4 Simulation and preliminary Results

2.4.1 Initial parameter guesses

Non-linear, local optimization as applied in our blind deconvolution requires initial guesses for all the parameters. We start the LS optimization with values deduced under the following considerations:

Since the PSF perturbation parameters are impossible to guess a priori they are always initialized to zero. The parameter $\sigma_{xy}$ defining the width of the Gaussian PSF body can be guessed by measuring the blur of the bead surface imaged in the equatorial slice of the bead image. Since the signal in this slice is very strong even at large depths, the initial values of $\sigma_{xy}^0$ turned out to be very accurate. In principle, the same procedure could be applied to determine $\sigma_z^0$ using the XZ or YZ center slice. However, for beads deep inside the thick section spherical aberration and scattering prevent a stable assessment (cf. Figure 5a). For this reason we set $\sigma_z^0 = \kappa \sigma_{xy}^0$, with $\kappa$ being the theoretical ratio between axial and lateral confocal PSF extension.
The same reservations apply to the estimation of the values $\rho^0$ and $\zeta^0$. $\rho^0$ can be deduced very precisely from the fluorescent ring occurring in the equatorial slice. For beads close to the top of the thick section, also $\zeta^0$ can be derived robustly. One can measure the distance between the top and bottom pole imaged to the XZ or YZ center slice and relate it to $\rho^0$. For beads at deeper positions the pole images fade dramatically such that a measurement of the pole-pole distance gets impossible. For these beads we have to set up a depth-dependent look-up table for $\zeta^0$.

### 2.4.2 Convergence

We tested the algorithm described in the previous section with synthetic data. We generated focal check bead fluorescence distributions with a lateral diameter of 31 pixels. To simulate focal shift effects we stretched the axial diameter by a relative scaling factor $\zeta$. In a first experiment we created a series of PSFs and convolved them with the specimen to create input images. Since all system parameters were known in this situation we were able to validate the correct convergence of the algorithm.

Results from such a performance test are presented in Figure 6. To fit the PSF aberration we applied a second order perturbation polynomial without mixed $rz$ coefficients. The panels a) through d) display the XZ center slice of the bead image stack. Panel a) contains the noisy input data (SNR = 31). Panel b) shows the model image $I_m(x, \omega^0)$, i.e. the image of a focal check bead that is obtained when applying the initial parameter guesses. Panel c) shows the same image after the first iteration, thus $I_m(x, \omega^1)$. Panel d) presents the model image computed based on the final parameter estimates $\hat{\omega}$, thus $I_m(x, \hat{\omega})$. Convergence was reached for this case after 26 iterations. The visual difference between the pictures c) and d) seems fairly minor. This indicates that the major parameter adjustment is accomplished in the few first iterations. This statement is supported by the convergence curve for parameter $\sigma_z$ depicted by panel e). $\sigma_z$ is the parameter which dominates the visual appearance of the vertical slices. Panels f) and g) compare the value distributions of the true and estimated PSF. The good quality of the fit is further demonstrated by panels h) through k). The first of these panels displays the true noise distribution that was
superimposed to the synthetic bead image. Panel i) shows the estimated noise field. There is no visually appreciable difference between them, suggesting a high quality fit. Nevertheless, we further examined the noise field for hidden systematic effects. Panel k) presents the difference image between panels h) and i). Indeed, there is a faint systematic effect induced by inaccurate fitting of mainly the bead scaling parameter. However, notice that the maximum value of this map is 1 gray-value, i.e. it corresponds to the quantization level of this data. We conclude that there is no significant difference left between the original data and the fitted model.

Figure 6
Object-constrained parametric deconvolution of a synthetic focal check bead image. XZ-center slices of: a) noise perturbed input data; of the model image of the bead after b) setting initial parameter guesses, c) 1 iteration, d) 26 iterations which corresponds with the terminal estimate. e) Development of the $\sigma_z$ estimates throughout the iterative optimization. f) True and g) estimated PSF value distribution. h) True and i) estimated noise distribution. k) Difference between the noise fields. See text for interpretation of this data.
2.4.3 Robustness in the presence of noise

We varied the SNR of the image and compared the parameter estimates with the corresponding true values that were known from data synthesis. This gave us the possibility to test the stability of the parameter estimation in presence of various noise levels. Figure 7 plots the behavior of 6 representative parameters: the center position $p_z$ of the Gaussian PSF body, and its lateral and axial 1/e-positions, $\sigma_y$ and $\sigma_z$ respectively; first ($b_1$) and second ($b_2$)order perturbation coefficients; and the relative bead scaling $\zeta$.

We have exposed the deconvolution to SNR values between zero noise (SNR = $\infty$) and SNR = 3. The latter represents an extremely low SNR mimicking the weak signals observed at the bottom of scattering and light absorbing thick sections. The known true parameter values are indicated by the dashed bold lines. The parameter estimates are plotted as thin solid lines. Overlaid are error bars propagated by the LS deconvolution scheme. Given the solution to Eq. 10 in the terminal iteration we can calculate 95% confidence intervals for any of the parameters in $\hat{\omega}$ by

$$t_{\nu}(95\%) = 1.96 \times 2\hat{\sigma}_n \sqrt{(A^T A)^{-1}}$$

with $\hat{\sigma}_n = (I - I_m(\hat{\omega}))^2 / N_{\text{pix}}$. $\hat{\sigma}_n$ describes the a posteriori variance of the image noise.

Apart from minor exceptions the parameters fall inside the confidence intervals around the estimated values. This is true even for high SNR suggesting high robustness of the parameter search method against noise. The parameters most susceptible turn out to be $b_1$ and $b_2$. For low SNR where the true value for $b_1$ falls outside the confidence range of the associated estimate we observe a clearly negative correlation between them. Overestimation of one of them is accompanied by underestimation of the other. This suggests that these parameters get weakly determinable as single values, yet as a parameter set they can still correct cooperatively for significant perturbation effects.
Chapter 2. One-Step Calibration of Focal Shifts

2.5 Discussion

In this paper we make a proposal for a simple, simultaneous calibration of focal shifts and spherical aberration in thick sections with refractive index mismatches. We have established a framework for object-constrained parametric blind deconvolution enabling the estimation of a depth-dependent PSF representation and the deformation of focal check beads, which is associated with focal shifts.

The results we show at this point of our study are very preliminary. We have confirmed the convergence and high robustness of the estimation algorithm in presence of low SNR. This is an essential first step, as light scattering and absorption cause typically very faint and noisy signals at the bottom of thick sections.

A central problem in our current prototype implementation is the relatively narrow convergence area for parameters describing the bead geometry - the actual target quantities in our calibration procedure. At this point the initial guesses for this parameter class needs to be within 5% of the final value. We can improve this situation in two ways: Application of more global parameter search strategies like Levenberg-Marquart. Improvement of gradient computation or replacement of gradient-based search strategies with non-gradient-based strategies. The weakest step in the current implementation is the calculation of the columns in the Jacobian $A$ that are associated with the bead parameters (cf. Eq. 12). This is mainly due to...
the limited resolution of the convolution procedure. In order to efficiently account for object symmetries in the large number of convolutions required for iterative LS optimization, we have restricted the bead surface representation to the discrete voxel raster. Subtle variations in the bead parameters as they are demanded for the computation of numerical derivatives cannot be handled correctly at this resolution level. Consequently significant artifacts are imposed to the calculation of the parameter gradients. We will examine refined gradient computation using more sophisticated interpolation schemes. Alternatively, we could replace the gradient-based search strategy by non-gradient-based optimization methods. Yet, we hesitate to go this route, as the computation time will enormously increase. Also, we would loose the possibility for error propagation and other statistical analysis of the parameter estimates, which are inherent features of gradient-based LS optimization. Another issue for further investigation concerns the PSF model itself. We plan to support the LS estimation of the PSF parameters with self-diagnostic tools, similar to the framework we have applied earlier to the calibration of stereo microscopes (Danuser(1999)). Insignificant and non-determinable parameters, which deteriorate the estimation stability will be automatically detected and excluded from the system. This will allow us to start the optimization with a relatively high order perturbation term in the mathematical PSF representation. Statistically significant, yet fine details in the PSF aberration will be maintained while physically irrelevant parameters introduced with the initial model will fall out.

The strength of our method is that once the PSF and focal shift parameters are established, noisy, thick section images of biological samples can be readily deconvolved using conventional techniques in each depth class separately. Our calibration algorithm is applicable to more complicated sections mimicking real biological tissues with a strong stratification of refractive indices. In certain cases where focal check beads can be implanted into the sample one can even think of an in situ calibration.

2.6 Acknowledgements

We acknowledge the financial and intellectual support from Carl Zeiss, Inc. Also, we are grateful to our friend Claude Berney who helped us in proof-reading the final manuscript.
Chapter 3

3. Automatic Fluorescent Tag Detection in 3D with Super-Resolution: Application to the Analysis of Chromosome Movement

Dominik Thomann¹, Daniel R. Rines², Peter K. Sorger² and Gaudenz Danuser¹∗

¹Bio Micro Metrics Group
Laboratory for Biomechanics
Swiss Federal Institute of Technology
Wagistrasse 4
CH – 8952 Schlieren, Switzerland

²Department of Biology
MIT Room 68-371
77 Mass Ave
Cambridge MA 02139


Relevance and own contribution

The following paper presents a novel algorithm for the detection and localization of fluorescent tags in 3D with super-resolution. It reports a first step towards the answers of the research questions of this thesis. The novel detection algorithm requires minimal user input and thus accelerates the process of tag localization for entire movies by several orders of magnitude as compared to the manually performed analysis of such data. But the gain in time is not the only advantage. Compared to cumbersome manual clicking of the center of the spots in 3D, the algorithm also increases the accuracy in position estimation immensely. In case of overlapping spots, a manual determination of the position is in many cases not possible at all. The presented algorithm extends the limits in resolution to a factor two beyond the classical optical diffraction limit for interfering point sources even in such low SNR data.

This paper is co-authored by Dan Rines and Prof. Dr. Peter Sorger of the Massachusetts Institute of Technology and my supervisor Prof. Dr. G. Danuser. Dan’s and Peter’s contribution were the setup of specialized 3D light microscopy with high time resolution, the culturing and labeling of yeast strains and filming of complete movies of viable cells. I visited the Sorger lab for two weeks to optimize the microscopy for our algorithms. My contributions to this paper consisted in:

- Simulated and experimental validation of a Gaussian approximation for the PSF in low SNR wide-field microscopy
- Development of a sensitive criteria for discriminating between spot features and noise related local maxima for the detection of spots in low SNR data
- Novel approach for the localization of interfering spots based on a PSF mixture model
Summary

In this paper, we describe an algorithmic framework for the automatic detection of diffraction-limited fluorescent spots in 3D optical images at a separation below the Rayleigh limit, that is, with super-resolution. We demonstrate the potential of super-resolution detection by tracking fluorescently tagged chromosomes during mitosis in budding yeast. Our biological objective is to identify and analyze the proteins responsible for the generation of tensile force during chromosome segregation. Dynamic measurements in living cells are made possible by GFP-tagging chromosomes and spindle pole bodies to generate cells carrying four fluorescent spots, and observe the motion of the spots over time using 3D-fluorescence microscopy. The central problem in spot detection arises with the partial or complete overlap of spots when tagged objects are separated by distances below the resolution of the optics. To detect multiple spots under these conditions, a set of candidate mixture models is built, and the best candidate is selected from the set based on $\chi^2$-statistics of the residuals in least-square fits of the models to the image data.

Even with images having an SNR as low as 5 to 10, we are able to increase the resolution two-fold below the Rayleigh limit. In images with an SNR of 5 to 10, the accuracy with which isolated tags can be localized is less than 5 nanometers. For two tags separated by less than the Rayleigh limit, the localization accuracy is found to be between 10-20nm, depending on the effective point-to-point distance. This indicates the intimate relationship between resolution and localization accuracy.
3.1 Introduction

One of the most powerful approaches to study cell biological processes has evolved in the form of dynamic analysis of sub-cellular structures such as organelles, vesicles, and the cytoskeleton by localizing and tracking labeled proteins using light microscopy. The development of techniques in which target proteins are fused to Green Fluorescent Protein (GFP) makes it straightforward to label specific macromolecular structures in vivo. Optical microscopes and digital imaging have advanced to the point where the recording of time-lapse series of 3D images at multiple wavelengths from living cells (frequently referred to as 5D imaging) has become routine in many labs. However, the methods for analyzing these images and for the extraction of accurate quantitative data of the localization and dynamics of biological structures are much less developed. This is particularly true in the case of structures that are small relative to the resolving power of optical microscopes and appear in images as diffraction-limited spots. Manual analysis is usually inappropriate in this situation.

Several methods have been reported for automated detection and localization of spots (Table 1). Most of these papers address the problem only in two dimensions and neglect the case of partially overlapping spots. This situation occurs whenever fluorescent markers are positioned at distances below the diffraction limit. The only paper listed in Table 1 that deals with overlapping spots is the one by (Netten, et al.(1997)). They apply a non-linear Laplace filter and thresholding to detect two overlapping spots, but since they are only interested in the number of present spots they do not localize the spots individually.

Most of the other methods listed in Table 1 rely on conventional linear image filtering techniques and some type of maximum search strategy, in which the position of a spot is associated with a peak in the filter response field. Cheezum, et al.(2001) and Thompson, et al.(2002) examined the fitting of a 2D-Gaussian to the sub-resolution feature in order to improve the localization. Although these approaches give good results in localization accuracy and precision they are usually limited to well-separated spots.
Table 1

Overview of the literature on spot detection and tracking. Most of the articles deal with 2D data only and with non-overlapping spots.

Legend: D=Detection, L=Localization, T=Tracking, SPM=Single Protein Molecule, SF=Subchromosomal foci

This limitation is a significant because in many biological systems it is necessary to detect and localize structures that are separated by less than the resolution limit of the optics. Santos & Young(2000) were, to our knowledge, the first who suggested using multiple Gaussians to improve the resolution. However, their study is limited to the 2D case and deals with synthetic data only. The only references in Table 1 addressing 3D spot detection are those by Bornfleth, et al.(1999) and Bornfleth, et al.(1998). In contrast to the approach presented in this article, they do not apply a PSF-model in the detection and localization procedure, but perform sophisticated image segmentation by a region-growing algorithm that includes prior knowledge from measured PSFs. This bears the advantage that they can handle spots representing images of sub-resolution tags as well as spots associated with larger structures at the same time. However, a region growing approach inherently fails in the segmentation of two or more spots that are located at less than 1-2 times the
Rayleigh distance apart, dependent on the SNR. The position estimation is accomplished by centroid calculation, which works fine for symmetric spots, but is significantly biased in case of asymmetry associated with asymmetric tag distribution in non-subresolution objects, aberrations or partial fusion of multiple spots.

For noise free images the distance of resolution for equally bright, fluorescent tags is given by the Rayleigh distances (Inoue(1995)):

\[
\begin{align*}
    d_{xy}^R &= 0.61 \frac{\lambda}{NA} \\
    d_z^R &= 2 \frac{\lambda n}{NA^2}
\end{align*}
\]

Equation 13

The two distances correspond to the first root of the point-spread function (PSF) in the lateral and axial direction, respectively. Also, note that the distances are inversely proportional to the cutoff frequency of the optical transfer function (OTF).

According to Equation 13, the resolution depends on the emission wavelength \( \lambda \), the numerical aperture \( NA \) of the objective lens and the refractive index \( n \) of the sample medium. For all calculations in this paper, we choose \( \lambda=525 \), \( NA=1.4 \), \( n=1.33 \) as these were the parameters of our microscope setup (see Section 3.2).

For many years microscopists have attempted to overcome the limits expressed in Equation 13 using one of two different but complementary sets of resolution-extending methods. The first category extends resolution by narrowing the main lobe of the PSF or by increasing the bandpass of the OTF. Representative PSF engineering methods include the confocal microscope, which uses a pinhole to narrow the PSF essentially in axial direction, or the more recently developed multiphoton microscopy and stimulated emission and depletion (STED) microscopy (Klar, et al.(2000)), which use non-linear optical effects in fluorescence excitation and emission to narrow the PSF. Other methods, such as 4Pi microscopy or fluorescence excitation with structured illumination can be better understood in the framework of OTF bandpass extension. For a comprehensive review of these optical approaches we refer to Gustafsson(1999).

The second category of methods does not change the optical characteristics of the microscope but attempts to recover spatial frequencies beyond the OTF cut-off using post-acquisition computational analysis. Since they do not increase but break the
optical band-pass of the microscope, we refer to such techniques as super-resolution methods.

The key to achieving super-resolution is the incorporation of prior knowledge, which constrains the search space in object reconstruction and permits the extrapolation of object-space frequencies beyond the cut-off of the OTF (Pask(1975)). Prior knowledge can be introduced at the level of the raw signal, as for example with the non-negativity constraint frequently applied to iterative deconvolution of fluorescence image stacks (Carrington, et al.(1995)), or on the higher, more symbolic level of visual information, where geometric and dynamic models of the observed specimen are included (Danuser(2001)).

In this paper we demonstrate the potential of super-resolution for a biological application: The tracking of chromosomes during mitosis in the budding yeast S. cerevisiae. Mitotic chromosome segregation is the process by which replicated sister chromatids are divided equally into two physically separated sets around which daughter cells form. Sister chromatids are created by DNA replication and bind to the microtubules of the mitotic spindle, a complex contractile engine comprised of several classes of microtubules and various motor proteins (Sharp, et al.(2000)). Chromatids bind to microtubules in pairs via kinetochores, multi-protein complexes that assemble on centromeric DNA. An essential step in mitosis is linking one set of sisters to microtubules emanating from one spindle pole and the other set of sisters to microtubules emanating from the opposite pole to achieve a state of bipolar attachment. The process of establishing and maintaining bipolar attachment involves the oscillatory movement of sister chromatid pairs back and forth along the spindle microtubules. When bipolar attachment is achieved during metaphase, sufficient tension is imposed on sister chromatids that the chromatin stretches and the sisters separate at their centromeres but not along their arms as shown in Figure 8 (He, et al.(2000), Tanaka, et al.(2000), Goshima & Yanagida(2000) in yeast and Sullivan(2001) in humans). The metaphase dynamics of sister centromeres in budding yeast involves transient separation by as much as 0.8 \( \mu \text{m} \) on a spindle ca. 2 \( \mu \text{m} \) long, whereas anaphase involves the dissolution of sister linkage and the sustained movement of chromatids towards the poles.
Our long-term goal is to characterize the movement of yeast chromosomes in sufficient detail in wild-type and mutant cells such that rates of acceleration can be measured, forces deduced and a mechanical model of the process developed. Ultimately this mechanical model will be combined with molecular genetics to determine the roles of various kinetochore proteins in force generation. Real-time analysis of chromosome movement in yeast is made possible by labeling individual chromosomes with GFP (see Section 3.2). The challenge is to track GFP tags with high precision and at different degrees of separation in 3D movies comprising 1000 or more 2D optical slices.

As a first step in our long-term program, we report the development of a fully automated tool for chromosome and spindle pole body (SBP) detection that incorporates super-resolution techniques. The prior knowledge necessary to formulate the constraints for super-resolution detection is the assumption that each spot visible in the 3D image represents the band-limited projection of a finite number of sub-resolution GFP tags. In other words, spots are comprised of a finite number of superimposed PSFs. We find that this weak constraint suffices to improve the resolution between a factor of two and three relative to the classical limit (Equation 13). In the future, we intend to use more sophisticated constraints based on a mechanical model of mitotic spindle dynamics. The better our mechanical model, the stronger the constraints and the greater the power of the super-resolution tracking. This illustrates the close and fundamental interaction between data analysis and biological modeling in the development of optimized analytic routines.

