Eur Biophys J (2011) 40:1167–1186
DOI 10.1007/s00249-011-0747-7

ORIGINAL PAPER

Automated multidimensional single molecule fluorescence microscopy feature detection and tracking

Daniel J. Rolfe · Charles I. McLachlan · Michael Hirsch · Sarah R. Needham · Christopher J. Tynan · Stephen E. D. Webb · Marisa L. Martin-Fernandez · Michael P. Hobson

Received: 18 February 2011 / Accepted: 23 August 2011 / Published online: 18 September 2011

Abstract Characterisation of multi-protein interactions in cellular networks can be achieved by optical microscopy using multidimensional single molecule fluorescence imaging. Proteins of different species, individually labelled with a single fluorophore, can be imaged as isolated spots (features) of different colour light in different channels, and their diffusive behaviour in cells directly measured through time. Challenges in data analysis have, however, thus far hindered its application in biology. A set of methods for the automated analysis of multidimensional single molecule microscopy data from cells is presented, incorporating Bayesian segmentation-based feature detection, image registration and particle tracking. Single molecules of different colours can be simultaneously detected in noisy, high background data with an arbitrary number of channels, acquired simultaneously or time-multiplexed, and then tracked through time. The resulting traces can be further analysed, for example to detect intensity steps, count discrete intensity levels, measure fluorescence resonance energy transfer (FRET) or changes in polarisation. Examples are shown illustrating the use of the algorithms in investigations of the epidermal growth factor receptor (EGFR) signalling network, a key target for cancer therapeutics, and with simulated data.

Keywords Single molecule microscopy · Fluorescence · Data analysis · Bayesian · FRET

D. J. Rolfe · M. Hirsch · S. R. Needham · C. J. Tynan · S. E. D. Webb · M. L. Martin-Fernandez (✉)
Central Laser Facility, Science and Technology Facilities Council, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot, Oxon OX11 0FA, UK
e-mail: marisa.martin-fernandez@stfc.ac.uk

C. I. McLachlan
Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK

M. P. Hobson
Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK

Introduction

Understanding how signal inputs and outputs are organised in functional protein networks is a major goal in biology and can ultimately contribute to the prediction of the outcome of therapeutic drugs. By following the spatio-temporal dynamics of individual molecules in cells, single molecule fluorescence imaging can reveal quantitative information on behaviours of interacting proteins species that would otherwise be masked by the cell ensemble. This includes combinatorial inter-protein interactions, unsynchronised changes in protein conformation and organisation (see for example Weiss 1999; Kapanidis et al. 2006; Leake et al. 2006; Teramura et al. 2006; Das et al. 2007; Zhang et al. 2009; Schaaf et al 2009; Hern et al. 2010), and the positive and negative feedback loops found during cell signalling, for example in the epidermal growth factor receptor (EGFR) family (reviewed in Citri and Yarden 2006). To derive this information, single molecule data must be collected from cells over time in a number of colour channels to follow molecular movements and interactions (see for example Roy et al. 2009; Fili et al. 2010), and over different polarisations to determine 3D molecular orientations, using single molecule fluorescence energy transfer (FRET) and polarisation imaging (Webb et al. 2006, 2008).

The analysis of single molecule cell data can be challenging. Signals emitted by individual fluorophores are

inherently weak and are superimposed on non-trivial backgrounds consisting of out-of-focus spots mixed with non-uniform intracellular cell autofluorescence, which can be reduced but not completely eliminated by total internal reflection (TIR) excitation. For each excitation laser, TIR creates an evanescent field on the glass coverslip to which the cells adhere; these fields reduce exponentially with depth, penetrating <300 nm axially into the sample. The limited depth of the evanescent field is essential to reduce autofluorescence to a level where the detection of individual fluorophores at the basal membrane of cells is possible. However, a side effect of TIR is that fluorophores located at varied axial distances in the cell membrane experience different excitation field strengths. The fluorescence from single molecules can therefore be much fainter and have a wider spread of signal-to-noise ratios than biomolecules immobilised on a 2D glass surface exposed to the same incident intensity.

Significant effort has been applied in the last few years to the development of single molecule analysis methods suitable for cell studies (Selvin and Ha 2007; Yanagida and Ishii 2009; Bräuchle et al. 2010, and references therein). The general approach is the detection and measurement of features in each image frame and the inferring of information from the changes in these features through time. Exploiting the temporal information is often achieved by tracking detected features from frame to frame and interpreting the resulting position and intensity traces versus time. In some cases though, information on protein diffusive behaviour has been derived by measuring correlation properties without first calculating temporal tracks, for example for the membrane-associated GTPase H-Ras (Semrau and Schmidt 2007). Mashanov and Molloy (2007) followed the movements of plasma membrane-bound pleckstrin homology (PH) domains and intact PLC-δ1 in live cells by separately considering the cases of moving and non-moving proteins, and performing simple tracking in the latter case. Sergé et al. (2008) implemented a likelihood-based feature detection algorithm to track quantum dot labelled transmembrane EGFR molecules in COS-7 cells. Recent advanced work concentrating on the feature tracking problem includes Jaqaman et al. (2008) and Yoon et al. (2008). Yoon et al. (2008) implement a sequential Monte Carlo Bayesian approach to tracking that can exploit both spatial and temporal prior information about the trajectories, is less sensitive than other techniques to control parameters, but that is very computationally expensive. Dunne et al. (2009) demonstrate this Bayesian approach in investigations of the diffusion of tandem fluorescence protein fusions of CD86, dimer CD28 and the T-cell receptor in Jurkat T cells.

Here, however, we present algorithms that enable single molecule data analysis of images that are both multidimensional and from cells. They have been developed in order to detect and simultaneously track any number of fluorophore-tagged protein species in a multiprotein signalling network. The single molecule microscope (Webb et al. 2006, 2008) uses image splitting optics to create multiple copies of the same field of view on the camera that differ only in their spectral range and/or polarisation. Registration of these copies matches the fluorophore positions in each spectral/polarisation channel and accounts for image aberrations. In contrast to the other data analysis schemes discussed above, our emphasis is on providing a complete solution for multicolour single molecule tracking and combined single molecule FRET and polarisation studies. In addition, our algorithms use Bayesian evidence-based feature detection, which is more robust for model selection than likelihood-based feature detection.

Description of the algorithms

Feature detection using Bayesian segmentation

Single channel feature detection

The reliability of feature detection in single molecule cell data depends on the ability to discriminate the images of individual molecules from noise and cellular background. Bayesian analysis is well suited to this task because it provides a formal approach to making statistical inferences from observations based on explicit assumptions and prior knowledge (e.g. physical models for those observations and known constraints on the model parameters). Bayes's theorem relates the *posterior probability* $\Pr(M|d, K)$ that a particular hypothesis (or model) M is true given some data d and background knowledge K to the probability $\Pr(d|M, K)$ that the data would have been measured if the hypothesis were true, given the same background information. $\Pr(d|M, k)$ is called the *likelihood* and is generally a lot easier to specify directly than the posterior. See "Appendix 1" for further explanation.

Features are detected in each image by comparing two alternative models for the image region (region of interest, ROI) around each pixel. One model, denoted H_0, is the null hypothesis that the image region around a pixel is described by pure background emission and stochastic noise. The alternative model, H_1, is that the image region around a pixel is described by a feature profile from a single point emitter centred in the middle of the pixel, plus background emission and stochastic noise. By specifying explicitly the assumed form of the background emission in each pixel, the feature profile and the probability distribution of the stochastic noise, Bayes's theorem can be used to

objectively compare the probabilities of the H_0 and H_1 models at each pixel. An *evidence map* of the relative probability of these models in each pixel is calculated, considering all possible model parameter values. This map is thresholded; pixels where model H_1 is the better explanation of the data are assumed to be the locations of features, a process known as *segmentation*.

For this work a Gaussian profile with fixed (and known) width is adopted as an approximation for the point spread function (PSF) of the microscope. While the true theoretical profile should be an Airy disc, approximating this with a Gaussian is common practice in this field (e.g. Sergé et al. 2008; Mashanov and Molloy 2007; Cronin et al. 2009; Mortensen et al. 2010). A Gaussian approximation is significantly easier to use and computationally faster to evaluate. We demonstrate in "Appendix 1" and Fig. 11 that whether the true profile is a Gaussian or Airy disc makes no significant difference to the reliability of this feature detection. If images of fluorescent beads are obtained, the best fit Gaussian bead profile can be used for feature detection. By default, the ROI considered around each pixel is a square of side approximately four times the assumed feature profile full width at half maximum intensity (FWHM).

The local background over the image region around each pixel is assumed to be spatially uniform, as in Hobson et al. (2009), and this correlates well with many observations. We therefore have model H_0 with parameter B_0 (background level, $0 \leq B_0 \leq B_{\max}$) and model H_1 with parameters B_1 (background, $0 \leq B_1 \leq B_{\max}$) and I (feature intensity, $0 \leq I \leq I_{\max}$). $I_{\max}$ and $B_{\max}$ are set to values based on the maximum and standard deviation of the whole image intensities.

The most probable parameters for each model, including the feature intensity, $\hat{I}$, are easily calculated using Bayesian parameter estimation. Thus once the locations of features have been determined, their most likely intensities are also given. This method provides an objective way of detecting features and works well in situations with noisy data of interest here. Root mean square uncertainties in the model parameters, including the error in feature intensity, $\sigma(\hat{I})$, and background $\sigma(\hat{B}_1)$, can also be estimated.

This Bayesian evidence-based feature detection algorithm (Hobson et al. 2009, originally based on Savage and Oliver 2007) is detailed in "Appendix 1". This method is closely related to the likelihood ratio feature detection used in Sergé et al. (2008), both methods having an objective probabilistic test for the presence of a feature that takes account of the noise and model feature profile, rather than simply thresholding the intensity. However the Bayesian evidence ratio test is a more robust method for model selection, since it accounts for the differing number of parameters in each model and those parameters' distributions, whereas the likelihood ratio test compares only the likelihood of the most likely parameters for each model.

In images with crowded fields, i.e. in which feature separation is comparable to one PSF, for example in cells overexpressing the protein of interest, and/or with significant background intensity variations on spatial scales similar to one PSF, e.g. in some tissues samples, the calculations will be incorrect. The detection and measurement limits depend on the feature size, signal-to-noise ratio and the size of the sub-image around a pixel used to determine the evidence ratio and feature parameters. In some cases overlapping/close together features are still individually detected by the algorithms, but their positions and intensities are distorted. Significant background variation invalidates the assumption that it is smooth on the spatial scale of the PSF. Various options are being explored to improve handling of both of these issues. One approach to the crowding problem is discussed briefly in the "Iterative improvement of detection for overlapping features" section.

Sub-pixel refinement

The feature detection method described provides locations specified to the nearest mid-pixel location. Spatial resolution close to the molecular scale can also be achieved from single molecule data, which is the basis for several super-resolution methods of biological importance, such as photoactivatable light microscopy (PALM, Betzig et al. 2006) and stochastic optical reconstruction microscopy (STORM, Rust et al. 2006). To assess the most likely feature position to higher precision, two methods have been implemented.

1. A recursive algorithm is used to increase the resolution of the evidence map around the feature location as follows. Once a feature has been detected in a pixel, a more detailed "zoomed-in" evidence map is calculated on a 3×3 grid within this pixel, with the peak of this map being adopted as the new best feature location. This "zooming in" process is repeated until a desired precision is reached, by default 0.01 image pixels, which represents ≈ 1.6 nanometers (depending on the wavelength of the emission). This precision is the tolerance limit of the of the iteration rather than the localisation precision of the feature.
2. A Gaussian profile is fitted to the image region centred on the initial location. The posterior probability of the intensity and background levels, position and feature size given the image data, $\Pr(I, B, X, Y, \sigma_t|\mathbf{d}, H_1)$, is used as the objective function (i.e. the function we wish to optimise) and maximised using the Downhill

Simplex algorithm (Press et al. 1992). The feature size can be fitted or remain fixed at the value assumed for the feature detection (typically equal to the PSF for each given laser wavelength). This is the same approach as implemented by others, e.g. Sergé et al. (2008). The precision of the measurement of the feature properties is limited only by the noise in the image and any difference between the assumed model and the real system.