The rest of the paper is organized as follows: In the following section we describe the materials and methods for 3D live cell imaging. Section 3.3 starts with mathematical preliminaries concerning the PSF and explains the spot detection algorithm in detail. In Section 3.4, we first present results on biological data and then assess the theoretical limits and performance of the algorithm using synthetic data. In the last section, we summarize and discuss the main findings of our work and give an outlook on future extensions.
3.2 Materials and Methods

Our analysis of chromosome dynamics in yeast is based on the construction of strains in which chromosomes can be tracked relative to the spindle axis (He, et al.(2001), He, et al.(2000)). The combined tracking of chromosomes and the spindle poles is essential to distinguish between chromosome movements driven by kinetochores and motion induced by the rapid movement of the nucleus under the influence of astral microtubules. Additional motion can arise from drift in the microscope stage and whole-cell relocations in the buffer due to volume variation of the mounting gel. All of these kinetochore independent motions are eliminated by using a spindle-centered reference frame.

Chromosomes are tagged with GFP by integrating a tandem array of 100-250 tetracycline (TET) operator sequences (p306tetO2X224) at a unique site in the genome and then expressing a fusion between the DNA-binding TET repressor (TetR) and GFP. A typical array spans 11 kb of DNA, or about 3.5 \( \mu \text{m} \) of linear B-form DNA, but in mitotic yeast cells the DNA is compacted approximately 200-fold. When chromosomes are not under tension, chromosome tags are estimated to span about 25 nm and appear as diffraction limited spots as shown in Figure 8. When chromosomes stretch they can become deformed into linear structures as much as 750 nm long. In this paper, we make the assumption that chromosome tags are well represented by diffraction-limited spots, as indicated in Figure 8, for both relaxed chromosomes and chromosomes under tension. We consider the issue of tag deformation as an exceptional case and will propose a solution to this problem in the Discussion.

TET operator arrays were inserted approximately 2kb (representing 5-10nm of compacted chromatin) from the CEN-IV. Thus, the chromosome tags allow proper visualization of the location of the kinetochores. The location of the mitotic spindle is determined by marking the spindle pole bodies (SPBs) with a GFP-fusion to an integral pole protein, Spc42p. Spc42p is present at ca. 1000 copies per SPB and is a component of the crystalline central plaque. EM reveals that this central plaque has dimensions of ca. 0.2 \( \mu \text{m} \) and it is therefore accurately represented as a point source in fluorescence images (Bullitt, et al.(1997)).
Figure 8

**Top row:** Schematic of the mitotic spindle in budding yeast.

**Bottom row:** Two-dimensional projection (maximum intensity projection) of 3D image data illustrating the states above. The data of this process has been acquired using a Delta Vision optical section microscope, equipped with 100X 1.4NA objective and Roper RTE camera.

**a)** Microtubules connect to the chromosomes through kinetochores. Black dots represent GFP tags embedded in the chromatids near centromeric DNA (CEN) and in the two spindle pole bodies (SPB).

**b)** Once bipolar attachment is established, the spindle exerts opposing tensile forces that pull the centromeres apart reversibly. However, for most of the time points, the kinetochore tags cannot be separated visibly.

**c)** Eventually, enough force is developed within the spindle to cause the sister centromeres to be visibly resolved, although the signal has become weak since the number of fluorophores are now physically cut in half and due to some amount of photo-bleaching.
Live cell imaging was performed using a Deltavision optical sectioning microscope on a Nikon TE200 base with a 100X/1.4NA objective and Roper RTE camera (Applied Precision, Inc.; Issaquah, WA). 3-D image stacks consisting of 16-20 optical slices separated by 0.2 μm with exposures ranging from 0.10-0.05 seconds were recorded every 0.5 to 30 sec to generate 3D movies. These short exposures were necessary to capture the high dynamics of the spindle; to minimize photo-bleaching so that movies could be recorded over 30 to 40 minutes; and to avoid photodamage, thereby keeping the sample viable for long periods. Critical illumination was used, giving a spatial resolution of about 0.23 μm parallel to the image plane and 0.71 μm in Z (by Raleigh criteria). Fluorescent filters used for live cell imaging were EGFP Chroma 41017 and Endow GFP Bandpass Set.

All S. cerevisiae strains used in this study were haploid and derived from S288C (MATa; trp1-1; leu2-3,12; his3-11, 15; ura3-52, ADE+). Live cell mounts were created by applying cells, suspended in a small amount of medium, directly to a microscope slide. A cover glass was applied and sealed in place with petroleum as described in Rines, et al. (2002).

3.3 Algorithm for 3D Spot Detection with Super-Resolution

In the following, we describe an algorithm for 3D tag detection with super-resolution. This system can detect and localize multiple overlapping spot signals. The algorithm runs frame-by-frame, and the detection is therefore independent of the history of spot detection in previous frames. In future work, we will employ the detector described here only for the initialization of a time-propagating tracker that utilizes mathematical and mechanical models of movement to determine tag displacements between consecutive frames.
Figure 9
Outline of the algorithm

a) Noise reduction with a matched Gaussian filter, where the parameter $\sigma$ is selected to match the width of a single GFP spot. b) Spot candidate detection by a local maxima operator. The significance of a local maximum is determined by the spot classification response $s$ defined in Equation 16. It is a measure for the concurrence of high local mean intensity and a convex intensity distribution. c) Cumulative histogram of the spot classification response $s$ over all local maxima found in one frame of a typical set of biological image data. The large separation between true spots and local maxima induced by background noise indicates that the spot classification response is a sensitive measure for robust spot selection. d) Detection of multiple overlapping spots. A single local maximum detected with the histogram analysis in c) may arise from a superposition of several spots. To find the actual number of underlying spots we fit a mixture model consisting of multiple kernels to each local maximum. With an F-Test of the $\chi^2$ statistics of the residuals the number of contributing kernels is determined. e) In each frame, the detected spots are classified into spindle pole body (SPB) and kinetochore (Kin) spots. The SPB spots are simply selected by the longest point-to-point distance. f) Spot tracking between consecutive time-points is achieved by a minimum search of a correspondence score, where the score for one spot correspondence configuration consist of the sum of a weighted minimal distance criterion between corresponding spots. The computer generated results of steps e) and f) can be verified and manually adjusted, if necessary, using a specialized graphical user interface.
3.3.1 A practical 3D model for the point spread function

Because both the kinetochore and SPB tags (in their undeformed states) are sub-resolution features their images can be approximated by the PSF of the microscope. The theoretical PSF in the focal plane of an aberration free microscope, normalized to unit intensity, is given by a Bessel function of the first kind:

$$PSF(r) = \left( \frac{2J_1(ar)}{r} \right)^2$$

$$a = \frac{2\pi NA}{\lambda}$$

Equation 14

It has been demonstrated that the PSF is well approximated by a Gaussian distribution (Santos & Young(2000)) for the 2D case. To validate a Gaussian approximation in 3D we compared it to the PSF model described by Gibson & Lanni(1991). We first fitted the Gibson-model to the experimental intensity distribution of a single and isolated GFP spot. Then, we fitted a 3D Gaussian to the same data. Using the fitting residuals we calculated the $\chi^2$ statistics for both models, yielding $\chi_{\text{Gibson}}^2 = 6.01$ and $\chi_{\text{Gauss}}^2 = 7.0$ in units of image grayvalues. A comparison of the two residual statistics with a Fisher significance test suggests that, given the relatively low SNR of our image data, the Gibson & Lanni model improves the simple 3D Gaussian model for the PSF with a probability of only 30%. Figure 10 displays the raw data (a); the Gibson & Lanni model (b); and the Gaussian model in XZ-sections. The residuals for the Gibson & Lanni model are presented in (d) in central XZ-, YZ- and XY-sections, and in (e) for the Gaussian model. As the numerical values of the $\chi^2$ statistics suggest, the residuals of the two models are in the same range and they look approximately random, which means that they largely reflect fitting errors due to noise. A minor exception of this conclusion has to be made for the residuals of the Gaussian model in Figure 10e. They exhibit a small systematic residual near the center, indicating a weak systematic error of the model.
Chapter 3. Automatic Detection of Fluorescent Tags

Figure 10

Comparison of a complex PSF model (Gibson & Lanni(1991)) and the more simple Gaussian approximation fitted to the image of a sub-resolution GFP tag.

The top row shows the central XZ-plane of
a) the raw data; b) the fit of the Gibson model; c) the fit of the Gaussian model.

The bottom row shows the central planes of the fitting residual of
d) the Gibson model; e) the Gaussian model. Confer text for a detailed analysis of the residuals.

This can be explained by the fact that in our current fitting of the Gaussian we introduce only the center position and the amplitude as free parameters while keeping the parameter $\sigma$ fixed (see below). In the presence of small spherical aberration, which essentially affects the axial (Z) direction, this leads to a slight imperfection in representing the image of a sub-resolution feature with a 3D Gaussian. Nevertheless, as discussed below, all the statistical tests employed for the extraction of mixture models are applied on a confidence level between 95% and 99%. Therefore, the difference of 30% between the two tested models is practically irrelevant. We chose the 3D Gaussian as our PSF model despite its minor imperfection. This choice is mainly motivated by computational considerations. The forward calculation of the Gibson model with a given set of parameters takes several seconds on a 1.2GHz AMD computer. Unfortunately, the procedure cannot be
parallelized. As the fitting of the model is a non-linear and thus iterative optimization process, the computation time for localizing one sub-resolution tag increases prohibitively. Matters get even worse when fitting mixture models of several PSFs (see below).

An additional simplification of the Gaussian PSF model is the absence of a zero-crossing. We choose the parameter $\sigma_{xy}$ – the lateral radius for which the intensity of the approximated PSF drops to $1/e=0.37$ times the peak intensity – as the value that provides an optimal least squares fit of the Gaussian to an aberration-free PSF model relying on Bessel functions. In this case the Gaussian decreases to about 1% of the maximum intensity at the position of the first root of the Bessel function. The exact percentage depends on $\lambda$ and the $NA$. Notice that this low intensity value sufficiently approximates the exact root of the Bessel function, the latter also representing the Rayleigh limit. We write the parameter $\sigma_{xy}$ of the Gaussian as

$$PSF(r_0 = 0.61 \cdot \frac{\lambda}{NA}) = 0,$$

$$g(r_0) = \exp(-\frac{r_0^2}{2 \cdot \sigma_{xy}^2}) = 0.0163$$

$$\sigma_{xy}(\lambda, NA) \approx 0.21 \cdot \frac{\lambda}{NA}$$

Equation 15

This parameter is then held fixed during the fitting of Gaussian PSFs to spot images. Using Equation 13, a similar calculation can be made for the axial direction:

$$\sigma_z(\lambda, NA, n) \approx 0.66 \cdot \frac{\lambda n}{NA^2}.$$

Consequently, for a given set of physical parameters ($\lambda$, $NA$, $n$) the ratio between $\sigma_{xy}$ and $\sigma_z$ is fixed. By mapping the continuous object space into the pixelated image space an additional ratio is introduced accounting for unequal sampling in lateral and axial directions (representing the typically fine spacing of pixels in the CCD camera and the coarser spacing of Z-steps between optical slices). This second ratio is determined by the lateral to axial side length of an image voxel ($P_{xy}/P_z$).
For the remainder of this paper, we define the parameter $\sigma := \sigma_{XY}$ in [$\mu$m] and relate to it:

$$\sigma = 0.3 \frac{NA}{n} \cdot \sigma_z.$$ 

The relationships between object space (in $\mu$m) and image space (in voxels) that follow from these equations are summarized in Table 2.

| Object space [\mu m] | \sigma_{XY} = 0.21 \cdot \frac{\lambda}{NA} | \sigma_z = 0.66 \frac{\lambda n}{NA^2} |
|-----------------------|----------------------------------------------|
| \sigma := \sigma_{XY} | \sigma_z \approx 3.1 \frac{n}{NA} \cdot \sigma |
| \sigma_{XY} \approx \frac{\lambda}{NA} | \frac{d_{XY}^2}{\sigma_{XY}} \approx 2.9 |
| \sigma_z \approx 0.66 \frac{\lambda n}{NA^2} | \frac{d_z^2}{\sigma_z} \approx 3 |

Table 2

Relationship between the image space and the object space and between the parameter $\sigma_{XY}$ used for PSF approximation. A Greek letter denotes that the value is expressed in the image space, i.e. in pixels as units. $P_{XY}, P_z$ are the lateral and the axial pixel size respectively.

### 3.3.2 Noise reduction by matched filtering

The requirement of short exposure times for each optical slice so that the sample is not subjected to excessive photo-damage results in data with typically an SNR $\approx 15$ or less (see Materials and Methods). Consequently, the reliable detection of fluorescent spots is a challenging task. The algorithm has to distinguish between true spots and intensity maxima arising from background noise. To suppress early on as many noise-related local maxima as possible, we filter the raw intensity data $I_r$ with a 3D Gaussian kernel $G(\xi, \sigma)$ with the parameter vector $\sigma=(\sigma_{\xi n}, \sigma_{\eta n}, \sigma_{\zeta})$ defined in the image coordinates $\xi=(\xi, \eta, \zeta)$ (see Figure 9a). The parameters are matched to the optics according to Table 2 and Equation 15 with $\int G(\xi, \sigma) d\xi = 1$. 
The convolution of the raw intensity data with this kernel results in an intensity volume $l_f(\xi)$ that exhibits high responses at the position of potential spots, but contains, compared to the raw image, fewer noise-induced local maxima. Notice that the Gaussian filter kernel has the characteristics of amplifying low frequency features at the expense of high frequency information and hence of spatial resolution.

### 3.3.3 Spot detection

A 3D local maxima detection is performed on the noise-filtered image $l_f$ to generate a set $L$ of features representing the positions of potential spots:

$$L = \left\{ \xi \in I_f(\xi) \left| I_f(\xi) > I_f(\xi + t), \forall t, \in \{-1, 0, 1\}, t \neq 0 \right. \right\}.$$  

Because the SNR of the images is low, noise is often recognized as a feature. Thus, the set $L$ typically contains many more candidate features than true spots present in the image. A partitioning of $L$ into true spots and the many noise-induced local maxima applying a simple threshold is not appropriate because of low SNR. Instead, structural information has to be employed, where the distinction is made based on differential geometric considerations in $\mathbb{R}^3$. For every local maximum $l_i \in L$ we calculate the mean intensity $\bar{l}_i$ of a $5\sigma \times 5\sigma \times 5\sigma$ volume surrounding the local maximum position $\xi$ and the curvature $\kappa_i$ of the intensity distribution in this point (Figure 9b).

Note that $\kappa_i$ is not the curvature of a 2D surface but of a 3D intensity distribution. The curvature is given by the determinant of the Hessian matrix $H$ of the intensity $I_f(\xi)$.

$$H(\xi) = \nabla \cdot \nabla^T I_f(\xi)$$

$$\kappa_i(\xi) = \det(H(\xi))$$

Based on these two numbers, we define a spot classification response

$$s_i = \bar{l}_i \cdot \kappa_i.$$ 

**Equation 16**

This value turns out to be very sensitive in discriminating between features associated with true spots and artifactual features associated with background noise. After matched filtering noise-induced local maxima typically exhibit a random distribution of intensity changes in all directions leading to a low local curvature $\kappa$. In contrast, true spots are characterized by a combination of convex intensity distribution in all directions plus a relatively high mean intensity. Thus, while the mean intensity alone is not sufficient to separate noise-induced local maxima from
true spots, a combined selection based on the concurrence of high intensity and correct curvature is very effective.

We run a statistical analysis of the data $s_i \forall i \in I$. We generate a cumulative histogram and select a threshold, $s_t$, at the position where the slope of the histogram significantly flattens. Figure 9c displays an example of such a histogram. In the case displayed, all classification responses except those associated with the three significant spots cluster around $s=0$, resulting in an almost infinitely steep slope. It is straightforward in such data to automatically define $s_t$ and partition $I$ accordingly into a set of noise-induced local maxima, $I_N$, and a set of tag-induced local maxima, $I_T$.

3.3.4 Separation of overlapping spots

When two tags are separated by a distance smaller than the optical resolution limit, the superposition of their images gives rise to a single local maximum only (Figure 9d). As explained in section 3.3.2, suppressing noise by filtering decreases the resolution limit further. We address the problem of unresolved tags using a module for multiple spot detection. The idea is to build a candidate set of mixture models, where the best candidate model is selected based on a goodness-of-fit criterion. The mixture model consists of a superposition of $n$ kernel functions and is formulated, in the case of diffraction-limited spots, as a superposition of $n$ Gaussians, each one representing a version of the PSF shifted in space and with variable intensity

$$M(\xi, a, b, c, n) = \sum_{i=1}^{n} a_i \cdot \exp \left( -\frac{1}{2} (\xi - c_i)^T \Sigma^{-1} (\xi - c_i) \right) + b$$

$$\Sigma = diag(\sigma_{\xi y}, \sigma_{\xi y}, \sigma_{c})$$

Equation 17

The free parameters of the model are the number $n$ of Gaussians, the center positions $c_i$ of each Gaussian, the amplitudes $a_i$ of the Gaussians and a common background value $b$. A number of techniques have been reported for fitting such mixture models to real data and for simultaneously determining the number $n$ of underlying kernels. The primary challenge is to trade the goodness of fit against the degrees of freedom of the model: the larger $n$, the better is the fit to real data, but the higher is the probability of introducing insignificant kernels. In our analysis of tagged chromosomes and SPBs, insignificant kernels are prohibitive as they define the
parameters of a non-existing tag. A recent approach to this problem is motivated by mutual information theory and starts with a large number \( n \) of kernels (Yang & Zwolinski(2001)). Each kernel is tested for its statistical dependency on the other kernels. Kernels that are statistically independent are judged to make a significant contribution to the full model and are retained. In contrast, kernels with strong mutual dependence are eliminated from the mixture model. Our approach to finding an optimum mixture model is quite different from that of Yang & Zwolinski(2001) in that it involves a bottom-up scheme. We start with a low number of already validated kernels and ask whether additional kernels are significant. Since our images contain four or fewer objects a small number of underlying kernels suffices for the formulation of a mixture model. Moreover, a bottom-up approach minimizes the inclusion of insignificant kernels. We fit a candidate model of the order \( n \) to the raw data and calculate the residuals \( R_n \) and the corresponding \( \chi^2 \) statistics:

\[
R_n(\xi) = M(\xi) - I(\xi)
\]

\[
\chi^2_n = \frac{\sum R_n(\xi)^2}{r_n}
\]

**Equation 18**

The degree of freedom \( r_n \) is defined as \( r_n = N - p_n \), where \( N \) is the number of voxels supporting the fit and \( p_n = 4n + 1 \) denotes the number of parameters. We formulate a statistical test

\[
H_0 : \chi^2_n = \chi^2_{n+1}
\]

\[
H_A : \chi^2_n > \chi^2_{n+1}
\]

\[
T = \frac{\chi^2_n}{\chi^2_{n+1}} \sim F
\]

**Equation 19**

to decide whether it is necessary to represent the data with \( n+1 \) kernels or if \( n \) kernels suffice (Bevington & Robinson(1992)). The test statistics \( T \) is Fisher distributed and we typically test on a confidence level \( \alpha = 95\% \). In practice, we perform these tests as follows: from \( L_T \) we build clusters of local maxima where each member in the cluster has a minimal distance of \( 9\sigma \) to all non-member local maxima in \( L_T \). The distance of \( 9\sigma \) guarantees that intensities of any non-member in \( L_T \) do not influence the mixed model fit of the cluster.
The number $m \geq 1$ of cluster members defines a starting value for $n$. Notice that the fitting problem is highly non-linear. The initial set of parameters, the center coordinates of the kernels and the amplitude are derived from the position and intensities of the associated local maxima. For the fitting we use the MATLAB non-linear least square module \textit{lsqnonlin}. This module implements a large-scale algorithm, which is based on the interior-reflective Newton method (Coleman & Li(1996), Coleman & Li(1994)). After fitting an $n$-order model, the procedure is repeated on an $n+1$ order model. In a random sequence, we assign to every one of the $m$ original cluster members a candidate partner spot that may have been lost in the previous noise filtering and local maximum search steps. For initial coordinates, we simply use the same values for the original and the partner spot and set the initial amplitudes equal. Applying the test procedure (Equation 19), we determine if the number of cluster members has to be incremented by one new kernel. If so, this additional kernel is included in all the partner search tests for the remaining spots in the cluster. The procedure is repeated for all clusters found in $L_T$, mapping the set $L_T$ of local maxima into a new set $S_T$ of detectable tags. The size of $S_T$ is equal or larger than the size of $L_T$.