Estimation of image noise

The stochastic noise in each image pixel is approximated with a Gaussian distribution. The variance, σ_{ij}^2, of this distribution in pixel (i, j) is estimated as the mean square of the values in a rectangular region around pixel (i, j) in a "noise image". This noise image is calculated by smoothing the image (with a normalised square box smoothing kernel size 20 pixels), then subtracting this smoothed image from the original image. This is a good approximation in smooth regions of the image, but will make errors in rapidly changing regions of the image. Sergé et al. (2008) also make the assumption of Gaussian noise, but use an alternative method to assess σ_{ij}^2.

In reality the true noise will be more complicated, including components due to Poisson counting statistics and specific detector properties, e.g. quantum efficiency, which vary between different detector types and individual detectors. A physical noise model, under development, will provide a robust replacement for the Gaussian approximation, allowing the benefits of a Bayesian technique to be fully realised. Work such as that of Mortensen et al. (2010) shows the sort of benefits to be gained from better modelling of the noise likelihood.

Nevertheless, for feature detection the current approximation works quite well (see examples in "Examples" section and tests in "Appendix 1"), particularly given that the evidence threshold can be varied if necessary; however it does mean the estimates of the errors in feature intensity $\sigma(\hat{I}_1)$ and background intensity $\sigma(\hat{B}_1)$ are only approximate.

Iterative improvement of detection for overlapping features

Where nearby features are detected but interfere with each other (resulting in the calculation of erroneous intensities and positions), it is possible to iteratively refine the features. Each pass of this process involves cycling through the detected features one by one, subtracting the current best fit profile for all other features from the real image, so that only the image of the current feature plus the residuals of other features after subtraction are in the image, then refining the current feature against this image using the first method in the "Sub-pixel refinement" section, thus improving the estimate for that feature. (This process does not help in situations where overlapping features are too close to be separately detected.) Hobson and McLachlan (2003) and Sergé et al. (2008) implemented related approaches to improving the handling of overlapping features, by repeating the feature detection on the residual image until no new features were detected (also referred to as deflation (Sergé et al. 2008)). This allows the detection of faint features that were initially hidden by nearby features, and in the case of Hobson and McLachlan (2003), simplifies and speeds up an intensive Bayesian multi-object detection algorithm. However, it does not allow correction of mischaracterised features since it never refits the parameters of existing features. Both approaches therefore help with the challenges of crowded fields, but do not solve them completely. Holden et al. (2011) recently demonstrated the use of astronomical source detection software for the analysis of crowded fields in STORM images, combining the iterative refitting and deflation approaches. Testing this software for our purposes and/or making similar extensions to the algorithms described in this work will be performed in the near future.

Single molecule data analysis in multiple channels

So far the methods have been described for feature detection in a single image channel. Multidimensional data can also be acquired, in which multiple copies of the same field of view, which differ in spectral and/or polarisation content, are imaged. These channels are spread across one or more images, using image-splitter optics, acquired on one or more detectors, and may not be simultaneously imaged (Webb et al. 2008).

Two extensions are needed to enable analysis of multichannel data. The first is the determination of the transformations required to align channels. The second is the extension of feature detection and measurement to handle the transformations and channel information.

The only restriction on the channel arrangement imposed by the multichannel methods described in this section is that the images in each of the channels have approximately the same rectangular size. A particular channel is always imaged with the same detector at the same location in a single data set, but multiple channels may be present at different locations within each image. Multiple images may be taken with different sets of channels, either with the same detector at different times (time-multiplexing) or with different detectors at the same time. In order to handle the multiple different possibilities of experimental setup, a channel description model has been developed that specifies:

- which channels are in which images,
- and for each channel
  - the properties of the illumination (e.g. polarisation),
  - the location in the image,
  - the target fluorophore,
  - the components in the optical path,
  - and the detector used.

Correctly specifying this information allows automatic analysis of data sets for a variety of different experimental configurations.

Registration transformation

The first step in multichannel analysis is determination of the registration transformations. When registering channel images for ensemble fluorescence work, images of a micron-scale grid are typically used. These provide insufficient precision for single molecule images, and so images of a fluorescent bead smaller than the microscope point-spread function are used instead. Either a single bead can be scanned across the field of view using a piezoelectric microscope stage (Selvin and Ha 2007) or, as here, a single well-separated sample of beads can be used. Channel registration also addresses image aberrations due to differing optical components between channels. One channel is adopted as a reference channel with which all other channels must be aligned (registered). This is achieved by

1. imaging, in all channels, one or more areas of a sample of fluorescent beads distributed throughout the field of view (Fig. 1a, b),
2. using the feature detection algorithms described above to obtain sub-pixel position measurements for each bead in each channel,
3. using the Scott and Longuet-Higgins algorithm (Scott and Longuet-Higgins 1991) to identify corresponding beads in each channel,
4. determining the optimal affine transformation to align the beads with those in the reference channel consisting of
   - a translation, $[D_x, D_y]$,
   - a scaling by factors S_1 and S_2 in two orthogonal directions at orientation, θ_s to the image axes,
   - a rotation through angle θ.

Figure 1 illustrates several aspects of this process.

The form of the affine transform used to register positions has six parameters for each transformation, $\mathbf{T} = [D_x, D_y, S_1, S_2, \theta, \theta_S]$. If a position in the non-reference channel is represented by the vector $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ and the corresponding point in the reference channel is $\mathbf{r}_{\text{ref}} = \begin{pmatrix} x_{\text{ref}} \\ y_{\text{ref}} \end{pmatrix}$ then the transform function $\mathbf{A}$ is defined by $\mathbf{r}_{\text{ref}} = \mathbf{A}(\mathbf{r}, \mathbf{T})$.

Upper limits on the alignment error can be estimated using the residual separation of registered and reference channel bead positions.

See "Appendix 2" for more details of transform determination.

Multichannel feature detection

Feature detection in multichannel data sets is performed by creating a single evidence map of feature locations that combines information from all simultaneously imaged channels. This is done by calculating the evidence map and the optimal model (H_0 or H_1) parameters in each channel at all points equivalent to the mid-pixel positions in the reference channel. The inverse registration transformation $\bar{\mathbf{A}}(\mathbf{r}, \mathbf{T})$ (i.e. $\bar{\mathbf{A}}(\mathbf{A}(\mathbf{r}, \mathbf{T}), \mathbf{T}) = \mathbf{r}$) is used to determine the point positions in the non-reference channels. The feature profile shape used to calculate the ratio of the posterior probabilities for models H_0 and H_1 is scaled when transforming to each non-reference channel. This anisotropic scaling, $\bar{\mathbf{S}} = \begin{pmatrix} 1/S_1 \\ 1/S_2 \end{pmatrix}$ (from the registration transform $\mathbf{T}$), which may not be aligned with image axes, is approximated by an isotropic scaling factor $\bar{S} = 1/\sqrt{S_1 S_2}$. Since the anisotropy is always small (S_1, S_2 are both typically within a few percent of each other), this is a suitable approximation that ensures the transformed profile has the correct area, thus giving the correct intensities for features measured in each channel. It is important to calculate the evidence maps and feature properties using the untransformed images, transforming the feature model as described here, rather than by transforming the images. This is because the affine transformation mixes contributions from different pixels, introducing statistical inhomogeneities and correlations in the noise properties of transformed images and hence artefacts and spurious detections (illustrated in Fig. 2). This is particularly important in data where noise contributions are significant, such as in single molecule microscopy.

Having calculated aligned evidence and parameter maps for each channel, it is possible to apply the feature detection algorithms already described separately to each channel. For co-localisation experiments, where features are tracked separately in different colours and then compared, this is the right approach. When doing experiments where the intensities of features in different channels are interdependent, for example FRET experiments, this is not necessarily the best approach. It is possible to compare feature positions between channels and identify those aligned to within a specified tolerance to be the same feature, though this adds

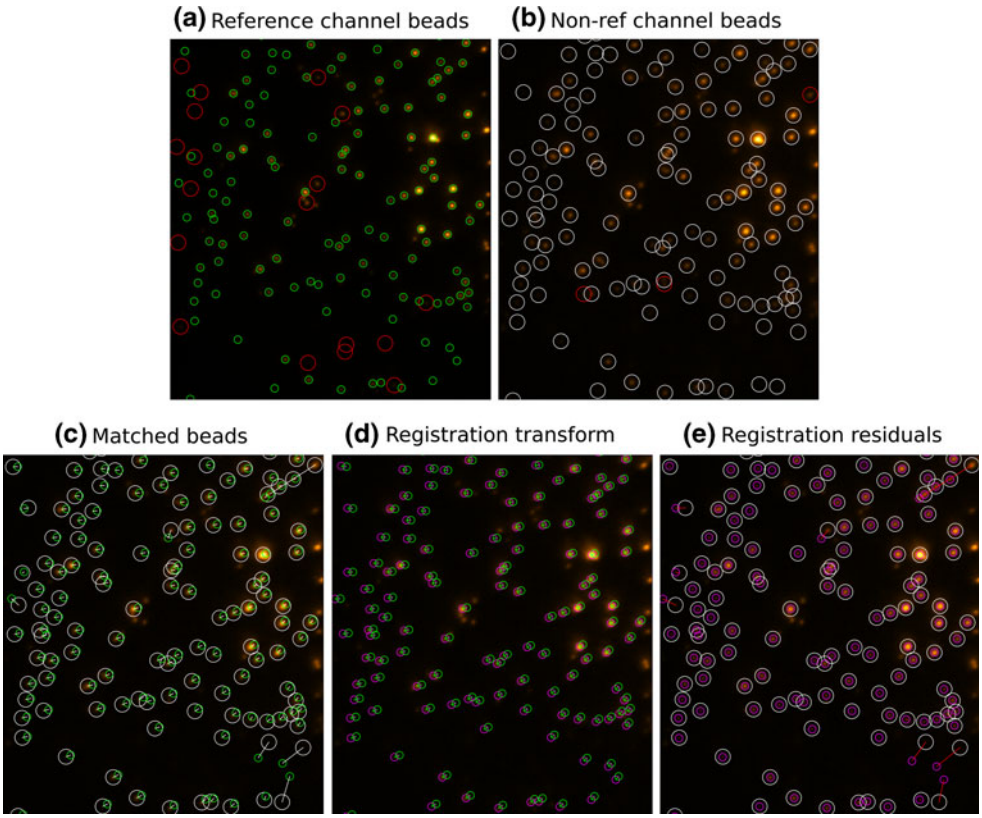

Fig. 1 Registration using beads. **a** and **b** show corresponding unregistered images of beads in the reference channel and a non-reference channel. *Circles* show detected bead locations in each image. *Green/white circles* show bead locations for which the Scott and Longuet-Higgins algorithm identified matching pairs in the reference and non-reference channels, while *red circles* show bead locations for which counterparts were not detected. **c** shows the non-reference bead image with detected bead locations from **a** and **b**, joining with *white lines* the matched pairs. Note the incorrect matches (*long white lines*). In **d** the reference channel measured bead locations (for matched beads only) have been plotted (*green*) on the non-reference image, and the same reference locations have been transformed using the determined registration parameters and plotted in *purple*, thus showing the registration transformation. **e** shows the transformed reference bead locations and the detected non-reference locations, with *red lines* joining the centres of each matched pair. The *red lines* are therefore the residuals of the registration transformation

an additional analysis step and its associated control parameter. Additionally, carrying out feature detection separately in each channel does not allow easy exploitation of prior knowledge about the multichannel properties of feature emission. For example, if it is known that the molecules should be visible in a particular set of channels, then the presence of signal at the same location in multiple channels can be exploited to improve detection efficiency.

The approach used so far in this work has been to exploit channel information at the feature detection stage because it has mostly involved multichannel characterisation of individual features. At a given position in the sample we have a choice between the 'null' and 'feature' models (H_0 and H_1 described before) in each channel. We can identify the most likely multichannel model as the combination of the most likely models in each channel. For each channel we choose model H_1 if the evidence ratio is greater than 1, otherwise H_0. The product of the evidences for the most likely models in each channel is therefore the evidence of the most likely model, $\prod_{\text{all channels}} \Pr(\mathbf{d}|H_{\mathbf{d}\,\text{most likely}})$ (where $H_{\mathbf{d}\,\text{most likely}}$ is the most likely model (H_0 or H_1) in a particular channel). The multichannel null hypothesis is that there is no feature in any channel, the evidence for which is the product of the evidences for H_0 in each channel, $\prod_{\text{all channels}} \Pr(\mathbf{d}|H_0)$. The multichannel evidence ratio comparing the evidence of the most likely model with the evidence of the null hypothesis is thus

$$
\begin{aligned}
E &= \prod_{\text{all channels}} \frac{\Pr(\mathbf{d}|H_{\mathbf{d}\,\text{most likely}})}{\Pr(\mathbf{d}|H_0)} \\
&= \prod_{H_0\,\text{most likely}} \frac{\Pr(\mathbf{d}|H_0)}{\Pr(\mathbf{d}|H_0)} \prod_{H_1\,\text{most likely}} \frac{\Pr(\mathbf{d}|H_1)}{\Pr(\mathbf{d}|H_0)} \\
&= \prod_{H_1\,\text{most likely}} \frac{\Pr(\mathbf{d}|H_1)}{\Pr(\mathbf{d}|H_0)}
\end{aligned}
$$