The method described above stands in contrast to the mutual information theory approach, in which a few significant kernels have to be extracted from a large number of insignificant kernels. Without formal proof, we suggest that for data with a small number of overlapping spots and low SNR, the bottom-up approach is more stable. In particular, our strategy allows us to initialize $n+1$ order fits with almost perfect initial guesses from an $n$ order fit, and only one weak starting parameter set for the new kernel needs to be introduced. In top-down approaches, initializations for many unknown and potentially non-significant kernels have to be made. Depending on the quality of these parameter initializations, the mixed model fits can drift badly and the inclusion of non-significant kernels can damage a model containing highly significant kernels. This problem of estimation drift gets more prominent when the SNR of the data is low.

3.3.5 Classification of spots

We classify the features in $S_T$ and associate them with particular objects in the data: GFP-tagged kinetochores and SPBs (Figure 9e). While the algorithmic steps described above are generally applicable to any kind of spot detection, the
procedures for classifying spots are specific to the biological system and involve the application of prior knowledge. Because the trajectories of kinetochores and SPBs are highly characteristic, it is our impression that human-assisted spot finding makes considerable use of time-dependent information. However, our current spot finding algorithm works frame by frame and it is not yet possible to incorporate time-dependence in the analysis. We have, therefore, relied on a geometric argument to classify SPBs and kinetochores. In general, chromosomes lie inside the spindle poles and the longest point-to-point distance among four GFP spots is the spindle axis. Shorter distances are the SPB-kinetochore and kinetochore-kinetochore distances. This simple approximation appears to be quite reliable in wild-type cells, but is not as reliable with mutant cells in which the spindle is often shorter due to an imbalance in mitotic forces. In this case, the system requires operator support in classifying spots. In the future, we intend to develop more sophisticated approaches to automated spot classification.

3.3.6 Motion analysis

To track spots from one frame to the next, we have implemented an algorithm which selects a weighted minimal distance of all spots in two consecutive time points. A list of all possible correspondences of spots in time point t to spots in time point t+1 is generated. The sum of weighted squared distances is computed for each configuration of correspondences and the configuration with the minimal score is selected. For an unequal number of spots in subsequent time points the software distinguishes between two cases: With three and four spots present the correspondence mapper assumes the fusion or separation of kinetochores and the scores are calculated including all possible fusions or separations of spots. In all other cases only the spindle pole tags are mapped. The weight matrix favors spot correspondences within a typical range of displacement distances and is introduced to decrease the effect of outliers or falsely detected spots. The score weight for a spot i mapped from time-point t into a spot j in time-point t+1 is calculated according to the following equation:

$$ w_{ij} = e^{-\frac{(x_i - x_j)^2}{\sigma^2} - \frac{(y_i - y_j)^2}{\sigma^2}} $$

Equation 20
The parameter $\sigma_\theta$ is an estimate of the expected variation in spot intensity, e.g. due to bleaching, the parameter $\sigma_\rho$ the expected range of displacements. In future work, we will replace this procedure with a proper differential tracking of spot configurations between consecutive frames. This will allow us to directly extract trajectories and is expected to further increase the resolution (see Discussion). Since the spindle can move and rotate during the acquisition, we transform the coordinates from a microscope-centered Cartesian frame into a spindle-centered Cartesian frame in which the chromosome movements relative to the SPBs are easier to analyze. Because this transformation relies on a correct assignment of SPB and kinetochore tags, we have to eliminate possible errors in the above described classification scheme. Currently, we use a specialized graphical user interface where the classification and the mapping of the spots can be verified and interactively adjusted, if necessary.

3.4 Results

3.4.1 Performance on biological data

We first present results from the analysis of biological data and then focus on the performance of the algorithm using synthetic images. The algorithm is designed to work robustly on large data sets with minimal demand for parameter adjustment between movies. We did not tweak the parameters for the performance tests, but instead ran the software on synthetic data under the same conditions as on real imagery.

Initially, the software reads a definition file with optical parameters and the pixel sizes. From this, the parameter $\sigma$ is calculated and the PSF is formulated according to Equation 15. The corresponding parameters $(\sigma_{\rho}, \sigma_{\theta})$ in the image space are derived from the relations listed in Table 2. In addition to the PSF definition, the operator has to set two parameters manually:

1. The threshold $t_s = \frac{1}{s_{i+1} - s_i}$ for distinction between local maxima associated with true and artifactual spots (see section 3.3.3). The quantity $s_i$ represents the spot classification response (Equation 16) of the first spot in the cumulative histogram considered significant and $s_{i+1}$ denotes the response of
the next higher ranked spot. This means that the algorithm determines the index \( t \) of the spot in the cumulative histogram for which the slope falls below the operator-defined threshold \( t_s \). As mentioned before, the parameter \( s \) represents a very sensitive characterization of local maxima in terms of their probability of being a true spot. In the histogram true spots do not cluster with the large majority of false spots, resulting in a characteristic decrease of the histogram slope by a factor 100 or more (Figure 9c). Therefore, the software is very tolerant to non-optimal settings of this slope threshold.

2. The confidence parameter \( \alpha \) used for the statistical testing of mixture models (see section 3.3.4).

The detection of isolated spots poses no problem for the algorithm even in the case of low SNR. This is illustrated in Figure 11a. In this frame the cell is still in metaphase, the sister chromatids are closely linked and hence the two kinetochore GFP tags are imaged into a single spot. Figure 11b shows a frame where the sister chromatids are separated. Here the power of a super-resolution scheme becomes evident. Thanks to the fitting of mixture models, the software detects two overlapping spots with high positional accuracy.

Figure 11c displays the detection result in a time point where the chromatids are pulled towards the SPBs. In this state the four fluorescent spots are all isolated. Typically, the image sequences we have analyzed last more than 30 minutes. Acquiring image stacks every 30 seconds over this extended time period is accompanied by photo-bleaching, leading to a significant decrease of SNR. Thus, towards the end of the movies the detection of spots becomes problematic even though the spots may be well separated. The influence of bleaching on SNR and spot detectability is illustrated in Figure 12. The graphs show the number of spots detected (black line) in comparison to the mean SNR of all spots (blue line) and the SNR of the weakest spot (red line) in each frame. Figure 12a presents the result from a movie with 100 frames over 50 minutes of data acquisition. The graph in Figure 12b displays the same analysis for a movie with only 60 frames, i.e. 30 minutes of data acquisition.
Chapter 3. Automatic Detection of Fluorescent Tags

Figure 11
Results of the spot detection algorithm during yeast mitosis. a) Cell in metaphase: The two SPB spots are isolated and give a strong signal. The sister chromatids are attached to each other giving rise to an image of the two kinetochore tags as in a single spot. b) Super-resolution detection of two kinetochore spots. Thanks to the fitting of mixture models, the software detects two kinetochore tags with high positional accuracy although the associated spots overlap. c) The chromatids are separated and pulled towards the SPB. In this state, the four fluorescent tags are all separated by more than the Rayleigh distance and well visible in the image as isolated spots.

Both movies start with four detected spots and a mean SNR of 10. In the example in Figure 12a bleaching is very prominent. After 40 frames the mean SNR drops to 5 and the weakest spots are detected at an SNR level of about 3. Once this low level is reached, the algorithm breaks down and only 1 or 2 spots are detected for which the SNR is above 4. Presumably, they originate from the SPB tags, which contain
more fluorophores than the kinetochore tags. In the example in Figure 12b the SNR is much more stable. Over the period of 60 frames, the mean value drops from 10 to 8 allowing the algorithm to detect either 3 or 4 spots for most of the frames. The SNR of the weakest spot never falls below 4. Notice that in two frames, five spots are detected. The inset in the graph displays the maximum projection of the second of these frames. With regular contrast settings, the kinetochore tags are barely visible by eye. Yet, the software extracts one kinetochore tag position near SPB1 and two kinetochore positions closer to SPB2. We know that this is biologically not possible (see Figure 12b). There are two reasons for such outliers: i) The algorithm may split one spot into two spots depending on the confidence level of the statistical test (Equation 19). For biological data, we set this level to about 90%, with an error of the first kind of 10%. Notice that for the used statistical test, the error of the second kind, i.e. the error that artifactually split spots are classified as significant, is the same as the error of the first kind. Thus, we can decrease the frequency with which outliers are generated by increasing the confidence level, but this can be accomplished only at the risk of loosing significant splits that are detected under the present settings. ii) Chromosome tags are known to get stretched mechanically during metaphase He, et al. (2000) Goshima & Yanagida (2000), Sullivan (2001). This leads to objects whose images do not match the PSF kernels in the mixture model. Obviously, for stretched tags the mixture model tends to prefer two kernels instead of one. This problem will be addressed in a future version of the software, where spots will be tracked differentially with mapping models that not only account for tag translocation and brightness variation between consecutive frames, but also for mechanical deformation.
The influence of bleaching on SNR and spot detectability is shown for two movies. The graphs show the number of detected spots (black line) in comparison to the average SNR of all spots (blue line) and the SNR of the weakest spot (red line). 

(a) The long light exposure in this 100 frame movie leads to strong bleaching. Below an SNR of about 4 the spot detection algorithm breaks down. The one or two spots remaining after frame 50 originate from SPB-tags, which usually are stronger than the kinetochore tags, even when the latter are superimposed.

(b) This movie is more stable in terms of SNR. The average SNR remains around 8 and the SNR of the weakest spot never falls below 4, allowing the software to detect 3 or 4 spots in almost all frames. Two outlier frames are observed (arrows), where 5 spots are detected. The inset displays the maximum projection of the second of these frames. A detailed interpretation of this effect is given in the text.

Table 3 summarizes the results of three representative movies we have analyzed in detail. The left column specifies the number of frames contained by the movie. These values are divided in column 2 into percentages of frames with 0, 1, 2 or more spots detected. Only the third movie contains 2 frames with 5 spots and thus obvious outlier data (see Figure 12b). From this, we conclude that our parameter settings are sufficiently conservative that only few outliers are generated. When all movies are considered, the majority of the frames contain either three or four spots.
### Table 3

Statistical analyses of three biological data sets. The left column specifies the number of frames contained by the movie, which are divided in column 2 into percentages of frames with 0, 1, 2 etc. spots detected. The 3rd column indicates the percentage of spots detected by the mixture model fit. Columns 4 and 5 contain the mean over all frames and the minimum of the shortest tag-to-tag distance per frame. The statistics is done separately for isolated spots not requiring a mixture model fit ($d_{nm}$) and between spots extracted in mixture models ($d_{mm}$). The numbers indicate the improvement in resolution due to mixture model fitting.
The large number of frames in which we find only three tagged objects rather than the four present in cells suggests that despite mixture model fitting many of the kinetochore tags still cannot be resolved. Nevertheless, the benefit of the mixture model fits becomes evident from the results in columns 3 to 5. Column 3 specifies the percentage of spots extracted with mixture model fits. For most of the movies about 25% of the detectable spots can only be found with a mixture model. And even in the second movie, where most of the frames contain only 3 spots, 15% of all spots are only detectable because of the mixture model approach. Columns 4 and 5 indicate the improvement due to mixture model fitting in geometric terms. The mean over all frames of the minimal distance between isolated points not requiring a mixture model fit amounts to 0.4 – 0.7 μm, the shortest distance found in a whole movie to 0.3 – 0.4 μm. Mixture model fitting systematically decreases these numbers. The mean minimal distance between mixture model spots lies between 0.3 – 0.45 μm. Notice that while these values are still larger than the lateral Rayleigh distance, they are below the axial limit of resolution. We did not split the statistics into axial and vertical components. The essential result is given by the absolutely minimal distance found per movie. For the first movie this is 250 nm, right at the lateral Rayleigh limit of our optics, while for the two other movies in the table the value drops to 130 nm, a factor 2 below the Rayleigh limit. Notice that this estimate of a super-resolution factor represents a lower bound, as some of the measured distance vectors have a significant axial component.

From the estimated tag positions in each frame, kinetochore and SPB trajectories can be reconstructed (see section 3.3.6). In Figure 13a spot trajectories of a small part from a typical movie are displayed. The kinetochore trajectories split and rejoin several times during this period (cf. arrows for an example). For one of the time-points where kinetochores split, Figure 13b shows the spot configuration and the image stack projection to emphasize the need of super-resolution to recover such data. Notice that the rejoining of the kinetochore spots does not necessarily imply a physical rejoining of the tags. It only means that they are separated by a distance below the limit of detectability of our algorithm.
Figure 13

Kinetochores (red, green) and SPB (blue) trajectories for a small part from a 60 frames movie. a) The trajectories are displayed in the microscope reference frame, where the movement is a superposition of i) drifts of the microscope stage; ii) whole-cell relocations in the buffer due to volume variation of the mounting gel; iii) spindle movements relative to the cell cortex and iv) kinetochore movements relative to the SPBs. The arrows indicate positions where the kinetochores separate before rejoining later again. Figure b) shows a typical example where the two kinetochore spots overlap (displayed in the maximum intensity projection). The algorithm can still estimate the positions of both tags. c) Trajectories of a different movie are shown in a reference system with a fixed SPB-SPB axis. The kinetochores remain in a small central region between the SPBs suggesting that the cell is still in metaphase. d) The SPB-SPB distance as a function of time. The SPBs move apart with a constant rate.
The measured spot displacements are due to four factors: i) drifts of the microscope stage; ii) whole-cell relocations in the buffer due to volume variation of the mounting gel; iii) spindle movements relative to the cell cortex driven by microtubules and cortical motors; iv) kinetochore movements relative to the SPBs. Only type (iv) movements are driven by kinetochores. To visualize the kinetochore-dependent motion, we transform the trajectories from the microscope-centered Cartesian frame into a reference frame with one fixed SPB and the SPB-SPB axis defining the x'-axis. For any time point, the y' axis spans the plane defined by the SPB-SPB axis in the very first frame and momentary SPB-SPB axis. The z' axis is then chosen to form a right-handed coordinate frame. Figure 13c displays the trajectories of a transformed movie. For this movie, the two kinetochores perform a joined dance, relatively stable at half the distance between the two SPBs. Also, vertically, they always stay above the axis, an arrangement that appears to be typical of metaphase. Throughout the movie the SPBs move apart at a constant rate, another characteristic of metaphase (Figure 13d). The SPB-SPB distance estimates contain one outlier, which has been eliminated from the trend analysis of spindle growth. As shown in the next section, the accuracy with which well-isolated spots can be localized reaches 10 nm in movies such as this one with an SNR 8. The total change in the SPB-SPB distance over the course of the movie is in the range of 100 nm, and is therefore statistically significant and clearly visible in the data.

3.4.2 Performance on synthetic data

Before initiating large-scale biological studies in which protein function is linked to chromosome trajectories in wild-type and mutant cells, we investigated the performance of our algorithm using synthetic data. Synthetic spots mimicking the image of sub-resolution GFP tags were created using the Gibson-PSF-model (Gibson & Lanni(1991)). The parameters of the model were determined with a fit of the model to experimental data (see Section 3.3.1). The spots were positioned in a virtual 3D image space filled with a constant background value and additive white noise was superimposed to obtain test data at a predefined SNR. To study the resolution performance and the interplay between resolution and localization accuracy we prepared model trajectories of two tags converging towards one
another. For every SNR setting and point-to-point distance, 100 Monte Carlo simulations were run to accumulate statistical data on the detectability of spots and the accuracy with which they could be localized.

![Graphs showing the performance limits of the algorithm](image)

Figure 14

Performance limits of the algorithm analyzed with synthetic data. a) Resolution as a function of SNR and the tag-to-tag distance, calculated for equally bright tags and tags with a brightness ratio 2:1. The distance is indicated as a factor of the parameter $\sigma$ that characterizes the Gaussian PSF approximation. The curves display the boundary lines between the resolvable and unresolvable domains. Above the curve, spots are detected with a confidence of approx. 95%. The gray area gives an estimate for the range of confidence probability of 95-100%. b) Resolution in Z-direction for equally bright tags.

The localization accuracy depends on two factors: i) The SNR. ii) The tag-to-tag distance. The accuracy is analyzed by calculating an error statistics (Equation 21) of 100 random experiments for each parameter setting. c) The localization error of a single spot as a function of the SNR. d) The localization error as a function of the point-to-point distance calculated for three different SNR values.
**Resolution of tag separation**

Figure 14a displays the resolution performance of the algorithm as a function of SNR and distance. Here, we define resolution as a statistical measure: Given a certain SNR, the resolution is determined as the tag-to-tag distance, \( d \), at which the tag pairs can be separated with a predefined confidence. For our analysis, we chose a confidence level of 95%. For example, the graph in Figure 14a indicates that an SNR of 20 permits our algorithm to separate tags of a distance \( d = 1.9 \sigma \) with 95% confidence. Consequently, the area above and to the right of the distance-SNR curve represents the 'resolvable' domain, while the area below and to the left of the curve represents the 'unresolvable' domain. The boundary line between these domains is drawn for two cases:

i) Equal brightness of tags, \( l_1 = l_2 \) and

ii) A brightness ratio 1:2, hence \( l_1 = 2 l_2 \).

As expected, the algorithm performs somewhat less well in the second case. There exists a limit below which the mixture model breaks down independent of the SNR for both cases. We did not explore this limit further and suspended our simulations for SNR values larger than 40, because values beyond this are irrelevant for biological fluorescence microscopy.

Figure 14b illustrates the result of the simulation for the Z-direction. Comparing the curves for XY and Z direction we find similar resolution factors (i.e. the ratio between the resolvable distance and the Gaussian parameter \( \sigma \) for the respective direction), this although the Gibson model includes aberration terms which mainly affect the axial (Z) direction and which are neglected by our Gaussian PSF representation. This confirms again that in practice the Gaussian PSF is sufficient to represent high-resolution and low SNR images of sub-resolution tags. Notice that the apparent difference between the simulations in a) and b) are not significant. The locus of the boundary line at a certain distance is found by adaptive adjustment of the SNR until 95 to 100% of the spots are correctly identified in 100 Monte Carlo runs. Therefore, the curve does not represent an absolute boundary but the mean position of algorithm breakdown within a certain tolerance band. The gray area underlain in Figure 14a) and b) indicates the width of this 5% variation band.

For both cases \( l_1 = l_2 \) and \( l_1 = 2 l_2 \) we found that the algorithm can resolve tags at sub-Rayleigh distances (domain on the left side of the dashed line in Figure 14a). Recall
that the inverse of the Rayleigh distance corresponds to approximately the cut-off frequency of the OTF. Thus, the ability to resolve sub-Rayleigh distances requires the recovery of frequencies beyond the OTF passband, indicating the super-resolution performance of the algorithm. Super-resolution is achievable for SNRs as low as 5. Interestingly, for SNR values below 5 the algorithm fails to detect the spots independent of distance. This lower limit for SNR is defined by the confidence limit controlling the switch from a single kernel to a two-kernel mixture model. Below a certain SNR, the test statistics in Equation 19 will never reach the quantile associated with the predefined confidence probability. Neither a single- nor a two-kernel model will accommodate the noisy signal and thus the improvement in the goodness of fit for the two-kernel model is marginal. Potentially, we could push this breakdown to lower SNR values, but only at the risk of accepting more false spots. For real data, we usually choose 90% instead of 95% confidence. The breakdown observed with real data was therefore a little lower than with the synthetic data, corresponding to an SNR of 4.

For super-resolution factors larger than 2, SNR values larger than 30 are required. For the image data presented here, this is rarely achieved, yet the graph indicates the potential power of the algorithm in applications with brighter tags.

**Accuracy in point localization**

Not only the separation of tags but also the accuracy with which positions can be estimated is of practical interest. Positional accuracy is influenced by two factors, i) the SNR and ii) the point-to-point distance.

To address the influence of the SNR, we ran simulations on isolated spots. We define the statistical localization error as

$$err = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |\hat{x}_i - x_0|^2},$$

Equation 21

where $\hat{x}$ is the estimated position of the tag, which is the center position of the Gaussian PSF approximation, and $x_0$ represents the true tag position known from the simulation settings.

Figure 14c displays the localization error as a function of the SNR. Notice that the horizontal axis points from high to low SNR. The graph confirms the well known fact
that center positions of features with a known intensity distribution, in our case the
distribution of the PSF, can be determined with sub-20nm precision (Bobroff(1986),
Inoue(1989)). For SNR>15 the precision even reaches the single nanometer range.
Interestingly, this high precision can only be maintained for tags separated by at
least the Rayleigh distance. With shorter separation distances the accuracy
decreases, as suggested by the graph in Figure 14d (the Rayleigh limit is denoted by
a dashed vertical line). An explanation for this may be found in the weaker
determinability of the extent to which the two signals contribute to the super-position.
Notice that in a 1D mixture model for distances larger than the Rayleigh distance, the
rising edge of the mixture signal corresponds almost exclusively to the rising edge of
the first kernel, while the falling edge coincides with the falling edge of the second
kernel. Therefore, half of the intensity measurements in support of each kernel are
uniquely associated with this kernel. Below the Rayleigh distance this is no longer
true. The fitting algorithm needs to partition every intensity measurement into a
contribution of the first and the second kernel. In the presence of noise, this is not
unambiguous, leading to decreased accuracy in fitting.

To conclude, the performance tests on simulated data clearly demonstrate the
super-resolution potential of mixture model fitting even in the case of low SNR.
However, the results also show that high precision single-kernel fits cannot be
maintained in the sub-Rayleigh domain. Although the spots are still separable by the
algorithm, the localization accuracy drops remarkably. However, in no relevant case
is there a crossover between resolution and accuracy: the minimal distance
resolvable is always larger than the associated measurement error. In other words,
once two tags are detected as separate, the uncertainty in the distance estimate is
small enough to render the difference in the tag coordinates statistically significant.
This represents a condition imperative for any further statistical analysis of tag
positions.

3.5 Discussion

In this paper, we report a generic algorithm for automated fluorescent tag detection
with super-resolution. Both the possibility of attaining resolution beyond the
diffraction limited passband of a microscope as well as methods for efficient and
unsupervised spot detection by computer vision have been discussed. The novelty of
our contribution is in the rigorous extension of such methods to 3D, where manual measurements are especially cumbersome. Furthermore, we demonstrate the performance not only on synthetic data but also on a relevant set of biological data in which distance measurements below the conventional resolution limit are required.

**Biological Applications**

In the immediate future, the primary biological application of the methods described in this paper is automated motion analysis of GFP-tagged chromatids during yeast mitosis. We plan to exploit our algorithm to perform systematic cell-based analysis of proteins involved in force generation by kinetochore-microtubule attachments. For every genetically modified yeast strain tens of movies, representing several GByte of image data, will be collected and processed automatically to generate information on chromosome trajectories. We aim then to extract principal motion components from the trajectories and compare trajectories among wild-type and 50 to 100 mutant yeast strains. To achieve the high-throughput necessary for this analysis, a large degree of automation is imperative, and the chromosome tracking software must perform reliably without the supervision of an operator. Furthermore, for robust principal component analysis by statistical means the generation of outliers must be reduced to an absolute minimum. Thus, the segmentation and tracking algorithms we have built are tuned with conservative thresholds and high confidence levels. Our philosophy is that it is better to miss some details of a trajectory than to introduce erroneous positional information on artifactual spots. Importantly, despite the conservative nature of our approach we show in this paper that sister chromatid separation can be detected at distances as small as 100nm laterally and 200nm vertically. This corresponds to a super-resolution factor of better than 2 in both directions.

**Performance of the Algorithm**

Using synthetic data, we have rigorously tested the performance of our algorithm. We can report the following findings (see also Sec. 3.4.2):

1. Tags of approximately equal brightness can be distinguished at a distance two times lower than the Rayleigh criterion. The actual limit of our algorithm is governed by the SNR. For SNR > 20 the resolution can be pushed to a factor 3. With biological samples such SNR is rarely obtained and a super-resolution
factor of 2 is typical. For SNR < 5 the algorithm breaks down completely. The breakdown level is governed by the confidence probability applied to statistical tests. In some cases with real data we have lowered the confidence to make it possible to process data with an SNR below 5. However, this can only be done at the risk of detecting a higher percentage of false spots. Also, the system is susceptible to differences in tag brightness. For a brightness ratio 2:1 one can loose up to 50% of the super-resolution performance.

2. The localization accuracy for isolated spots is in the single nanometer range for SNR > 20 and 10 – 20 nm for the range 5 < SNR < 20. Interestingly, this accuracy cannot be maintained for tag pairs separated by less than the Rayleigh distance. This arises because of the intimate relationship between resolution and accuracy. The smaller the resolvable distance is the less accurate the estimate of position. However, for relevant SNR values there is no cross-over between resolution and accuracy, consequently the shortest distance resolvable will always be several times larger than the uncertainty of the distance estimate. This is a condition necessary for any further processing of point-to-point distances.

The central component of our algorithm is a routine that fits a mixture model with multiple PSF kernels to clusters of features potentially corresponding to overlapping spots. The exact number of kernels to use is determined in a statistical test. We weigh the significance of an improvement in the goodness of fit of a mixture model containing more kernels against the lower degree of freedom, and hence greater robustness, of a mixture model with fewer kernels. Ideally, the statistical test will pick a mixture model with the smallest number of kernels necessary to represent all significant components of the data. The test we apply in Equation 19 tends to loose sensitivity with an increasing number of kernels and with larger sets of input data, i.e. larger image volumes treated with one mixture model. This behavior is in agreement with our paradigm of conservatism. For example, to add a fifth kernel to a four-kernel mixture model will require much more evidence in the data than to add a third kernel to a two-kernel model. The flipside of this behavior is the demand that the amount of input data is kept as small as possible. In principle, one could fit a mixture model to the entire 3D stack and simultaneously fit all significant local maxima plus potential partner spots, no matter if all kernels in the mixture model are mutually overlapped.
This would simplify the algorithm in that there would be no need to group local maxima in clusters prior to mixture model fitting (see Sec. 3.4.). However, the chance that the software would correctly find multiple spots underneath one of the local maxima would be dramatically diminished. The $\chi^2$-statistics for testing an $n$-order model against an $(n+1)$-order model would be extracted from the residuals in all voxels of the stack. Consequently, the majority of tested residuals would originate from data that is not affected at all by an increment in the order of the model. Local improvements in the model arising from the addition of a supplemental kernel would be considered insignificant compared to the overall fitting quality. To circumvent this problem, we apply mixture model fitting to small clusters of spots, where only kernels are embraced in one mixture model that do indeed overlap.

Despite the strengths of our current mixture model method, we believe that we have not yet arrived at the ultimate limit of resolution for spot detection problems. In a completely different application, we have recently shown in theory and in practice that dynamic processing can increase super-resolution performance (Danuser(2001)). Using the concept of information theory, we have proven that the simple assumption that two objects move relative to one another is sufficient prior knowledge to improve the resolution by a factor 2. In the analysis of yeast chromosome dynamics we should be able to adopt this concept as follows: Starting with a frame in which 4 tags can be localized using the static data processing scheme described in the paper, spots will be tracked differentially forward and backward in time into frames of the movie in which only three spots can be detected. This differential tracking will be supplemented by the prior knowledge that the number of actual objects – and thus the number of tags – is preserved over time. Supposing this, one can seek the optimal parameters, including tag displacement and intensity changes, of a function that maps two of the originally four spots into one merged spot. Such differential tracking is not only expected to improve the resolution but carries the advantage of directly providing chromosome trajectories rather than generating a list of unconnected tag coordinates. In addition, differential tracking allows one to account for currently neglected, complex effects such as the mechanical deformation of tags between frames.
Acknowledgments

The authors acknowledge support from the Swiss National Science Foundation (grant No. 21-59452.99 to G.D.) and from NIH (grant No. GM51464 to P.K.S.)
4. Automatic fluorescent tag localization II: Improvement in super-resolution by relative tracking

Dominik Thomann\textsuperscript{1}, Jonas Dorn\textsuperscript{1}, Peter K. Sorger\textsuperscript{2} and Gaudenz Danuser\textsuperscript{1*}

\textsuperscript{1}Bio Micro Metrics Group  
Laboratory for Biomechanics  
Swiss Federal Institute of Technology  
Wagistrasse 4  
CH – 8952 Schlieren, Switzerland

\textsuperscript{2}Department of Biology  
MIT Room 68-371  
77 Mass Ave  
Cambridge MA 02139

Submitted to the \textit{Journal of Microscopy-Oxford}
Relevance and own contribution
The following paper presents a novel algorithm for tracking fluorescent tags with super-resolution in 3D time series. It is based on the detection algorithm reported in the last chapter and therefore a logical continuation towards the thesis goals.

With regard to the automated information extraction, the additional component introduced by the tracker is the full time series trajectory of the tags. In 3D analysis the estimation of tag correspondence in subsequent time frames is nontrivial. The problem is solved by choosing the optimal configuration over all tags in both frames subject to minimal displacement of the tags and similarity of intensity amplitude. This part of the algorithm was designed and implemented by Jonas Dorn, the first co-author of this paper. The other co-authors are Prof. Dr. Peter Sorger of the Massachusetts Institute of Technology and my supervisor Prof. Dr. G. Danuser.

My main contribution to this work was the implementation of a novel multi-template matching algorithm that can now track features, in our case tags, in relative manner and under signal interference from nearby tags. This algorithm boosts the resolution and accuracy of our framework for chromosome movement analysis. In summary my main contributions consist in:

- Implementation of a module for tracking strategy selection
- PSF model based separation of interfering 3D tag images
- Design and implementation of novel iterative multi-template matching approach including significance testing of the results
- Supervision of Jonas Dorn
**Summary**

We present an algorithm for the 3D tracking of multiple fluorescent sub-resolution tags with super-resolution in images of living cells. Recently, we described an algorithm for the automatic detection of such tags in single frames and demonstrated its potential in a biological system (Thomann, et al. (2002)). The algorithm presented here extends the tag detector with a module for relative tracking of signals between frames. As with tag detection, the main problem in relative tracking arises when signals of multiple tags interfere. We propose a novel multi-template matching framework that exploits knowledge of the microscope PSF to separate the intensity contribution of each tag in image regions with signal interferences. We use this intensity splitting to reconstruct a template for each tag in the source frame and a patch in the target frame, which are both free of intensity contributions from other tag signals. Tag movements between frames are then tracked by seeking for each template-patch pair the displacement vector providing the best signal match in terms of the sum of squared intensity differences. Because template and patch generation of tags with overlapping signals are interdependent, the matching is carried out simultaneously for all tags in an iterative manner.

We have examined the performance of our approach using synthetic 3D data and observed a significant increase in resolution and robustness as compared to our previously described detector. It is now possible to localize and track tags separated by a distance three times smaller than the Rayleigh limit with a relative positional accuracy of better than 50 nm. We have applied the new tracking system to extract metaphase trajectories of fluorescently tagged chromosomes relative to the spindle poles in budding yeast.
4.1 Introduction

In recent years, the tracking of fluorescently tagged objects has become a major goal in live cell microscopy. This development has been accelerated especially by new sensitive electronic imaging devices and the possibilities of labeling specific proteins or protein complexes with Green Fluorescent Proteins (GFP) and GFP-derivates to follow their dynamics under physiological conditions inside the cell. The amount of information gained by such analysis is overwhelming and it turns out that in many cases the bottleneck of live cell imaging experiments is not any more the cell preparation or filming, but the actual extraction of the dynamics. Consequently, many research labs have set off to facilitate this procedure by computational tools for semi-automatic or automatic tracking of fluorescent markers, adopted from conventional computer vision. A recent review comparing the performance of some of the most popular algorithms can be found in Cheezum, et al. (2001).

A particularly important and frequently encountered tracking problem in fluorescence microscopy occurs with the observation of one or several markers concentrated in a sub-resolution region. The image of one such region corresponds to the point spread function (PSF) of the optical system. In this paper we refer to the labeled sub-resolution region as a fluorescent tag whereas we call the image of a tag a spot.

Also for the case of spot tracking, many algorithms have been developed and applied to a variety of biological problems. In our recent paper Thomann, et al. (2002), we have assembled a table with references to the relevant literature and we have discussed the main features of these algorithms in detail. Briefly, most algorithms found are limited to the detection and tracking of well separated spots, i.e. they do not address the problems that arise when signals from two or more tags interfere. Also, they are usually restricted to two-dimensional data only.

For this reason we have initiated the implementation of a new tracking system, which will cope with these problems. The developments are driven by our demands for analyzing large numbers of 3D movies displaying the dynamics of chromosome segregation during mitosis in budding yeast. Chromosome segregation is the process by which replicated sister chromatids are divided equally into two physically separated sets around which daughter cells form. Sister chromatids are
created by DNA replication and bind to the microtubules of the mitotic spindle via a specialized organelle known as the kinetochore. The kinetochore comprises over thirty proteins, many of them thought to be involved in the generation or regulation of the forces necessary to move the chromosomes (He, et al. (2001)). Forces are also generated by the spindle apparatus itself, which represents a complex contractile engine with several classes of microtubules and various motor proteins (Sharp, et al. (2000)). The relative importance of these various components and how they interact to guarantee reliable separation and transport of the chromatids is, however, largely unknown.

Our goal is to build a high-content screen relying on extended statistics of chromosome trajectories and molecular genetics to identify the mechanical function of the kinetochore proteins as well as the force-generating components of the spindle. Real-time analysis of chromosome movement in yeast is made possible by labeling the centromeres (cen) of pairs of sister chromatids with two GFP tags. Two additional tags are employed to mark the position of the spindle pole body (spb) and then the cells are filmed in 3D using a Delta Vision (Applied Precision, Inc.; Issaquah, WA, U.S.A.) imaging system. The details of the biological preparation and microscopy are described in Thomann, et al. (2002).

In terms of tracking, the challenges are to automatically detect and follow the position of the tags in 3D movies containing 1000 and more 2D optical slices. The difficulties are increased by the typically low SNR (~10 and less) of the data, as well as the substantial interference of the tag signals. In many time points, the chromosomes are not sufficiently separated to exhibit two resolvable centromere tags. Otherwise, when the chromosomes are separated, the centromeres tend to localize at only short distances to their associated spindle poles such that signal interference occurs again between the centromeres and spindle pole body tags.

In Thomann, et al. (2002) we have addressed the problem of tag detection under these circumstances. We have shown that with the prior knowledge of the microscope PSF and of the fact that one 3D frame contains the images of an integer number of tags, it is possible to detect and localize tags at sub-Rayleigh distances. We found that with our algorithm it was possible to resolve distances at half the limit of diffraction, which corresponds to a super-resolution factor two. The localization accuracy for a tag imaged into an isolated, non-interfering spot was determined to be
in the range of few nanometers. For overlapping spots the accuracy was found to decrease to about 20 nm.

In the present paper we report an extension of the algorithm to a true tracking of overlapping spots using a multi-template matching approach. Notice our strict distinction between tag detection and tracking. We refer to detection, when we talk about algorithms that segment the raw image data of a single time point into a set of features, in our case spot-like, including information about the feature position. In contrast, we refer to tracking, when we talk about algorithms that follow the positions of already detected features over time. Therefore, with tracking we deal with a relative position measurement, whereas the problem of detection involves absolute position measurements. From the perspective of physics it is well known that relative measurements are more accurate. Specifically for spot tracking problems, one may, for example, realize that absolute measurements of the spot position are susceptible to PSF aberrations, while relative measurements cancel out such effects as long as the aberration is invariant over the tracked distance. Moreover, the use of prior knowledge of the object dynamics, e.g. by supporting the tracking with a model of the potential movement of the feature, will allow an increase in resolution. For bright-field reflection light microscopy of semi-transparent objects we have shown the increase in resolution due to prior knowledge of the object dynamics in theory and experiment (Danuser(2001)). The motivation of the present paper is to test this concept with a novel algorithm for super-resolution tracking of fluorescent tags, which has a much broader impact for particularly biological applications. Although we confine our tests to the specific problem of chromosome tracking, all algorithmic steps and performance analyses are readily adapted to other applications of fluorescence microscopy.

The paper is organized as follows: We start with a brief overview of the tracking algorithm, followed by an in-depth description of each of the main steps (Section 2). In Section 4.3 we first demonstrate the performance of the algorithm on synthetic data and compare it to the spot detector laid out in Thomann, et al.(2002). In the second part of Section 4.3 we show results of tracking spots in biological data. We focus our presentation on the performance and not on the interpretation of the biological data. A manuscript with a detailed analysis of our findings with respect to chromosome movement is in preparation. We conclude with a discussion of the
power and limitations of the current algorithm and give suggestions for further improvements (Section 4).

![Flow chart of the tracking algorithm. The numbers identify the Sections in which the associated algorithms are described.](image)

Figure 15

Flow chart of the tracking algorithm. The numbers identify the Sections in which the associated algorithms are described.
4.2 Algorithm for 3D Spot Tracking with Super-Resolution

In the following section we describe a tracking algorithm based on a frame-to-frame co-matching of multiple templates with partial overlaps. The algorithm consists of three independent modules, each comprising several sub-steps. Figure 15 provides an overview of the processing steps and depicts the data flow between the modules. Input data for the module (2.1) are the raw image data of a 3D movie.