Ideally the formulation of a combined posterior function is desirable, but the difficulty in forming a sensible prior in

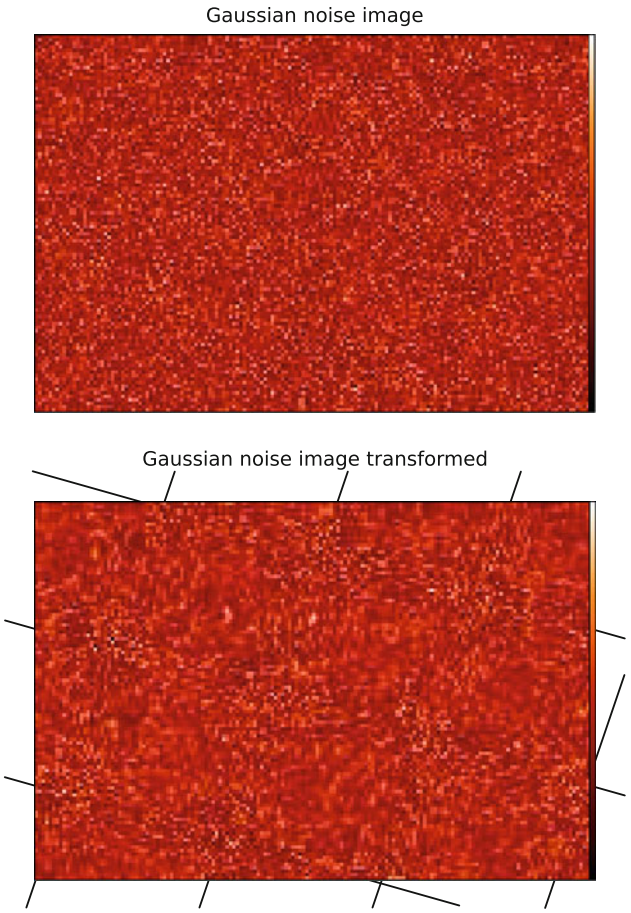

Fig. 2 Image transformation artefacts. *Upper panel* shows an image consisting purely of statistically homogeneous, uncorrelated Gaussian noise. *Lower panel* shows an image consisting purely of statistically homogeneous, uncorrelated Gaussian noise, which is then transformed according to a typical registration transformation. The extent of pixel mixing varies periodically across the image, with areas of minimum pixel mixing along the indicated *grid lines*, and regions between these lines looking blurred. These blurred regions are where the transformed and untransformed pixels are least aligned, leading to maximum pixel mixing and highly correlated noise in the transformed image

each channel is further complicated by the many possibilities that depend on the experimental details. For example, how many molecules do we expect to be present? How many will be visible at each time point? In all channels? In the case of donor/acceptor channels in a FRET experiment, the channel intensities may be correlated, anti-correlated or unrelated. The multichannel evidence ratio E expression adopted here combines multichannel information without requiring further prior information. An alternative possibility would be to compare the total evidence for all possible non-null hypotheses (i.e. all combinations for which at least one channel has H_1 more likely) to the multichannel null hypothesis.

Feature detection is then performed as in the single channel case for E. E is the product of evidence ratios >1, and the threshold used for E will always be $\gg 1$; therefore each channel that contributes to E increases the likelihood of a feature being detected. Cases where there may be no detection in a single channel can now combine to increase the chance of detection in the combined evidence. At locations where a feature is detected in E, the existence of a multichannel feature is inferred and the intensity in each channel is $\hat{I}$. A multichannel feature therefore has a location, and an intensity vector containing an intensity for each channel, $\mathbf{I} = [\hat{I}_{\mathrm{ch1}}, \hat{I}_{\mathrm{ch2}} \ldots]$. We also perform sub-pixel refinement of the detected features as in the single channel case by recursively zooming in on the multichannel evidence map.

Registration of images

As discussed in the previous section, transformation of images to align them is not carried out for analysis. They are, however, transformed for visualisation purposes only. This can be performed using either a simply bi-linear interpolation or by recursively subdividing pixels and re-binning image counts to improve registration precision and image count conservation.

Tracking

Tracking is the process of following molecules through time. It enables single molecule data to be analysed in live cells, where the molecules move, both to measure molecular photophysics over many frames and to measure diffusion rates. It is performed as follows:

- A track is seeded from each feature in the first frame.
- For each subsequent frame at time t in turn
  1. A connection probability, $P(\mathit{feature}_i|\mathit{track}_j)$, is calculated for each feature i in the frame and each existing track j. $P(\mathit{feature}_i|\mathit{track}_j)$ is the likelihood of measuring feature i as part of track j.
  2. This set of probabilities is used to determine which features to assign to existing tracks, with at most one feature per track and one track per feature.
  3. Features not assigned to existing tracks seed new tracks starting in that frame.

We assume that features are more likely to be linked to the tracks closest in spatial and temporal proximity and cannot be linked to features separated by more than l and τ in space and time respectively. l and τ are control parameters that must be tuned to achieve acceptable results for particular data sets. Thus the connection probability presently used is a simple combination of temporal and spatial Gaussian distributions with cutoffs, $P(\mathrm{feature}_i|\mathrm{track}_j) \propto$

$\exp\left(-\frac{\Delta r^2}{l^2}\right)\exp\left(-\frac{\Delta t^2}{\tau^2}\right)$ if $\Delta r < l$ and $\Delta t < \tau$, $P(\text{feature}_i|\text{track}_j) = 0$ otherwise, where $\Delta r = |\mathbf{r}_{\text{feature } i} - \mathbf{r}_{\text{last in track } j}|$ and $\Delta t = t - t_{\text{last in track } j}$. $\mathbf{r}_{\text{feature } i}$ is the position of feature i. $\mathbf{r}_{\text{last in track } j}$ and $t_{\text{last in track } j}$ are the position and time of the most recent feature in track j. This simple formulation allows for tracks that briefly disappear because of the single molecule fluorescence phenomenon of blinking to be picked up again when they reappear within time τ (and within separation l), but it takes no account of how the expected spatial proximity l might vary with t. More physically meaningful forms for the probability could be used, but these will depend on the nature of the motion of individual tracks.