The data is filtered, the spots are segmented and their positions are estimated. All processing steps are performed on a per-time-point manner. The output of the first module is a coordinate list of all spots detected, along with an intensity fingerprint allowing the distinction between bright tags and dim tags. In the second module (2.2) this information is exploited to identify corresponding spots in consecutive frames. As a result we obtain a connectivity graph over the entire time series. The last module (2.4) uses the coordinates and connectivity list to: i) develop an optimal tracking strategy (2.3); ii) to separate the signal contributions in areas where spots overlap; and iii) to perform a simultaneous tracking of all tags between frames pairs applying our new scheme for multi-template matching. In the following we revisit each of these steps and provide the technical details.

4.2.1 Spot Detection

In Thomann, et al. (2002), we proposed an algorithmic framework for spot detection and performed an extensive analysis of its performance. Because of the low SNR of our live-cell data (~10 or less) and the potential overlap of spots, even the initial detection and approximate localization of tags requires sophisticated image processing. The detector starts out with 3D low-pass filtering of the raw image stack to reduce random background noise. For a filter kernel we apply a representation of the PSF. This enhances the signal wherever the intensities of neighboring voxels are spatially correlated according to the intensity distribution of the PSF, whereas noise, assumed to be spatially uncorrelated, is suppressed. In a careful, quantitative comparison of various PSF models we found that for high-resolution, but low SNR data, a Gaussian satisfactorily represents the PSF provides satisfying results (Thomann, et al. (2002)).

After image filtering, possible locations of isolated and overlapping spots are determined by local maxima detection. In practice, many of the local maxima that are found are still associated with noise. To separate local maxima associated with the
signal of an actual fluorescent tag from local maxima due to noise, a "spottiness"-value is computed, which characterizes the similarity of the intensity signal in the neighborhood of a local maximum to the intensity signal of a theoretical spot. It turns out that the "spottiness"-value provides a very robust indicator of spots associated with a real tag.

When two or more tags are separated by less than the Rayleigh distance their combined images give rise to only one local maximum. We test whether the signals of those local maxima with significant spottiness are generated by a single tag or multiple tags. This is achieved by fitting the image signal surrounding each significant local maximum with an n-order mixture model $M_n$ of PSF kernels, where $n$ is the number of kernels in the model. Again, we approximate the PSF with a 3D Gaussian intensity distribution, and write the n-order mixture model as

$$
PSF_i(x, a, c_i) = a_i \cdot \exp \left( -\frac{1}{2} (x - c_i)^T \Sigma^{-1} (x - c_i) \right)
$$

$$
\Sigma = \text{diag}(\sigma_{x_1}, \sigma_{x_2}, \sigma_z)
$$

$$
M_n(x, a, b, c) = \sum_{i=1}^{n} PSF_i(x, a, c_i) + b
$$

Equation 22

with $a_i$ denoting the amplitude of the kernel $i$, and $c_i$ its center coordinates. The parameter vector $\sigma_{x_1}, \sigma_{x_2}, \sigma_z$ describes the lateral ($\sigma_{x_1}$) and vertical ($\sigma_z$) distances from the kernel center, measured in the image coordinate system, for which the PSF intensity drops to 1/e times the center intensity. Given the numerical aperture and magnification of the microscope optics, as well as the dimensions of the lateral and axial image sampling, these values are known a priori (Thomann, et al. (2002)). The number $N$ of actual kernels present in the data, their positions $c_i$ and amplitudes $a_i$, as well as a common background intensity $b$ for all kernels are estimated by sequentially fitting mixture models to the local signal. Starting with $n=1$ the number of kernels is increased until the $\chi^2$-statistics of the residuals of model $M_{n-1}$ is not significantly improved with respect to the $\chi^2$-statistics of model $M_n$. In other words, the system adds as many kernels as needed to obtain a satisfactory goodness of fit, without introducing unnecessary degrees of freedom. Once mixture models have
been fitted to all local maxima the spot detector terminates and returns the number $N$ of tags found, along with their positions and signal amplitudes.

Figure 16
Spot linking between frames. a) Overview of sequence of frames presenting all practical challenges for linking spots: $t_0 \rightarrow t_1$: Linking two frames with an equal number of spots; $t_1 \rightarrow t_2$: Linking two frames with a decrease of the number of spots (but not the number of tags present in the data); $t_2 \rightarrow t_3$: Linking two frames with an increase in the number of detected spots (but not the number of tags).

b) The same series as in a) but shown for real data.
c) Graph construction for the situation of equal number of spots in source and target frame (see text for further discussion).
d) Graph construction in case of spot fusion (first and second layer) and in the case of spot separation (second and third layer).
See text for how the algorithm copes with these cases.

4.2.2 Spot Linking Between Frames
Spot connectivity graphs establish the tag trajectories, i.e. they unambiguously identify the tags in all time points of a movie. This information is critical to both the
relative tracking and subsequent kinematic analysis of tag movements. Figure 16a depicts the problem of spot linking as encountered in our data (with the corresponding image data in Figure 16b). For the sake of simplicity we show an image series with only three separable tags but illustrate all the cases that the linker has to cope with. In terms of our biological application this means that the chromosomes never separate. Three spots are visible in the frames $t_0$ and $t_3$. In frame $t_1$ the signals of spots A and B fuse so that only two signals are visible. Despite this, the mixture model fitting applied during spot detection resolves the two tags that give rise to the fusion spot and reports three tags to the linker. Thus, in frames $t_0$, $t_1$, and $t_3$ are identical for the spot linker, and the task is to identify the connections between tags in two sets of equal size, as for example between frame $t_0$ and $t_1$. In contrast, in frame $t_2$, the spot detector has failed to resolve all three tags and reports only two spots. Here, the linker must cope with a spot fusion, and determine which of the spots arises from the superposition of two tags. Subsequently, the linker also has to cope with the re-separation of spots between frame $t_2$ and $t_3$. As it will become clear throughout this paper, it is essential that links are established not between spots (the image) but between tags (the underlying fluorescent objects).

**Elimination of outlier frames**

Before the computation of a connectivity graph, the linker must eliminate time points in which the detector has missed one or several spots or has included false spots. Such outlier frames can occur when a tag moves out of the image stack or when very low SNR precludes a stable partitioning of significant and insignificant local maxima based on the spottiness measure. To identify frames with missing or extra spots we calculate the time series of the sum of fitted intensity amplitudes

$$l_{tot}(t) = \sum_{i=1}^{N(t)} a_i(t), \quad t = t_1 \ldots t_T.$$  

In a series with no outlier frames, $l_{tot}(t)$ decreases, on average, monotonically because of dye bleaching; but the intensity does not change with spot fusion and splitting as the mixture model fitting of $a_i(t), i = 1 \ldots N(t)$ correctly accounts for the superposition of signals. We run a linear least median square (LMS) regression of the series $l_{tot}(t), \quad t = t_1 \ldots t_T$ (Rousseeuw & Leroy(1987)). The slope of the regression line determines the degree of bleaching. In theory, the bleaching of an
isolated fluorophore is described by an exponential decay of the fluorescence intensity, but we have found the linear approximation representing our data appropriately. Using an LMS estimator for the regression analysis yields a break point of nearly 50%, i.e. up to 50% of all frames may suffer losses of spots or false spots without affecting the estimation of the regression line.

The high break point of LSM regression permits the detection of outlier frames based on the residual statistics of $I_{tot}(t)$. According to Rousseuw & Leroy (1987),

$$\hat{s}_0 = 1.4826 \cdot \sqrt{\text{med}(v_n^2(t))}$$

is a robust estimator for the noise-induced standard deviation of the intensities $I_{tot}(t)$, $t = t_1, \ldots, t_T$. $v_i(t)$, $t = t_1, \ldots, t_T$ denote the intensity residuals relative to the regression line. Residuals associated with outlier frames are assumed to be significantly larger than the noise level. Therefore, frames whose total intensity $I_{tot}(t^*)$ deviates more than $Q_\alpha \cdot \hat{s}_0$ from the regression line are discarded.

The parameter $Q_\alpha$ describes the $\alpha$-quantile of an approximate Student's $t$-test, with $\alpha$ denoting the confidence probability.

**Assumptions for the computation of tag links between frames**

For the frames that remain after outlier removal, the connectivity between tags is established based on three assumptions: i) The number of tags giving rise to various spot configurations over time is constant; ii) The fluorescence intensity of one tag varies only due to noise and bleaching; iii) Tag displacements between consecutive frames are small. Accordingly, the linking module searches for the tag connectivity between two frames that maximizes the similarity in intensity for each spot and minimizes the displacements. Note that this multi-objective problem has to be solved in a global sense, i.e. for all tags simultaneously, to avoid topological conflicts, which occur when linking individual tags.

In the following, we describe the linking procedure for the three cases that occur in our data (cf. Figure 16): i) the linking of frames with an equal number of spots found by the detector modul; ii) the linking of frames with a decrease in the number of detected spots (spot fusion); and iii) the linking of frames with an increase in the number of detected spots (spot separation).
Linking two frames with an equal number of spots

We build a two-layer graph where each tag in layer $t_i$ (a vertex, in the terminology of graph theory) forms all possible candidate connections with all tags in layer $t_{i+1}$ (an edge, in the graph theory). An example of such a graph is illustrated in Fig. 2c. The length of an edge between any vertex $n$ in layer $t_i$ and any vertex $m$ in layer $t_{i+1}$ is defined by the multiplication of two probability density functions, describing the distributions of intensity differences $\Delta I$ and displacements $r$:

$$
\ell_{nm} = \text{pdf}_{\Delta I}(\Delta I_{nm}) \cdot \text{pdf}_r(r_{nm}).
$$

Equation 23

The variables $\Delta I_{nm}$ and $r_{nm}$ define the actual intensity difference and 3D Euclidean distance between the tags associated with vertex $n$ and $m$.

We assume zero mean, normal distributed, noise for the fluorescence intensity with a variance $\sigma^2_I$. Consequently, the intensity differences between corresponding tags in two frames are also zero mean, normal distributed. To ensure a zero mean distribution, the intensity values have to be compensated for photo bleaching. We adjust the intensity value of each tag of the second time point according to $\tilde{a}_k(t+1) = a_k(t+1) - \beta \cdot (t_{i+1} - t_i)$, where $\tilde{a}_k$ denotes the adjusted amplitude of the tag signal, $a_k$ the original amplitude as reported by the spot detector, and $\beta$ the slope of the LMS regression line. The residual histogram of the regression provides the statistics for determining the variance $\sigma^2_I$.

In a similar way we have to determine the distribution of the tag displacements between two frames. Tag movements are, in our case, initiated by the mechanical forces in the spindle, superimposed with thermal fluctuations and global drift of all tags, caused by refoCUSing during filming or by changes in the position of the nucleus relative to the cell cortex. The global drifts are mainly translational and can be compensated by subtracting the shift in the center of gravity between $t_i$ and $t_{i+1}$ from all tag coordinates $C_k(t_{i+1})$.

In a first approximation, the drift compensated tag movement can be considered a random walk, which leads to a chi-squared distribution in the squared norms of the displacements. Obviously, this approximation is an oversimplification as tag movements are induced by the forces generated in the kinetochore and spindle apparatus and therefore have definite directionality. The penalty of zero
displacements imposed by the chi-squared distribution is unrealistic, as counter-balanced forces damp the thermal fluctuations and induce larger deterministic displacements. Until we have a complete mechanical description of chromosome dynamics, which is the long-term goal of our study, we use a normal distribution (positive half-space only) as the probability density function to approximate the probability density function (pdf(r)) of the drift corrected displacements. The variance $\sigma^2_r$ of pdf(r) is determined from the width of the first mode in the histogram of all candidate tag displacements between any pair of consecutive frames throughout the movie (data will be shown in a subsequent paper). Substitution of the normal distributions into Equation 23 yields

$$\ell_{nm} = \exp\left(-\frac{\Delta l_{nm}^2}{\sigma^2_r}\right) \cdot \exp\left(-\frac{r_{nm}^2}{\sigma^2_r}\right)$$

Equation 24

As will become clear in Equation 25 normalization of the probability density functions is not critical.

Given the length of all edges in the graph, the linker module now seeks the sub-graph $G_k$ with the highest joint probability, i.e.

$$G_k = \max_{G_k} \left( \prod_{j \in G_k} \ell_j \right).$$

Equation 25

Here, the index j enumerates all edges belonging to the sub-graph $G_k$. Combining Equation 24 and Equation 25, we define the cost of a sub-graph as

$$c(k) = 1 - \exp\left(-w \sum_{j \in G_k} r_j^2 \right) \cdot \exp\left(-(1-w) \sum_{j \in G_k} \frac{\Delta l_j^2}{\sigma^2_r}\right)$$

Equation 26

and then chose the sub-graph with minimal cost. Only the sub-graphs that establish an unambiguous relationship between the two tag sets are considered, i.e. each tag in $t_i$ must be linked exactly once and only once to another tag in $t_{i+1}$. Figure 16c illustrates the concept of working with sub-graphs. In red, we highlight the correct solution $G_k^*$; in blue another valid sub-graph but with higher cost.
The building of sub-graphs and the search for the one with minimal cost becomes computationally expensive for situations with more than 10 tags. This is not a problem in our current application because we have only four tags present in the images of the yeast spindle. However, in other applications involving many tags, it will be necessary to use sophisticated graph parsing algorithms to find the sub-graph with minimum cost (Sedgewick(2002)). However, by exploiting the fact that our application involves only four tags we do not lose generality and it is computationally feasible to calculate the cost of all possible sub-graphs and to accept the one with the lowest cost as the best match.

Linking two frames with a decrease of the number of detected spots

Figure 16d illustrates the graph between frames with an unequal number of tags as reported by the spot detector. First, we focus on the two top layers \( t_1 \) and \( t_2 \) (cf. also corresponding image data in Figure 16b). The only inputs the linker obtains from the detector are the positions and amplitudes of the spots. From intensity regression the algorithm receives the information that no loss of tags has occurred between \( t_1 \) and \( t_2 \), but that one of the two spots must represent the images of two tags. In such cases of spot fusions, we build graphs with independent vertices for each tag and vertices for tag combinations in all permutations (see layer \( t_2 \)). The associated graph edges are shown in gray and the correct solution is highlighted in red. To calculate the cost of an edge pointing from a single tag vertex in \( t_1 \) to a combined vertex in \( t_2 \) we register each contributing tag in the combined vertex separately. For example the vertex A/B obtains a virtual entry for tag A and tag B. They share identical coordinates and their intensities are propagated from the fusion spot amplitude \( a_{AB}(t_2) \) with

\[
\begin{align*}
  a_A(t_2) &= a_{AB}(t_2) \cdot \frac{a_A(t_1)}{a_A(t_1) + a_B(t_1)} \\
  a_B(t_2) &= a_{AB}(t_2) \cdot \frac{a_B(t_1)}{a_A(t_1) + a_B(t_1)},
\end{align*}
\]

where we assume conservation of the amplitude ratio of A and B in \( t_1 \) and \( t_2 \). In the same way, configurations with a fusion of the tags A and C and a well-separated tag B, as well as with a fusion of tags B and C and a well-separated tag A are calculated. Again, we build only the set of sub-graphs with unambiguous relationships. For
example Figure 16d, each vertex in layer $t_1$ has one edge leaving and two edges link
to a combined vertex (A/B), which automatically defines the target vertex of the third
edge (C). All possible sub-graphs are constructed in this manner and the one with
minimal cost (Equation 26) is selected.

**Linking two frames with an increase of the number of detected spots**
The problem of spot separation, as illustrated in Figure 16 for $t_2$ to $t_3$ can be
reduced to one of the cases already discussed:

i) All fusion spots are identified and their splitting into multiple tags is
completed in the first step, so that the number of *tags* is equal between the
two frames. This problem is thus the same as linking two frames with an
equal number of spots, with the simple modification it is not spots but tags
that are the entities to be linked. This situation is shown for the layers $t_2$
and $t_3$ in the graph of Figure 16d.

ii) The number of tags assigned to spots in $t_1$ is smaller than the number of
spots in $t_{i+1}$. In this case the problem can be reduced to the one of spot
fusion simply by reversing the time direction of linking. The resulting
identification of one or several new fusion spots in $t_1$ will be propagated
backwards along already existing links to previous frames and the
corresponding vertices updated in all layers $t_h$, $1 < h < i - 1$.

When linking is complete, each tag has a unique path in the connectivity graph from
the first to the last frame. Our qualitative assessment yields a success rate of
typically better than 90%. However, one mismatch will propagate throughout the
entire movie. We have decided to implement a graphical user interface for an
efficient user-supported verification of the connectivity graph. The user is allowed to
modify links between tags in consecutive frames and to remove entire frames with
unclear configurations. The verification of a 3D movie with 100 frames takes less
than 5 minutes and thus is affordable even in a screening application.

### 4.2.3 Tracking Strategy

Actual tracking begins with an analysis of the connectivity graph to determine an
optimal tracking strategy. As will become clear during the discussion of multi-
template matching (see below), the robustness of the tracker depends on the
distribution of the spots in the source and target frames. It is advantageous to track
spots from frames with minimal spot overlap to frames with greater overlap. The
best results are obtained when the spots are completely isolated in the source frame.
In this case, accurate matches will be found in the target frame even for overlaps in
which the spot distance is in the range $1\sigma$ of the Gaussian PSF (cf. Equation 22).

Due to this asymmetric performance, a tracking strategy that analyzes frames
progressively starting with the first time point is sub-optimal. Instead, a tracking
strategy module has been implemented to examine the number of spots detected in
each frame and then determines pairs of source and target frames that promise an
optimal solution.

In our application of chromosome tracking in yeast, we make use of prior
knowledge that cells contain four tags and thus, no more than four spots need to be
tracked. However, this information is not required to run the module. In applications,
in which the total number of tags is unknown a priori, the same search for the best
tracking strategy can be employed by assuming that the maximum number of spots
detected in any of the frames represents the actual number of tags.

The tracking strategy module first generates a list $N_{\text{spots}}(t)$, denoting the
number of spots in each time point $t=1...T$ returned by the detector. Next, it extracts
from $N_{\text{spots}}(t)$ a set $T_{\text{start}} = \{t_1^*, ..., t_A^*\}$ of A time points that are quasi-local maxima in
$N_{\text{spots}}(t)$, i.e. they fulfill the requirement:

$$T_{\text{start}} = \left\{ t_j^* \in 1...T \mid \exists s \in 0...T, N_{\text{spots}}(t_j^* - 1) < N_{\text{spots}}(t_j^*) = \ldots = N_{\text{spots}}(t_j^* + s) > N_{\text{spots}}(t_j^* + s + 1) \right\}, \forall j = 1...A$$

Equation 27

The set $T_{\text{start}}$ contains the starting frames for tracking. Beginning with the first entry
in $T_{\text{start}}$ the tags are tracked in forward time direction using a multi-template matching
scheme (see below), until a target frame is encountered where the number of spots
is greater than in the source frame. Then, the tracking direction is reversed, starting
again with the first entry in $T_{\text{start}}$. The backward tracking is again stopped as soon as
the source frame contains fewer spots than the target frame. This procedure is
repeated for all subsequent entries in $T_{\text{start}}$. It is completed when the forward-
backward tracking is finalized for the last entry in $T_{\text{start}}$. Notice that in Equation 27,
the end point of backward tracking starting with the first entry in T_{\text{start}} is given by the very first frame of the movie. Analogously, the forward tracking branch starting with the last entry in T_{\text{start}} ends with the very last frame of the movie. For all other entries in T_{\text{start}} the forward and backward tracking branches of consecutive entries guarantee a perfect handshake.

4.2.4 Tracking Tags Using Multi-Template Matching
After selection of source and target frame pairs, it is possible to track the tags using multi-template matching. The idea of our multi-template matching is to generate as many intensity signal templates as tags have been identified in the source frame and to estimate the parameters of a geometric mapping function for each template, such that the transformed templates are in best match with the intensity signal of the target image. This concept is illustrated by the panels in the top row of Figure 17a. In this example, four tags are detected in the source image S-I, where the images of P_2, P_3 significantly overlap each other. Voxels belonging to the image of one tag are assumed to transform coherently between the source and target image, i.e. all source voxel coordinates x_j^S(S), \forall j \in P_i convert into target coordinates x_j^T(T), \forall j \in P_i according to one mapping function x_j^T(T) = f_j(x_j^S(S), \xi_i). The vector \xi_i denotes the unknown, free parameters of the mapping function. In a system with tags generating completely isolated spots in source and target image, the parameters could be estimated independently for each function f_i, using an ordinary least squares template matching (Danuser, et al. (2000))

$$\xi_i = \arg\min_{\xi^i} \sum_{j \in P_i} \left( l^S(x_j^i) - l^T(f_j(x_j^i, \xi_i)) \right)^2.$$  

Equation 28

Equation 28 requires a list of voxels unequivocally associated with tag P_i. Thus, the tracking module must build a mask M_i that excludes voxels from the background and from spots associated with other tags. For isolated spots only the first condition is important and voxels belonging to P_i can be selected using any spot segmentation method. In our case spots are the image of a sub-resolution tag. Thus, we use a very simple segmentation strategy based on a PSF definition:
Chapter 4. Super-Resolution by Relative Tracking

\[ M_i = \left\{ j \in 1 \ldots K \left| \text{PSF}(x_j, a_i, c_i) > \kappa \cdot \max_{i \in S}(a_i) \right. \right\}. \]

Equation 29

The user-defined parameter \( \kappa \) determines the percentage of the maximum amplitude of all spots contained in frame S for which a voxel located at \( x_j \) is still accepted to be a member of the mask \( M_i \). The parameter \( a_i \) denotes the amplitude of the PSF centered at the spot position \( c_i \). Both amplitude and position are known either from the spot detector or a previous step of tracking in which the current source frame served as an earlier target frame. The condition for voxel index \( j \) in Equation 29 is tested for all voxels in the source image \((1 \ldots K)\).

The generation of masks via a PSF-definition has two advantages over simply thresholding the signals in the source image: i.) Independent of the choice of \( \kappa \), the mask will never contain any purely background voxels although, due to the contribution of noise, many of them would be above the threshold \( \kappa \cdot \max_{i \in S}(a_i) \). This is especially critical in low signal-to-noise image data. ii) The mask is guaranteed to form a convex and completely connected volume.

Reconstructing templates from sources frames with overlapping spots

Template matching fails when several image features overlap in either the source or target frames or in both. Our strategy is to separate the signals of interfering spots in source and target frames prior to multi-template matching (Equation 28). The spots \( P_h \) and \( P_i \) of two tags are considered to be interfering when the mask intersection is not empty, \( M_h \cap M_i \neq \emptyset \). For the example in Figure 17a this holds for the tags \( P_2 \) and \( P_3 \) (note the overlapping red and blue masks in Figure 17a, panel S-II). Signal separation in the region of interference can by achieved voxel-by-voxel using the rules of image superposition: For each voxel \( j \), we calculate a weight \( \rho^j \), indicating the fraction of its total intensity the voxel is contributing to the spot associated with \( P_i \).
Figure 17

Relative tracking of interfering tags using multi-template matching and intensity splitting. For graphical simplicity the algorithms is illustrated in 2D.

a) S-I shows the source image with four detected tags. The center positions are estimated by a mixture model fit (see Thomann, et al. (2002)) and prior knowledge of the microscope PSF permits the assignment of a mask $M_i$ to each of the tags (S-II). Obviously, the signals of tag 2 and 3 interfere. For pixels in this area the signal contribution of each tag can be calculated from the intensity ratios $\rho_1 = l_2/(l_2 + l_3)$ and $\rho_2 = l_3/(l_2 + l_3)$, respectively. The concept of intensity splitting is demonstrated along a profile connecting the mask centers (dashed line in S-II). In our case the tags have spatial extension below the diffraction-limit. Thus, the intensity ratios are a priori known from the PSF shape and their relative positions. For each tag a parametric model $x_i'(T) = f_i(x_i'(S), \xi_i)$ describes the transformation of all pixels of the tag image. Applying this transformation to each mask generates a similar configuration with potential mask overlaps in T-II (not shown). Again the signal contributions from several tags are resolved in areas with interferences of tag images. The goal of multi-template matching is to estimate the parameters $\xi_i$ of each transformation such that the isolated signal in source (template) and target (patch) frame are as similar as possible. This problem is inherently non-linear and must be solved by iterative optimization. Because of the intensity splitting template and patch generation of tags with signal overlaps are mutually dependent, thus requiring a simultaneous matching of all tags.

b) Example of two synthetic and isolated spots in source frame (S-I) tracked into two overlapping spots in the target frame (T-I). T-II displays the propagated mask distribution (blue for $M_1$ and red $M_2$, where voxels with a ½:½ - intensity split appear in violet), allowing the computation of ratio maps $\rho_1$ and $\rho_2$. Multiplication of the maps with the original target image T-II yields two patches $J_1^T$ and $J_2^T$ for both tags. By seeking the best match between the patches and the corresponding template signals in the source frame (blue arrows) the tag displacements $v_1$ and $v_2$ are estimated.
The set of weights \( \rho_i^j \forall j \in M_i \) for the spot \( P_i \) is called the ratio map \( \rho_i \) and is calculated as

\[
\rho_i(x') = \rho_i^j = \frac{\text{PSF}(x'_i, a_i, c_i)}{\sum_{j=1}^{N_{\text{tags}}} \text{PSF}(x'_i, a_i, c_i)}, \ \forall j \in M_i.
\]

Equation 30

The index \( N_{\text{tags}} \) denotes the total number of tags present in the source frame (see also Section 4.2.1).

The bottom row of Figure 17a illustrates the concept of intensity splitting for the noise-free, synthetic data shown in panel S-I. Although we explain the idea based on noise-free 2D the algorithm is implemented to cope with low SNR 3D imagery. Figure 17a S-I displays a situation with two isolated \( (P_1, P_4) \) and two overlapping \( (P_2, P_3) \) spots. Figure 17a S-II shows the corresponding masks labeled with \( M_1 \ldots M_4 \). Intensities of pixels belonging to more than one mask need to be split and distributed correctly between interfering signals. For further graphical simplicity, we limit the illustration of intensity splitting to a 1D profile connecting the centers of the two masks \( M_2 \) and \( M_3 \) (cf. dashed axis in panel S-II). The corresponding image signals represent perfect PSFs, i.e. \( I_2(x) = \text{PSF}(x, a_2, c_2) \) and \( I_3(x) = \text{PSF}(x, a_3, c_3) \) (bottom right panel of Figure 17a). In the case of noise-free signals, the observed signal \( I(x) \) is the sum of the two superimposed PSFs. Hence, the contribution of \( P_2 \) to the signal along the profile amounts to \( I_2(x') = \rho_2^j \cdot I(x') \) with

\[
\rho_2^j := \frac{\text{PSF}_2}{\text{PSF}_2 + \text{PSF}_3} = \frac{I_2(x')}{I_2(x') + I_3(x')}, \quad I_2(x') = I(x').
\]

In the presence of noise and image aberrations the PSF model and effective spot signal are not equal. Here, the ratio map \( \rho_i^j \) provides a noise- and aberration-independent estimate of the contribution of tag \( P_i \) to the total intensity measured in voxel \( j \).

To distinguish between the original source image \( I^s \) and the split signals, we denote the separated image of tag \( P_i \) with \( J_i^s(x') = \rho_i^j \cdot I^s(x') \). In the following we refer to \( J_i^s \) as the reconstructed templates of tag \( P_i \). Since reconstructed templates are only weighted readouts of the original data, they are still perturbed by noise and
aberrations. The PSF model merely enforces the rule for intensity splitting but does not influence the actual characteristics of a template.

**Reconstruction of patches from target frames with overlapping spots**

In the following we describe how the reconstructed templates and intensity splitting rules are applied to tracking each spot from the source into the target frame. In the example of Figure 17a the signal overlap of $P_2$ and $P_3$ is extended to an overlap of $P_2$, $P_3$, and $P_4$. As with ordinary template matching, coordinates of all voxels in the reconstructed template transform into the target frame coordinates according a parametric model $x^j(T) = f(x^j(S), \xi)$. Our current transformation model accommodates the simplest case of a 3D displacement vector $v$, i.e.

$$x^j(T) = x^j(S) + v,$$

with $\xi = v$. More complex models, which would for example account for tag deformations or constrained movement, are straightforward to implement with this approach. However, we have not yet tested their numerical stability. It appears that the pure translational model is suitable for almost all situations in our data, in particular because rotation of sub-resolution tags is not observable.

To estimate the parameters of the transformation model each reconstructed template in the source frame is matched with the signal of the target frame. This requires the extraction of a corresponding patch $J^j_T$, achieved by image interpolation using the transformed coordinates $J^j_T \left( x^j(T) \right) = I^j \left( f(x^j(S), \xi) \right)$. Again, patches of different tags may overlap. Intensity splitting in the target frame is performed as in the source frames (cf. Equation 30), except that the ratio map $\rho^j_T$ for tag $P_i$ in the target image is calculated using the transformed PSF models

$$\rho^j(x^j(T)) = \rho^j(f_i(x^j(S), \xi)) = \rho^j_T = \frac{PSF(f_i(x^j(S), \xi), a_i, f_i(c, \xi))}{\sum_{k=1}^{N_{vox}} PSF(f_k(x^j_k(S), \xi_k), a_k, f_k(c_k, \xi_k))} \quad \forall j \in M_i,$$

Equation 31

where $f_i(c, \xi)$ denotes the transformation of the PSF center coordinates from the source into the target image, $f_i(x^j(S), \xi)$ the coordinate transformation of other voxels contained in $M_i$, and $f_k(x^j_k(S), \xi_k)$ and $f_k(c_k, \xi_k)$ represent the transformations
of all other spots interfering with the target signal of $P_i$. By analogy to the reconstructed templates we obtain a set of corresponding reconstructed patches, whose intensities are given in the voxel raster of the source frame:

$$J_i^T(x_i^j(S), \xi_i) = \rho_i^{jT} \cdot I^T(f_i(x_i^j(S), \xi_i)).$$

Equation 32

**Multi-template matching**

The now separated tag signals are tracked simultaneously until a best match is found between reconstructed templates and corresponding reconstructed patches. Rewritten in the form of ordinary least squares template-matching we estimate the unknown transformation model parameters as

$$\hat{\xi} = \arg\min_{\xi} \sum_{i=1}^{N_{tag}} \sum_{x \in M} \left( J_i^T(x_i^j(S)) - J_i^T(x_i^j(S), \xi) \right)^2.$$

Equation 33

A simultaneous matching of all spots is necessary because of the mutual dependence of the ratio maps. Equation 33 formulates a non-linear optimization problem because of the dependence of the reconstructed patches $J_i^T(x_i^j(S), \xi_i) \forall i = 1...N$ on the unknown model parameters. The problem must be solved iteratively with an update of the ratio map and new interpolation of the patches after each improvement of the parameter estimates. We start the iteration with an initial guess $[\hat{\xi}_1^0 ... \hat{\xi}_N^0]$ derived from the results of the previous spot detection and linking. Typically, these values provide excellent initial parameter estimates, so that the optimum search is started close to the global maximum of the objective function of Equation 33. This allows us to use a Gauss-Newton gradient-based method with rapid convergence. The iterative procedure is stopped when the changes in the components of all translation vectors fall below a user-defined numerical resolution. Details of this optimization method applied to image template matching are described in Danuser, et al.(2000).
To account for the intensity decay due to bleaching when tracking in forward time direction, or the intensity gain when tracking in reverse time direction, Equation 32 is in practice supplemented with a compensation term

$$J_T^T (x_i(S), \xi_i) = \rho_{iT}^T \cdot I_T \left( f_i(x_i(S), \xi_i) \right) - \beta \cdot (t_T - t_s),$$

where $\beta$ denotes the slope of the LMS intensity regression over the entire movie (see Section 4.2.2), and $t_s$ and $t_T$ the absolute time points of the source and target frames respectively. This term gets especially important when tracking spots over time steps larger than the original acquisition rate of the movie. Those situations frequently occur with frame deletions by the spot linking module.

Figure 17b) summarizes the multi-template matching procedure with an example of two synthetic spots isolated in the source image (S-I) but moving in the horizontal direction to generate significant overlap in the target image (T). The distance of the two tags in target the frame is two-fold below the Rayleigh limit of the simulated imaging system. Notice that because of lack of overlap in the source frame we do not explicitly plot the reconstructed templates $J^5_2$ and $J^2_2$. However, the figure demonstrates the splitting of patches in the target frame. To the right of the target frame T we present a color-coded image (T-II) of the two super-imposed PSF masks $\text{PSF}(f_1(x_1(S), \xi_1), a_1, f_1(c_1, \xi_1))$ (in blue) and $\text{PSF}(f_2(x_2(S), \xi_2), a_2, f_2(c_2, \xi_2))$ (in red). Voxels with equal contributions from both masks appear in violet. Each of them is also printed separately above and below T-II. The ratio maps presented above and below the target frame are calculated according to Equation 31 using the final parameter estimates $\hat{\xi}_{1,2}$. A voxel-per-voxel multiplication of these maps with the original target frame yields the reconstructed patches $J^5_T$ and $J^2_T$ that are matched to the reconstructed templates (indicated by blue arrows). The final estimates of the spot displacements $\hat{v}_1$ and $\hat{v}_2$ are depicted by dashed yellow lines.

4.2.5 Testing the Significance of Template Matching Results

After matching, the results of Equation 33 must be tested for statistical significance. First, we compare the a posteriori estimate of the image noise $\hat{\sigma}_p(M_i)$, calculated for
each tag \( i \) from the matching residual with the \textit{a priori} estimate of the noise level \( \hat{\sigma}_0 \), obtained from a statistical analysis of a background area of the image raw data.

\[
\hat{\sigma}_p(M_i) = \sqrt{\frac{\sum_{j \in M_i} \left( \frac{J_i^S(x_j(S)) - J_i^T(x_j(S), \hat{\xi}_j)}{\rho_i^S(x_j(S)) + \rho_i^T(x_j(S), \hat{\xi}_j)} \right)^2}{K(M_i) - p}},
\]

Equation 34

In Equation 34 \( \rho_i^S(x_j(S)) \) and \( \rho_i^T(x_j(S), \hat{\xi}_j) \) denote the ratio maps for template and patch reconstruction, respectively, the latter relying on the final parameter estimates \( \hat{\xi}_j \). The variable \( K(M_i) \) specifies the number of voxels used for the match and \( p \) denotes the number of components in \( \xi_j \). In the absence of systematic matching errors, resulting from an inaccurate transformation model or intensity splitting, the two noise estimates are statistically equal. We test the hypothesis \( H_0 : \hat{\sigma}_p(M_i) = \hat{\sigma}_0 \) against \( H_A : \hat{\sigma}_p(M_i) > \hat{\sigma}_0 \) using Fisher-distributed statistics \( T = \frac{\hat{\sigma}_p^2(M_i)}{\sigma_0^2} \) against a user-defined confidence level \( \alpha \) (usually 90%). When the hypothesis \( H_0 \) is rejected, tracking is aborted and the original coordinates estimated by the detector are returned.

A second test compares the estimated distance \( \hat{d}_y \) between each pair of tags in the target frame with the propagated uncertainty of the distance estimate \( \hat{\sigma}_{d_y} \). (For a general treatment of error propagation in non-linear least squares optimization we refer to Koch(1988); and to Danuser, \textit{et al.}(2000) for its specific application to image template matching). It is logical that, for tracked coordinates to be meaningful, the estimated distance between any two tags has to be greater than the uncertainty associated with positional estimates. Otherwise, an unambiguous identification of the tags is impossible. We test the hypothesis \( H_0 : \hat{d}_y = 0 \) against \( H_A : \hat{d}_y > 0 \) using the test statistics \( T = \frac{\hat{d}_y}{\hat{\sigma}_{d_y}} \), which obeys a Student t-test distribution, again on the user-defined confidence level \( \alpha \).

In case when \( H_0 \) is rejected, the two tags are considered unresolved and the original coordinates determined by the spot detector/linker are returned instead.
4.3 Results and Discussion

4.3.1 Performance Test on Synthetic Data

To determine the performance of the tracking algorithm we analyzed test movies with synthetic spots in which the position, intensity and noise are known for every frame. Spots representing the image of sub-resolution GFP tags were synthesized using the PSF-model proposed by Gibson & Lanni (1991), which includes aberration terms. The parameters were determined with a fit to experimental data (Thomann, et al. (2002)). Spots were then positioned in the 3D image space and the background was filled with a constant intensity value. The amplitude of each spot above background defines the signal magnitude of the tag. Given a nominal SNR for the movie, the noise level was defined as \( \sigma_N = \frac{\bar{a}}{SNR} \), where \( \bar{a} \) denotes the mean spot amplitude for the test movie, and \( \sigma_N \) represents the standard deviation of an additive Gaussian noise field, simulated with a normal distributed random generator and superimposed on the image.

In all simulations we measured the displacement of one spot between two time points under various SNRs and under the influence of a second, interfering spot in the source or target image. An error statistic was calculated over \( N=20 \) simulation repeats:

\[
\varepsilon = \sqrt{\frac{1}{N} \sum_{i=1}^{N} | \bar{d}_i - d_0 |^2 },
\]

Equation 35

with \( \bar{d} \) denoting the estimated displacement vector and \( d_0 \) providing the reference displacement used for movie synthesis. All the error data has been converted to the object domain, and thus expressed in nanometers, using the same voxel size and magnification of our actual experimental setup.

Figure 18 summarizes results of these runs. The performance of the detector is shown in red, whereas that of the tracker is overlaid in blue. Recall that by comparing the results of the detector and tracker we evaluate the differences between an absolute and relative measurement of tag positions.
Figure 18

Performance analysis using synthetic data. For all simulations the displacement of one tag between two consecutive time points were measured and compared to the reference displacements. Error statistics were obtained according to Equation 35 based on $N = 20$ repeats. Errors in lateral displacement are specified by the left ordinate, errors in axial displacement by the right ordinate.

a) Displacement error of a tag generating isolated spots in both time points as a function of SNR. There is no significant difference in the performances of the detector and the tracker.

b) Displacement error as a function of the tag-to-tag distance in the target frame. The respective signals were non-interfering in the source frame. The bottom abscissa shows the tag-to-tag distance normalized with respect to the lateral ($\sigma_x$) and axial ($\sigma_z$) spreading of the PSF (see text for details), which permits a comparison of the breakdown of detector and tracker relative to the Rayleigh limit (black line), independent of the spatial direction of the tag movement. Top abscissa: corresponding values in lateral (XY) and axial (Z) direction.

c) Displacement error as a function of SNR. Tags with non-interfering signals in the source frame are tracked to positions separated by a distance half of the Rayleigh limit in the target frame. The dashed line shows the breakdown of the detector at SNR 8.

d) Displacement error as a function of SNR when tracking tags separated by a distance half of the Rayleigh limit in both source and target frame.
**Accuracy in tracking isolated spots**

In the first run (Figure 18a) we assessed the accuracy of displacement measurements for spots that were isolated in both source and target frames, as a function of the SNR. No significant difference in performance is observable between the tracker and detector. With both approaches we can localize tags with an accuracy better than 30 nm laterally and 120 nm axially, even for an SNR as low as 5. The break-down of both the detector and tracker at SNR = 4 is induced by the initial statistical test used in mixture-model fitting during spot detection. For SNR > 12, which unfortunately is rare in our experimental data, the lateral localization accuracy reaches the single nanometer domain.