The simple formalism with $P(\text{feature}_i|\text{track}_j)$ allows more prior information and understanding of track properties and fluorophore/molecule behaviours to be easily incorporated into the tracking, e.g. the likelihood of a particular intensity for the next feature in a track.

Linking of the features in frame t with existing tracks can be carried out using one of two algorithms.

1. For simple *greedy* pairing, the feature and track with the highest (non-zero) $P(\text{feature}_i|\text{track}_j)$ are considered linked, so feature i is appended to track j and both are no longer considered for linking in frame t. The next most likely track and feature are then linked, and this continues until there are no further non-zero likelihood links. This method is very quick, but has the disadvantage that a single good linkage could preclude multiple links with only slightly inferior probabilities.
2. $P(\text{feature}_i|\text{track}_j)$ is effectively a proximity matrix, and so the Scott and Longuet-Higgins algorithm (Scott and Longuet-Higgins 1991) previously discussed in the "Registration transformation" section can be used to find an optimal set of single matches. This is more computationally expensive, but is less likely to sacrifice many good matches for one excellent one. To speed up this method, the highest $P(\text{feature}_i|\text{track}_j)$, denoted by $P_{\max}(\text{track}_j)$, is calculated for each track, and only the $2n_{\text{features}}$ tracks with the highest values of $P_{\max}(\text{track}_j)$ are considered for linking in frame t, where n_{features} is the number of features detected in the frame. This can make a big difference to speed when tracking in frames when there can be many more existing tracks than features. For efficiency, tracks whose most recent time point has $\Delta t > \tau$ are excluded from consideration for linking with more features.

The result of tracking is a set of tracks, each of which consists of a time series of detected molecules, but in which some frames may be missing (due to blinking or non-detected features), and where each molecule has one or more intensities at each time point depending on the number of channels present at each time point.

Gaps in tracks are then closed by performing a simple linear position interpolation between the positions before the gap and after the gap onto each intermediate frame. At each interpolated position, the channel intensities for model H_1 are used as the intensities for that feature. Negative intensities are not precluded at this stage (ignoring the prior on intensity). The assumption of linear interpolation in position can be inappropriate, but for fixed or slowly moving molecules it enables intensity traces to include an estimate of the intensity when features are not detected. This is useful particularly for plotting intensity traces of blinking molecules, where the feature detection (correctly) does not find a feature during blinking (Fig. 3). Intensity traces can be extrapolated before the first and after the last detected features, using the positions at those time points, in order to illustrate the switching on of fluorescence and final photo-bleaching step (Fig. 3). Interpolated/extrapolated time points are clearly marked so that they may be included or excluded as appropriate for interpretation and further analysis.

This tracking method described is simple compared to the most advanced algorithms now available (e.g. Jaqaman et al. 2008; Yoon et al. 2008), and its reliability depends on tuning parameters l and τ for a particular problem, and on the feature density, motion properties and frame rate of the data. We are currently developing and testing a significantly more advanced tracking solution, to be reported elsewhere, which will be incorporated into this suite. However, the current solution has proved sufficient for some tracking problems, e.g. tracking EGFR molecules in live cells (see "EGFR tracking" section).

Track analysis and statistics

The steps to be followed after feature detection and tracking depend on the experiment being performed. Each track is analysed to determine various properties and statistics that can be used to filter and sort them, and help with

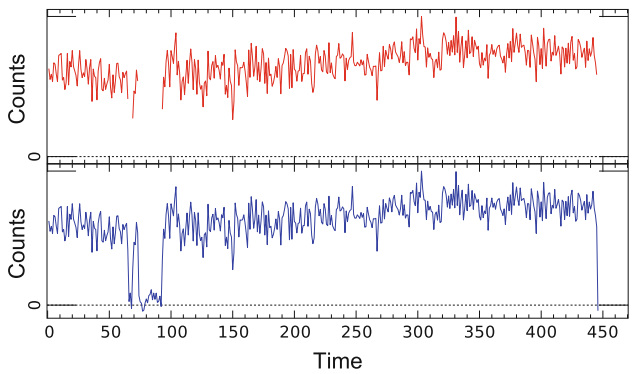

Fig. 3 A single molecule trace excluding (*red*) or including (*blue*) interpolated and extrapolated points. The *blue curve* shows the photo-bleaching and blinking more intuitively

their interpretation. Examples of properties include the maximum acceptor intensity, in a FRET experiment, and the number of intensity levels, for identifying the number of molecules in a track—some quantities are calculated only for particular experimental configurations.

For each channel in each track, intensity quantities calculated include:

- minimum, $I_{\min}$
- maximum, $I_{\max}$
- mean, I_{mean}
- standard deviation, $\sigma(I)$
- mean local standard deviation and $\sigma_{\text{mean local}}(I)$. This is the average of local standard deviations calculated at each time point, including five frames either side, and is a quick rough estimate of the noise standard deviation.
- $\sigma(I)/\sigma_{\text{mean local}}(I)$, a measure of the significance of any intensity variations. A value of 1 suggests there are no real intensity variations, only noise, while greater values correspond to variations in addition to random noise. For multichannel data sets the maximum value of $\sigma(I)/\sigma_{\text{mean local}}(I)$ from all channels is calculated.