The fact that the detector, which relies on the independent fitting of a PSF model to the intensity signal of both source and target frame, performs as well as the tracker, which relies on a relative matching of the signals between the two frames, confirms the high stability of PSF fitting in the case of sub-resolution tags. Despite the systematic deviation of the aberration-free Gaussian PSF model, used for spot detection, from the more complete Gibson-Lanni PSF representation, used for data synthesis, errors in detector displacement estimates are almost free of systematic effects. This can be understood based on the shift-invariance of the optical system. The PSF, and therefore also any incompleteness in the associated model does not change with a shift in the position of the tag and thus, inaccuracies are cancelled out when relative changes in position are determined. Although absolute position estimates determined by the detector do suffer from the incompleteness of the Gaussian PSF representation. Therefore, in this specific application to sub-resolution tag tracking there is no advantage of a relative tracking approach. However, relative tracking would be advantageous as soon as the objects dimensions exceed the diffraction limit, so that object rotation and deformation become measurable. The same holds for image analysis in a linear, shift-non-invariant microscope (Thomann, et al., 2001).
Accuracy in tracking deteriorates with signal interference but the tracker is less sensitive to it than the detector

In a second test using synthetic data (Figure 18b), we analyzed the influence of overlapping second spot in the target image on the accuracy of the displacement estimates. For this we tracked two tags generating isolated spots in the source frame into an overlapping configuration in the target frame. The lateral tag-to-tag distance in the target frame is specified in units of the parameter \( \sigma \) of the Gaussian PSF model, expressed in the object space (abscissa at the bottom of Figure 18b). As derived in Table 2 in Thomann, et al. (2002), \( \sigma \) relates to the numeral aperture NA and emission wavelength \( \lambda \) with \( \sigma = 0.21 \cdot \frac{\lambda}{NA} \). In axial direction, the spreading of the PSF is approximately three times larger. We introduce a second parameter \( \sigma_z = 3.1 \cdot \frac{n}{NA} \sigma \), with \( n \) being the refractive index of the sample medium, and specify axial tag-to-tag distances in units of \( \sigma_z \). Rayleigh limits in the lateral and axial direction, i.e. the first roots of the PSF in these directions, as expressed in units of \( \sigma \) and \( \sigma_z \) are \( d_{x,y}^R = 2.9 \cdot \sigma \) and \( d_z^R = 3 \cdot \sigma_z \), and thus essentially the same (see black solid line in Figure 18b). This permits the presentation of distance-dependent performance data independent of the spatial direction. For a better interpretation of the results, tag-to-tag distances are converted to the lateral (XY) and axial (Z) domain by the top abscissa of Figure 18b, using \( \lambda = 525 \text{ nm}, \ NA = 1.4 \), and \( n = 1.33 \) in accordance with the optical parameters of our experimental setup.

The results of three runs with different SNR are shown. A comparison of the error characteristics of detector (red) and tracker (blue) reveals that at all SNR levels the detector breaks down significantly earlier than the tracker. As explained in Thomann et al. (2002) the break-down of the detector is defined as the tag-to-tag distance at which the mixture model fitting of the corresponding spot rejects the addition of two or more tags as a significant improvement to the goodness-of-fit. In contrast, the tracker breaks down when the propagated error of the tag-to-tag distance precludes an unambiguous identification of the tags, that is, when the distance between tags is smaller than the uncertainty in the estimation of positions (see Section 4.2.4). At all SNR levels, tracking is consistently possible down to tag-
to-tag distances three times below the Rayleigh limit. However, shorter tag-to-tag
distances in the target frame are accompanied by an increased error in the
displacement estimate. At the breakdown of the tracker, the tag separation is almost
the same as the displacement error, rendering the re-identification of tags in the
target frame impossible. Indeed, the statistical tests of the tracking results correctly
detect this weakness and reject the resolution of two tags in this case. Of course, the
statistical tests applied after tracking (see Section 4.2.4) cannot exploit the error
measures presented in Figure 18b. The latter are derived from an evaluation of the
differences between estimated displacement and reference displacement posterior
to tracking. In contrast, the tracker relies on a propagation of the matching parameter
uncertainty into an estimate of the tag-to-tag distance error, which appears to
excellently agree with the here presented error analysis on synthetic data.

The tracker is stable with signal interference in source and target frame, but
the detector is not
In Figure 18c) and d) we examine how accuracy and resolution limits vary with tag
separation in the source frame as a function of the SNR. Figure 18c presents the
situation where the tags generate isolated spots in the source frame. Figure 18d
displays the results of a run where the tags start out at a distance of $\sim 0.5 \cdot d_{xy}^R$ and
thus with strong signal interference. For both graphs the tags are tracked in lateral
direction into a target frame configuration with a tag-to-tag distance of $0.5 \cdot d_{xy}^R$.

The comparison of graph Figure 18c with Figure 18a indicates the effect of
spot overlaps in the target image. Two effects are observable: i) For almost all SNRs
the displacement error is twice as high with overlaps than without, indicating that the
signal interference introduces uncertainties in spot localization for both the detector
and the tracker. ii) For a tag distance half of the Rayleigh distance the detector
breaks down at SNR 8, a conclusion that had been drawn previously (see Figure 7a
in Thomann, et al. (2002)). The dashed line connecting the graphs in Figure 18c and
a emphasizes that there is no such break-down for the detector with isolated spots.
The tracker overcomes the resolution limit of the detector and does not break down
until SNR=4. The displacement error at this point measures $\sim$60nm, still sufficient for
tag re-identification with 90% confidence in a configuration with a tag-to-tag distance
of 115nm.
The gain in resolution and accuracy using relative tracking is even more prominent in the results displayed in Figure 18d. With signal overlaps in source and target frame the detector breaks down at SNR=10 and the error increases rapidly between SNR = 10 and SNR=15. This suggests that the detector suffers substantial localization biases in both source and target frame, which do not compensate but instead add up in the calculation of displacement from positional differences. The tracker is clearly more stable. The error of the estimated displacement increase only marginally in comparison to Figure 18c and again does not break down until SNR = 4.

4.3.2 Influence of Initial Values
To test the robustness of the multi-template matching algorithm we examined the convergence of the iterative solution of Equation 33 with various initial parameters. For the source image the input parameter set consists of the center coordinates \( c_i(S) \) of the tags, the fitted or propagated amplitude \( a_i \) and the parameter \( \sigma \) of the Gaussian PSF. For the target frame the initial parameters comprise the center coordinates \( c_i(T) \) of the tags, which are subsequently updated by the matching procedure. All simulations were carried out with an SNR=20. Since the detector localizes the tags with an accuracy of 20nm laterally and 80nm axially for the SNR range relevant to our data, we confined the simulated perturbation of initial values to these limits. Notice that during the evaluation of an entire movie, most tag positions in the source frame are determined by the result of a previous tracking step. As seen in the results in Figure 18, the tracking is at least as accurate as the detector. Hence, our choice of boundaries for the variation of initial values is also relevant in this situation.

**Perturbation of initial tag coordinates in the target frame**
Figure 19a displays the error of a lateral displacement as a function of the initial guesses of X- and Z-coordinates of one of the two tags in the target frame. For the graph shown, the nominal tag-to-tag distance in the target frame was set to \( 0.5 \cdot d^R \), starting with non-interfering tag signals in the source frame. The planar and horizontal error surface indicates that independent of the initial guess the template matching converges consistently to the same displacement measurement with an
accuracy in the range found in Figure 18c for the same data conditions (SNR, tag-to-tag distance in source and target frame). We have repeated the simulation for various SNR levels and spot overlaps in source and target frame and always found the same error plane, merely positioned at different \( \varepsilon_{xy} \)-levels, as one would expect from the results presented in Section 4.3 (data not shown). This finding demonstrates the required robustness for our matching approach. The guesses of tag positions in the target frame are obviously the weakest of all initial parameters, but simulations show that they have little influence on the final solution.

**Perturbation of initial tag coordinates in the source frame**

Figure 19b) – d) draw a completely different picture for perturbations of the initial X- and Z-coordinates of the two tags in the source frame. In contrast to the tag positions in the target frame, these values, once set, remain unchanged throughout the matching. Again, we display the error in measuring a lateral displacement as a function of coordinate shifts in one of the two tags. For all three panels the lateral tag-to-tag distance in the target frame was half the Rayleigh limit. Panel b) presents the result for non-interfering tag signals in the source frame, panel c) for tags separated by \( d_{xy}^0 \), and panel d) for tags separated by half the Rayleigh distance.

The closer the tags in the source frame are, the more robust the matching appears to be against deviations of the initial tag coordinates. At first sight this finding seems counterintuitive. It can be explained, however, by taking a closer look at the role of the initial tag coordinates in the source frame: They are employed to set the center of the masks \( M_i \) and \( M_j \), which define the two reconstructed templates. Shifts in the mask position yield imprecise template signals. As explained in Sec. 4.2.4 the multi-template matching algorithm iteratively maps the masks of all tags from the source into the target frame, using the currently best guess of each tag displacement to reconstruct best matching patches. When tracking two isolated spots into two isolated spots a mask that is shifted at the start has no effect since the same mistake is made during template and patch reconstruction (data not shown). However, when tracking two isolated spots onto two overlapping spots, the reconstructed template and patch differ significantly and the supposedly optimal transformation can yield wrong displacement estimates.
Chapter 4. Super-Resolution by Relative Tracking

Figure 19

Influence of initial values on tracking accuracy.

a) Variation of the initial guess of the tag displacement from source to target frame. The nominal tag-to-tag distances in source and target frame were kept constant at half the Rayleigh limit.

b) – d) Perturbation of the tag position in the source frame reflecting to a potential error in tag localization associated with the spot detector or a preceding tracking. The nominal tag-to-tag distances in the source frame were varied from a non-interfering configuration (a), to a distance corresponding to the Rayleigh limit (b), and half the Rayleigh limit (c). The nominal distance in the target frame was kept constant at half the Rayleigh limit. See text for an interpretation of the graphs.

The transformation of artificially shifted mask into the target frame causes errors in intensity splitting, accompanied by invalid intensity contributions of the first tag to the second patch and vice versa. The error in the displacement estimate is most dramatic when initial shifts occur in the direction of tag movement, i.e. in our case for mask shifts in the X-coordinate (Figure 19b).

The closer the tag configuration in source and target frame, the less severe is the influence of a shifted mask. For example in Figure 19c the signal overlap in the source frame (tag-to-tag distance equals to $d^R_{xy}$ ) is very similar to the configuration in the target frame (tag-to-tag distance equals to $0.5 \cdot d^R_{xy}$ ). This greatly improves the error performance. Obviously, the error is minimal when the tag configuration in
source and target frame are identical (Figure 19d). However, in contrast to the situation with non-interfering tag signals, the error is in no case immune against mask shifts. Whenever spots overlap, mask shifts unavoidably lead to mistakes in the intensity splitting accompanied by mismatches between the reconstructed template and patch.

**Perturbation of the size of the PSF**

For a correct intensity splitting the choice of the parameter $\sigma$ of the Gaussian PSF turns out to be as crucial as the setting of the initial mask positions in the source frame. As mentioned in Section 4.3, we currently employ a $\sigma$-value calculated from the theoretical optical parameters of our microscope (Thomann, et al.(2002)). Deviations of the theoretical value of more than 10 % from the actual PSF lead to weak and unstable convergence of the matcher, with consequent displacement errors. To avoid this problem we have begun to estimate a $\sigma$-value for every movie.

### 4.3.3 Influence of the Objective Function

The sensitivity of the tracker is governed by the ratio between the magnitude of gradient of the objective function relative to the tag displacements and noise. Stable convergence is ensured as long as variations in displacement alter the objective function above noise-induced fluctuations. To investigate the theoretical limit in determining displacements, we constructed maps of the objective function for two tags exactly co-located in the center of an image. We synthesized a set of auxiliary frames where the two tags were systematically displaced from the center position by various lateral (x, y) and axial (z) distances. These frames can be understood as intermediate results of the iterative least squares optimization, where the vectors $c_1 = [c_{x1}, c_{y1}, c_{z1}]^T$ and $c_2 = [c_{x2}, c_{y2}, c_{z2}]^T$ describe locations visited by tag 1 and tag 2 on their way to the final position. We define the image of one tag as a Gaussian $G(x,c)$ with unit intensity amplitude. The vector $x$ denotes the 3D coordinates of a voxel in either the target or an auxiliary image and $c = [c_x, c_y, c_z]^T$ the center of the tag.
Figure 20

Least-squares objective function for tracking two noise-free tags with unit intensity, co-localized at the center of the frame. The color-encoded values of the function display the square root of the sum of the squared intensity differences between an auxiliary image where the two tags are displaced from their nominal positions by the amount indicated on the panel ordinates and abscissas (expressed in voxel units) and the target image, divided by the number of voxels contributing to the sum. See text for a detailed explanation of the construction and meaning of the panels. Nomenclature and definition of the displacement coordinates are indicated by the panel on the bottom right.

According to these definitions noise-free auxiliary and target images are then written as $A(x, c_1, c_2) = G(x, c_1) + G(x, c_2)$ and $T(x) = G(x, 0) + G(x, 0)$. The associated objective function is calculated by the sum of squared intensity differences for all voxels for which $G(x, 0) > 0.05$:

$$\phi(c_1, c_2) = \sum_{x \mid G(x, 0) > 0.05} (T(x) - A(x, c_1, c_2))^2.$$
By analogy to Equation 34 we normalize the least squares objective function with the total number of voxels \( K \) involved in the sum

\[
\chi^2(c_1, c_2) = \frac{\phi(c_1, c_2)}{K}.
\]

Equation 36

The square root of Equation 36 has the dimension of image intensity and thus can be compared to image noise. In the spatial domain \( \sqrt{\chi^2(c_1, c_2)} < \sigma_0 \) the optimum search is controlled by noise-induced local minima, which yields arbitrary results. Notice that since the amplitude of the Gaussian signal is set to 1, \( \sigma_0 \) is \( 1/\text{SNR} \) for a given SNR.

Figure 20 displays \( \sqrt{\chi^2(c_1, c_2)} \) for various pairs of displacement components in tag 1 and tag 2. Each of the panels represents a two-dimensional slice of the six-dimensional space for the minimum search. The bottom right panel depicts the definition of the coordinate system. The top-left panel color-encodes the evolution of the objective function when source tag 1 converges into the target tag position from the left and source tag 2 from the right i.e. \( \sqrt{\chi^2([-c_{x1}, 0, 0], [c_{x2}, 0, 0])} \). Analogously, the top-center panel \( \sqrt{\chi^2([-c_{x1}, c_{y1}, 0], [c_{x2}, 0, 0])} \) encodes the evolution when tag 1 converges at 45° between the negative x- and positive y direction and tag 2 from the positive x-direction.

The first conclusion to be drawn from the graphs is that, within the parameter ranges shown, the objective function decreases monotonically towards one global (and correct) minimum. Consequently, with initial guesses of the target coordinates not farther than \( \pm3 \) pixels (\( \pm150 \) nm laterally and \( \pm600 \) nm axially) from the true but unknown tag positions the iterative matching is guaranteed to converge towards the correct solution (deep blue).

Second, these graphs allow us to carve out the domain of deterministic convergence given a certain SNR level. For instance, with an SNR = 8 the domain where the objective function falls below 0.125 (transition between blue and dark green along the color scale) is dominated by noisy local minima and the actual point of convergence is random.
As these panels are calculated for two co-localized tags, Figure 20 represents a direct graphical representation of the ultimate limit in resolving two tag signals under noisy conditions using least squares matching. At SNR=8 the domain begins for most spatial directions at a distance ~1.5 pixel units from the global minimum. We conclude that reliable tracking of two tags less than ~1.5 pixel apart is essentially impossible with SNR = 8. Converted to the object coordinates, the domain boundaries are located at 75 nm laterally, and 300 nm axially. These values are in excellent agreement with Figure 18b, where we found the break-downs for tracking with SNR = 8 at ~90 nm laterally and ~290 nm axially. Although the statistical definition of the break-down utilized by the tracker is unrelated to the shape of the objective function (which is unknown, in general; see last paragraph of Section 4.2.4), it turns out that error propagation applied to estimate the difference in tag position is an appropriate means to assess the limits of matching two noisy signals.

4.3.4 Performance on Biological Data

The development of the super-resolution tracker presented here is motivated by our demand for accurate and complete chromosome trajectories in order to analyze the force-generating processes in yeast mitosis (He, et al.(2000); Thomann, et al.(2002)). Therefore, we conclude this section with a demonstration of how our new algorithm performs in tracking sub-resolution GFP tags in extended and noise perturbed time-lapse image series. We confine our presentation to the analysis of the algorithmic performance. A second paper is in preparation, where we evaluate several hundred movies of wild-type and genetically mutated yeast strains and present biological interpretations of chromosome trajectories.

More tags are resolved by the tracker than by the detector

Figure 21a analyzes the tracking performance for the same two movies examined in Figure 5a and 5b in Thomann, et al.(2002). The graphs compare the number of tags per frame extracted by the detector (red) and by the tracker (blue). Most frames with zero spots were automatically eliminated by the spot linker since they represent frames in which tags were missed during data acquisition or spot detection (Section 4.2.2). For both movies we found that accurate tracks can be extracted which bridge gaps of several missing time points.
Figure 21a indicates that in most cases where the frame was accepted, but only three spots were found by the detector, the tracker is now capable of recovering the missing tag. The statistics over both movies reveals that ~80% of the spots overlooked by the detector were correctly found by the tracker. Detector misses occur in two situations: i) When two tags are separated by less than the detection limit and in these situations, the new tracker can step in and improve the resolution limit; ii) When the isolated signal of a single tag is too weak to be considered a significant image feature above noise. In this situation the tracker generally cannot improve performance because the tracker is not designed to find new features and relies completely on the detector for the locations of significant signals (an exception of this rule is discussed below along with panel Figure 21b).

In Figure 21b two frames are highlighted in which the tracker was unable to recover the spot missed by the detector (green arrows). In this case the tracker would not accept that two tags exist as their separation was below the distance uncertainty propagated by the least squares matching framework (see Section 4.2.4).

**The tracker can recover undetected spots**

Figure 21b shows an example where the detector failed to extract one spot in the target frame because of an insufficient local SNR (panel T1). All panels show maximum intensity projections in z-direction after noise filtering. After these manipulations the spot seems visible to a human observer. However, a manual accurate localization in 3D by browsing the various slices of the image stack would be impossible. In the source frame S noise was less critical, so that the detector scheme correctly extracted all for tags. None of the tag pairs is separated by less than the Rayleigh distance and the SNR of each tag is above the critical value. Thus localization was straightforward. As explained in Section 4.2.2 the spot linker handles the spot disappearance between source and target frame with the introduction of a fusion spot (cf. T1). Yet, when tracking the tags from S to T using the initial coordinates of the red/orange tag in T1, the matching prefers spot separation and recovers the missing spot in T2.
The mutual influence of tags coupled in one objective function increases the robustness in localization

In addition to recovering spots missed by the detector, the tracker increases the robustness in localization, as illustrated in Figure 21c. Due to an interference with the spot marked with a green cross (the 'green spot' or 'green tag'), the spot at position T was missed by the local maximum filter applied by the detector module to initialize mixture model fitting (see Thomann, et al. (2002)). Nevertheless, because of the proximity of the second signal, the detector module recognized the presence of a second tag but estimated its coordinates at position D, far too close to the green tag. Using these coordinates as the starting values for the tracker, the least squares matching moved the tag position to the significantly more accurate position T.