For multichannel data sets additional quantities include:

- sample correlation coefficient between the intensity in each pair of channels (to see how similar the intensity trace is in a pair of channels)

$$C_{A,B} = \frac{1}{(N_{\text{frames}} - 1)\sigma(I_A)\sigma(I_B)} \times \sum_{i=1}^{N_{\text{frames}}} (I_{A,i} - \langle I_A \rangle)(I_{B,i} - \langle I_B \rangle).$$

For FRET data sets, the total intensity traces for donor and acceptor channels are calculated (FRET donor and acceptor emission may each be split into different polarisations to monitor their relative orientation during the FRET process). For data sets multiplexed in time, i.e. where an alternating subset of the channels is present at each time point, intensity traces in each channel are first interpolated (linearly) onto a common time series before adding them. From the total donor and acceptor intensity traces, quantities calculated are:

- mean and maximum donor and acceptor intensities
- sample correlation coefficient between donor and acceptor intensity. Strong correlations/anti-correlations between donor and acceptor intensity are indicators of FRET.

For both single and multichannel channel data sets, changes in the multichannel intensity vector **I** (see the "Multichannel feature detection" section) and temporal regions of constant intensity vector are identified. A change in intensity in any channel implies a change in the intensity vector. We define a step as a significant change in intensity vector between a pair of successive time points, and temporal regions between steps are vector intensity levels. In single channel data sets a vector intensity level is therefore just a constant intensity level. A step represents a sudden change in molecular behaviour, such as rotation or bleaching. The number of intensity levels in a single channel trace may be used to count the number of molecules within a fluorescent spot. We use the term "level" to refer to vector intensity levels for the rest of this discussion.

In a general case, if quantities x_1 and x_2 are drawn from normal distributions with the same mean and with variances $\sigma^2(x_1)$ and $\sigma^2(x_2)$ respectively, the probability that $|x_1 - x_2|$ is less than some value Δx ($\Delta x > 0$) is

$$P_{\text{diff}}(|x_1 - x_2| < \Delta x) = \text{erf}\left(\Delta x / \sqrt{2(\sigma^2(x_1) + \sigma^2(x_2))}\right), \quad (1)$$

where erf is the *error function*. So P_{diff} is the probability that a particular difference Δx is too high to be explained by chance, i.e. it is the probability that the means are different.

Steps and levels are are identified as follows:

- Equation 1 is used to determine whether the difference between successive intensities I_1 and I_2 in a particular channel is significant. The intensity errors from feature detection, $\sigma^2(I_1)$ and $\sigma^2(I_2)$, are used for the variances. Intensity changes are considered to be significant if $P_{\text{diff}} \geq 0.95$. A *step* (intensity vector change) occurs if the intensity change in at least one channel meets the criterion.
- The initial set of intensity levels consists of each temporal region of the trace separated by (but excluding) the identified steps.
- These levels are then compared using Eq. 1 to see if any are the same, using the intensity means and standard deviations of each level. If the most similar pair of levels are the same ($P_{\text{diff}} \leq 0.95$), they are linked, and the mean and standard deviation of the intensities comprising the combined level are calculated. For multichannel data sets, the similarity criterion must be met separately in all channels for levels to be linked. This merging process is repeated (including already combined levels) until no pair of levels meets the similarity criterion.
- Finally the levels are "cleaned" by removing insignificant steps. Any steps that separate two consecutive regions in the same level are checked for significance following Chauvenet's criterion. If neither intensity in the step in any channel is a significant outlier of the intensity distribution of that channel and level, then the points in the step are merged into the level.

An example of level identification is shown in Fig. 4. The thresholds used for this simple step/level analysis were determined by trial and error for a variety of data sets. This step and level identification works well for traces that are

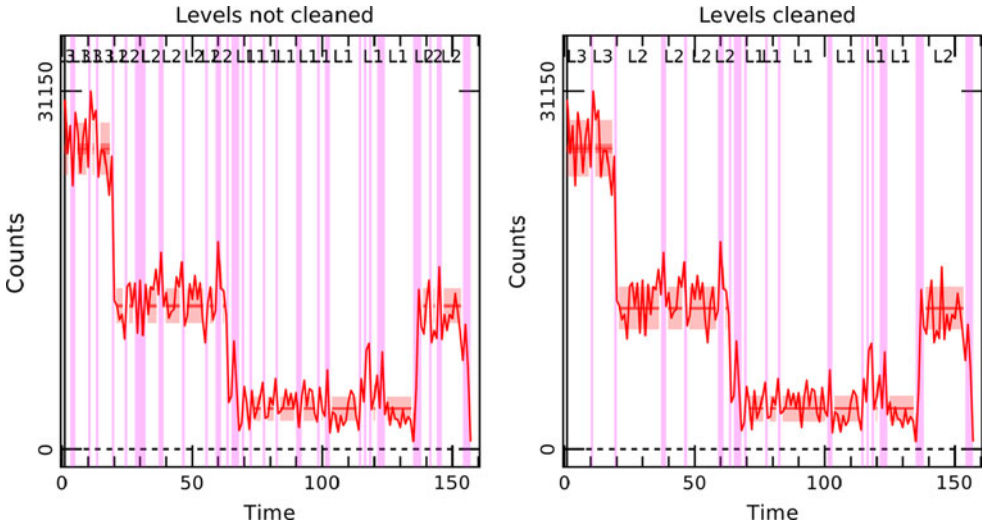

Fig. 4 Step detection and level identification in an intensity trace from a DNA ruler sample labelled with Atto647N dyes. *Horizontal bars* and *lines* show the intensity mean ± standard deviation and time interval of the three identified levels; *vertical bars* show the regions identified as steps. Insignificant steps have been cleaned in the *right panel*

well described by multiple constant intensity vector regions. Additional statistics make it very useful for quickly identifying tracks that clearly show the desired number of constant intensity vectors, which include

- The number of levels, N_{levels}.
- For tracks with at least two levels (including any extrapolated/interpolated points giving a zero intensity level), the fraction of the total track duration in an identified level (i.e. not in steps), f_{levels}. For other tracks, f_{levels} is set to 0.
- The reduced χ^2 of the determined levels, a goodness-of-fit statistic for how well the identified levels match the data at each time point.

$$\chi^2_{\text{levels}} = \frac{1}{\left(\sum_{\text{levels,frames,chans}} 1\right) - 3N_{\text{levels}}} \times \sum_{\text{levels,frames,chans}} \frac{\left(I_{\text{frame,chan}} - \langle I_{\text{level,frame,chan}} \rangle\right)^2}{\sigma^2(I_{\text{frame,chan}})}.$$

where the sums are over all channels in all frames that are part of a level. $3N_{\text{levels}}$ is the number of degrees of freedom in the model set of levels, which is 3 for each level—the intensity, start time and end time. If the uncertainties in intensity at each time point, $\sigma(I_{\text{frame}})$, are correct and the trace is well described by the intensity levels in the non-step regions, χ^2_{levels} should be about 1, increasing for worse fits.