The reason for this preference can be understood from the residual fields displayed in the second and third panels of Figure 21c. They present the center x, y - slices of the intensity difference $J^s_{\text{green}}(x^i_{\text{green}}(S)) - J^T_{\text{green}}(x^i_{\text{green}}(S), \hat{x}_{\text{green}})$, subject to the simultaneous tracking of a second tag whose initial guess of the position is given by the detector estimate D and whose final coordinates are at the tracker position T. The index 'green' in the difference equation indicates that the residuals are displayed for the green tag. Panel D+ shows the residual field after initialization of the matching procedure, the panel T+ after the last iteration of least squares optimization. The amplitude of the green spot is 15 gray values above background (using 8-bit image data). The residual in panel D+ exhibits a non-random distribution with magnitudes between +10 and -6. The systematic asymmetry originates from the subtraction of the wrongly suspected tag signal D in immediate proximity to the green tag $J^T_{\text{green}}(x^i_{\text{green}}(S), \hat{x}^0_{\text{green}})$. Notice the strong positive peak dominating the residual field at the initial position of the tag D (red arrow), which suggests that the reconstructed patch for the green tag had a too low an intensity compared to the unperturbed template. Conversely, the residuals to the right of the diagonal (yellow dashed line) are systematically negative. Because of the inaccurate splitting of the mistaken tag signal D, the reconstructed patch intensities for the green tag get, on average, too large in this region.
Figure 21

Performance of the tracking algorithm on biological data.

a) Comparison of the number of tags found per time point by the detector (red) and the tracker (blue) for two movies. The tracker is able to recover about 80% of the tags missed by the detector. Time points with zero tags were automatically removed by the algorithm because of incomplete data, e.g. when tags moved out of the image stack.

b) Example of the recovery of a tag missed by the detector. S: source image for tracking; T1: Tag positions for the target frame as identified by the detector and linker module. The orange tag is missed by mixture model fitting. T2: Recovery of the orange tag after running the tracker.

c) Example of the increase of robustness in tag localization due to tracking. Left panel: Overview of the target frame in maximum intensity projection. The detector accepts tag D as a significant secondary signal to the green tag, but misses the tag clearly visible to our eyes, and labeled with T, as it is properly recovered by the tracker. D+: Residual field of the green spot after initialization of the least squares matching with a tag at position D. Due to the off-splitting of a mistaken signal contribution from tag D, the residuals have a non-random distribution with magnitudes almost as large as the tag amplitude. T+: Residual field after completed least squares matching. The field has a random intensity distribution with a significantly lower magnitude. D,T: Residual field of the red spot after initialization (D) and after optimization(T).
The situation is sorted out during iterative multi-template matching and the residual field of the green tag after completed optimization (T+) does not contain such systematic distribution. Expressed in terms of $\hat{\sigma}_p^2(M_t)$ (Equation 34) the residuals decreased from $\hat{\sigma}_p^2(M_D) = 9.2$ to $\hat{\sigma}_p^2(M_T) = 6.3$. Panel D and T display the residuals of a central slice of the red tag before and after iteration. As one would expect, panel D delivers almost the mirror image of D+ with the slice center serving as the mirror point. Again, the systematic asymmetry disappears with the completed matching. To summarize, this example demonstrates the mutual influence of tags by including template-patch differences of all tags in one objective function.

**Summary: Visualization of the multi-template matching algorithm for two centromeres tags**

Figure 22 summarizes the concept of multi-template matching with adaptive intensity splitting for a biological example. In Figure 22a we present an overview of two frames where our new relative tracking breaks through the conventional limits of resolution in light microscopy. The distance between the two centromeres tags (cen1, cen2) in the target frame were estimated as $d_{xy}(T) = \sim 195\text{nm}$ and $d_z(T) = \sim 60\text{nm}$, starting from a distance in the source frame of $d_{xy}(S) = \sim 275\text{nm}$ and $d_z(S) = \sim 70\text{nm}$. Panel Figure 22b illustrates the mechanisms of intensity splitting similar to the synthetic example in Figure 17b. In the source frame S the detector localized two tags with interfering signals. In the target frame T only one spot representing the fusion image of two tags was reported by the detector. For both frames S and T we display the central xy-slice of the raw image stack to illustrate the challenge posed to the detector. Because of the extremely low SNR it failed to extract two spots in this image. The overlaid circles point out the positions, as recovered by multi-template matching.

We restrict the presentation of template reconstruction to the final two intensity fields $J_1^S$ and $J_2^S$ (center slices), whereas we again illustrate the entire reconstruction of the patches $J_1^T$ and $J_2^T$ for the last iteration of the matching: Using the estimate (at this point, the final estimate) of the tag displacements, the positions of the overlapping masks $M_1$ and $M_2$ are propagated from source to target frame (cf. color-coded illustration with $M_1$ in red and $M_2$ in blue resulting in a violet color for voxels
with approximately a $\frac{1}{2} - \frac{1}{2}$ intensity split. From their overlaps the ratio maps $\rho_1$ and $\rho_2$ are calculated, which serve as the multiplication factor for reconstruction in frame $T$.

a)

b)

Figure 22

Super-resolution performance demonstrated for the separation of two centromere tags in a yeast mitotic spindle.

a) Source and target frame where the centromere tags are located at a distance above (S) and below (T) the Rayleigh limit. The distance specified is measured in 3D. Notice that the tag signals interfere in both S and T.

b) Resolution of the two tags by multi-template matching and intensity splitting. The schematic is the same as in Figure 17b). Since intensity splitting is necessary also in S both the reconstructed templates $J_{12}^S$ and reconstructed patches $J_{12}^T$ are explicitly displayed for both tags.
4.4 Conclusion

We have presented a novel approach to tracking fluorescent signals of sub-resolution objects in 3D, whose images can be accurately represented by the PSF of the optical system. This paper represents a substantial follow-on to the work described in Thomann et al (2002), where we introduced a framework for the automatic detection and localization of such fluorescent tags. In our earlier work we were already able to localize tags with super-resolution, that is, at tag-to-tag distances beyond the cut-off frequency of the optical transfer function (OTF). With the new algorithm we improve the separation of two tags from a super-resolution factor 2 to 3. For the sake of simplicity we have not used the cut-off frequency as the measure of conventional resolution but compared all our data to the Rayleigh limit which is inverse proportional to the OTF cut-off and a more intuitive resolution criterion for most microscopists.

Besides resolution improvement the algorithm has proven to be very precise in tracking tag movements. The smallest displacement measurable is limited by the localization accuracy, which in turn depends on the SNR of the image and the degree of signal interference between the tag and other fluorescent objects. For our high-resolution microscope setup, the localization accuracy is of the order of a few tens of nanometers laterally and below hundred nanometers axially. Thus, localization accuracy is typically smaller than the distance resolvable between two interfering tags. The high accuracy is achieved by relative measurements using a geometric mapping of the noisy and aberrated tag signal between two frames. The parameters of the mapping functions are estimated by simultaneously matching of all tag images between source and target frame.

The problem with relative tracking is, however, that an initialization with approximate tag displacements and a complete map of the frame-to-frame correspondences of tags is required. To meet this requirement we employ the output of the spot detector and have supplemented the detection scheme with a module for tag linking. A special feature of this linker module is that it can cope with spot
fusions. It assigns multiple co-localized tags to one spot when the detector has failed to resolve all tags by mixture model fitting.

A further step towards the fully automatic analysis of tag movements has been accomplished by the automatic removal of outlier frames. The analysis of the evolution of the total tag intensity per frame throughout a movie delivers a robust estimate of the signal decay due to bleaching. Frames, which deviate from the general decay pattern, are reliably detected and eliminated. It appears that outlier frames arise when one or more tags come close enough to the edge of the image stack that a significant fraction of the PSF signal is lost.

Despite the robustness of the tracker, we observed small drifts in the position estimates of overlapping spots because of a slightly inaccurate splitting of the intensity data. For long time series such drifts add up as we use the output coordinates of the tracker from one frame pair as input coordinates for the tag positions in the source frame of the subsequent pair. Currently, the accumulation of such drifts is only stopped at frame pairs with an unequal number of detected spots, making a reinitialization of the tracker necessary. We will further improve the situation by a refinement of the module that defines the tracking strategy. In a future version we will restrict the selection of source frames as much as possible to time points with no signal interference. Instead of building source and target frame pairs from consecutive time points, the new strategy will be to combine frame pairs in a staggered fashion, i.e. by tracking a series starting at \( t_0 \) in the fashion \( t_0 \rightarrow t_1, \ t_0 \rightarrow t_2, \ldots, t_0 \rightarrow t_s \). This will lower the risk of cumulative drifts, but poses more challenges for the template matching itself, as signals over large time steps may be subjected to variations that are not included in the geometric transformation model and the bleach compensator. Therefore, we will have to define a maximum time step \( s \) for which staggered tracking is permitted.

Another measure to further improve the robustness of tracking is to incorporate additional prior knowledge of the tag movement. In our case we will soon have such information available, as we are beginning to build a mechanical model of the yeast spindle based on the existing tracks. Such a model will increase the robustness of the spot linker by suppressing mechanically non-sensical links. As mentioned in Section 4.2.2 we are currently forced to remove such problems by manual verification of the connectivity map of each movie. Although the procedure is
supported with a comfortable graphical user interface it is still too tedious to be tolerated in an image-based screen of thousands of movies. In addition, mechanical models will be valuable to constrain the search space of the transformation parameters $\xi_i$ of each tag.

**Acknowledgments**

The authors acknowledge support from the Swiss National Science Foundation (grant No. 21-59452.99 to G.D.) and from NIH (grant No. GM51464 to P.K.S.)
Chapter 5

5. Conclusions

The research in this thesis is characterized by an interplay of machine vision, super-resolution schemes in microscopy and analysis of sub-cellular processes. The thesis reports a consistent framework for the automated detection and tracking of fluorescent tags with super-resolution in 3D for the analysis of yeast chromosome movement during mitosis.

The general motivation behind the presented framework is to build a test application for an image-based screen of cell phenotypes. With today's knowledge and possibilities of cell biology in terms of molecular manipulation, e.g. by specific genetic mutation, combined with the advances in electronic imaging methods, the main challenge of image-based screens is often not the experimentation and observation, but the extraction of relevant data and its interpretation in terms of mechanistic or quantitative models of cellular processes. A particularly important example has turned up with the possibility of high throughput gene identification. In recent years the complete genome of many species have been identified. Although genes can nowadays be easily identified, their functions and importance for the organisms is largely unknown. One of the predominant questions in modern biology is therefore the decoding of the genomes in terms of function. Another example, closely related, is the problem of protein function mapping. The protein composition of many complexes in cells is known, although their individual function and importance is still an enigma.

One way to answer these questions is to compare large sets of image samples showing the live cell behavior of wild-type versus (well controlled) genetic mutant cells. The goal of such experiments it to delete specific proteins followed by a characterization of the cell phenotypes. An interpretation of the changes in the dynamic behavior relative to the wild-type behavior will then reveal some of the proteins function. As there are always variations between cells and the behavior is also dependent on the environment, such an interpretation is only possible when the statistics of tens to hundreds of samples is compared.
Such a system has two major impacts: a) It allows addressing some of today's most relevant scientific questions in cell biology and b) has an enormous economical potential. Companies are investing large efforts to implement software systems for robust and automated extraction and interpretation of light microscope data. The economic pressure for such tools is mainly exerted by the pharmacological industry, which aims at exploiting movies of dynamic cellular processes in high-content screens of cell phenotypes for more targeted drug discovery, as well as the efficient mapping of gene and protein functions.

The tools presented in this thesis aim at the specific application of protein identification responsible for force generation in chromosome segregation of budding yeast, as described in detail in Chapter 3. The scale on which the organelles (sister chromatids, kinetochores, microtubules and spindle pole bodies) in yeast cells interact is close to the boundary or sometimes even beyond what is observable with conventional light microscopy (<1 μm). For a complete analysis of these processes, it is necessary to push the limits imposed by resolution and accuracy as far as possible. The reported framework is capable of extracting the 3D trajectory of a finite number of fluorescent sub-resolution tags in low SNR data even when their signals strongly interfere.

### 5.1 Summary of the most important contributions

Chapter 2 addresses the issue of microscope calibration, as this is a mandatory prerequisite for high accuracy measurements. 3D imaging of thick biological samples introduces distortions in the image in two different ways:

a) Focal shifts are induced by local variations of the refractive index. And b) depth-dependent spherical aberrations introduce additional errors.

A calibration scheme is suggested that allows simultaneously correcting for these two effects in a single step. The described framework is based on measuring the deformation of fluorescent focal check beads at different depth levels. In an object-constrained parametric blind deconvolution the parameters of a depth-dependent
Chapter 5. Conclusions

PSF representation are estimated. This information can be used to correct for the error in biological image stacks in post processing. In various tests, the convergence and robustness of the estimation algorithm in presence of low SNR has been confirmed. The two drawbacks of the proposed scheme are, first, that it is computationally expensive and, second, that the optical properties of the calibration sample are required to be similar to the ones of the actual biological sample.

Chapter 3 presents an algorithm for automatic information extraction from noisy fluorescence 3D microscopy data. It provides a method to estimate the number and position of fluorescent sub-resolution tags with high accuracy even for cases where they are separated by less than the optical resolution of the microscope. The main novel contributions are:

a) The rigorous extension of such a detection algorithm to 3D, where manual measurements are inapplicable or very inaccurate.

b) The suggested PSF mixture model approach for estimating the number of kernels present in a given intensity distribution.

With the presented method, isolated spots can be localized with an accuracy of a few nanometers at SNRs as low as 5-10. For interfering spots the localization accuracy decreased to 10-20nm depending on the point-to-point distance of the interfering spots. For the above stated SNR range, it is possible to resolve tags separated at half the Rayleigh distance. As conjectured in Chapter 1, the factor of super-resolution achievable, turned out to be limited mostly by the SNR of the image. The algorithmic performance and limits were determined on synthetic data before applying the framework to chromosome tracking, where distance measurements below the resolution limit are required.

Chapter 4 reports an extension of the detection algorithm to a full relative tracker. In contrast to detection, where the image is segmented into features at a single time point, the tracking algorithm includes dynamic information of the object displacement. This results in an increase in resolution and robustness of the extracted information. In addition, it provides information of the complete object trajectory and transformation of an entire movie.

The novelty in this algorithm is the idea of establishing a multi-template matching subject to splitting interfering signals.
The super-resolution is increased from a factor two when using the detection algorithm to a factor three with the tracker. The increase in robustness is more difficult to quantify, as it is even stronger SNR dependent. Approximately 80% of the tags, which were missed or only inaccurately localized by the detector, were recovered with the tracker.

5.2 Outlook

Our plan for the immediate future is to apply the framework for a large-scale analysis of yeast mitosis as described in Chapter 3. The project involves the extraction and comparison of chromosome trajectories in mitosis of budding yeast of wild-type and mutant strains. In this thesis the biological interpretation was left out and the cell data was only used to show the applicability of our approaches to real world problems. It will be interesting to now test our hypothesis that it is possible to perform protein function mapping based on trajectory analysis. In a first step an automatic trajectory classification will be pursued that allows a statistical distinction between wild-type and mutant cells without interpretation of the observed trajectories in form of the function of the deleted protein. Based on that, mechanical models of the yeast spindle will be set up and parameters will be estimated and identified by fitting them to trajectory data.

On the algorithmic aspect of the thesis, future work will be directed towards a generalization of the algorithms in terms of objects. Instead of limiting the observation to a finite super-position of diffraction-limited tags, more complex objects will be implemented. As long as there is an efficient mathematical approximation of the observed objects available the splitting and tracking can be generalized in a straightforward manner.
References


References

39 Sedgewick, R. Algorithms in C++: Graph algorithms (Addison-Wesley, 2002).
Acknowledgements

First and most of all I would like to thank my supervisor Gaudenz Danuser for providing me with the possibility to work on such a fascinating topic. After finishing my physics studies I decided to change to a more interdisciplinary field, where my physics knowledge would still be of some use. I think I have found an ideal place for that at Gaudenzes lab. His scientific intuition and profound knowledge combined with experience have resulted in an optimal support during the entire thesis. After about one year into my PhD we started a collaboration with the biology lab of Peter Sorger at the Massachusetts Institute of Technology. This resulted in a slight turn of direction of my PhD. I have to admit that the cell biological questions introduced at that time boosted my personal motivation enormously and therefore I would like to thank Peter Sorger for giving me this opportunity. Also, I had the chance to spend two weeks at his lab and get hands on experience in a field totally new to me. A big thanks to all the members of his group, I had a great time at your place. The member from his lab I would like to mention especially is Dan Rines, thank you for preparing the samples and doing the image acquisition. You were also very helpful in sending (a lot of positive) feedback, bug reports and suggestions to improve the software. Another big “thank you” goes to all the members of the LfB/BMMG who provided a very positive working atmosphere. In particular I would like to thank Claude Berney for his support in Matlab, paper writing and other scientific questions. He has become a very close friend of mine and supported me also morally during times when things were not running as smoothly as I wished. Likewise, thanks to Toru Fischbach and our former lab member Michael Slameczka, they are both great friends who provided valuable input in form of technical support and discussions. Thanks to Jonas Dorn who contributed significantly to the last part of this thesis and I wish him the best of luck with the future work on this project. Last but not least a big thanks to my family and friends outside the scientific community without whose support and patience I wouldn’t have been able to perform this work.
Curriculum Vitae

Education
1999 - 2003 PhD student at the Bio-Micro-Metrics Group of the Laboratory for Biomechanics, ETH Zurich
1998 M.Sc. in theoretical physics (Dipl. Phys.), ETH Zurich
1993 - 1998 Undergraduate studies in physics, ETH Zurich
1995 - 1996 Undergraduate studies in physics at the University College Dublin (UCD) Ireland, two semester European student exchange program
1980 - 1992 Elementary and High School in Liestal

Research and Professional Experiences
2001 - 2003 PhD project II: Algorithms for Detection and Tracking of Objects with Super-Resolution in 3D Fluorescence Microscopy
1996 Internship at SIEMENS AG: Project in the area of high-speed telecommunication, regaining the clock signal from an incoming serial data stream using phase locked loops (PLL)
1995 Semester work: Scaling and rotation independent pattern recognition using neuronal networks
List of Publications

Full Papers


Conference Papers


Papers in Review


Abstracts


2. ASCB, Washington, DC, December 8-12, 2001, Automatic Fluorescent Tag Detection in 3D with Super-Resolution: Application to the Analysis of Chromosome Movement

3. Three-Dimensional and Multi-dimensional Microscopy: Image Acquisition and Processing VIII at BIOS 2001, SPIE; San Jose, CA, January 20-26, 2001; One-step Calibration of Specimen induced Focal Shifts and Spherical Aberration for quantitative 3D Microscopy