It should be noted that discrete intensity changes, as focussed on here, are not the only fluorescence changes that may be of significance. The emphasis on such steps here is a result of the type of problems the authors have been addressing, involving looking for photobleaching and blinking steps. In many cases more gradual changes may be of interest, for which additional analyses and heuristics may need to be developed.

Tracks can be filtered according to basic equalities/comparisons on the various track statistics and sorted according to any one statistic. This enables rapid identification of tracks of interest for particular problems. A useful method for finding good multi-step traces is the use of a filter of the form

$$\text{min levels required} < N_{\text{levels}} < \text{max levels required AND}$$
$$f_{\text{levels}} > 0.5 \text{ AND } \chi^2_{\text{levels}} < 2$$
(the exact thresholds sometimes need to be changed).

It is important when using such filtering/selection techniques to be aware of the limits of these methods. They are only as good as the the algorithms to generate the quantities on which the filtering is performed. Additionally, the user must understand clearly the reasons for making such selections and the effect they will have on their conclusions. If not used with sufficient consideration, this methodology could bias the conclusions of the researcher. As with many analysis techniques, it can be possible to reach conclusions that are controlled more by the choice of analysis parameters than the data if care is not taken.

There are many other statistics and quantities determined that are not detailed here. Work is continuing to identify the most appropriate statistics and filtering criteria for different purposes.

Simulations

To enable testing and assessment of the analysis algorithms, software has been developed to simulate pseudo-random single molecule model data sets of 1 or more channels including noise. The motion and intensity variation of single molecules are simulated using a simple model. The resulting model images can be analysed and the

results compared to the known "true" tracks. The model for the simulations is outlined in "Appendix 3".

The model data sets of sample and bead images can be analysed using the algorithms described to determine the registration vectors, and perform feature detection and tracking. The simulations can therefore be used for systematic testing of the properties of the various algorithms, and this will be performed in the future. Some tests of the feature detection are described in "Appendix 1" and Fig. 11. The model for the samples will also be improved using a more physical description of the fluorophore/molecule behaviours and channel properties. Figure 5 shows an example of an analysed simulated single channel data set, with an example blinking, bleaching, 5-fluorophore intensity trace and the position traces of many features.

Software implementation

The algorithms discussed in this paper have been implemented as part of a pipeline for automated analysis of multi-dimensional single molecule microscopy data. The pipeline takes single molecule image stacks with metadata and carries out the feature detection and tracking, returning tracks that can be visualised, filtered and sorted using a sophisticated graphical tool (Fig. 6) developed as part of this work. Tracks can be exported for further analysis in additional tools developed as part of this suite or as simple CSV files for more general use. Consideration has been given to the necessary metadata to enable automation of the process. The metadata describe the data set [including identifying bead images, the number and attributes of channels in images (described in the "Single molecule data analysis in multiple channels" section) and detector parameters] and specifies any options for the analysis (e.g. evidence thresholds for feature detection, FWHM of features to assume for feature detection). Metadata are provided in simple keyword-value text files and image headers. Images are currently stored in HDF5 files, though import/export of standard images (e.g. TIFF) is possible.

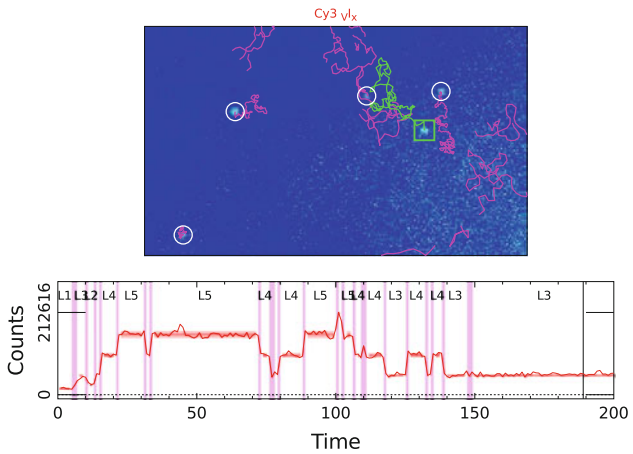

Fig. 5 Example of analysed simulated data. *Upper panel* shows a region of the field of view showing several tracked features present at the shown time [*white circles* and *green square* (highlighted)], with the position trace of all features up to the shown time shown in *purple* (*green* for the *highlighted trace*). *Lower panel* shows the intensity trace of the highlighted feature with steps and levels marked as in Fig. 4

The software has been developed on the GNU/Linux platform predominantly in C++ with some C, and some supporting tools developed using Python and scripting languages. The graphical visualisation tool uses the GTK+ toolkit (http://www.gtk.org) and Cairo graphical library (http://cairographics.org).

Examples

In this section we will briefly discuss and show examples of the use of the analysis software on some single molecule data from cells. Some examples have been published in Webb et al. (2008) and Vizcay-Barrena et al. (2011).

Multiplexed 8 channel FRET-polarisation

Inter- and intra-molecular distances in proteins are routinely measured using FRET. However, the FRET efficiency depends not only on the distance between donor and acceptor fluorescent probes (<10 nm), but also on the relative orientation between their dipolar moments (Stryer and Haugland 1967).

These distances and orientations can be inferred from single molecule FRET efficiencies and polarisation measurements respectively. This can be performed by switching the fluorescence excitation beam between two orthogonal polarisations in alternate frames and splitting the emission from the fluorophores (FRET donors and acceptors) into two colours and two polarisations (Webb et al. 2008). This results in eight channels to be analysed, each one corresponding to one of the two colours, one of two emission polarisations and one of two excitation polarisations. The multichannel infrastructure and methods in the analysis pipeline described in this paper were used to register channels, detect features and track molecules. They were then exported for further analysis (fitting of the FRET-polarisation model). Sorting the tracks by the donor-acceptor intensity correlation coefficient and looking for the greatest anti-correlation enabled the traces to be sifted. Figure 7 shows example traces.

Receptor separations

Ongoing work on measuring the separations of receptor proteins to determine their stoichiometry has also made use of the algorithms described. Single channel single molecule

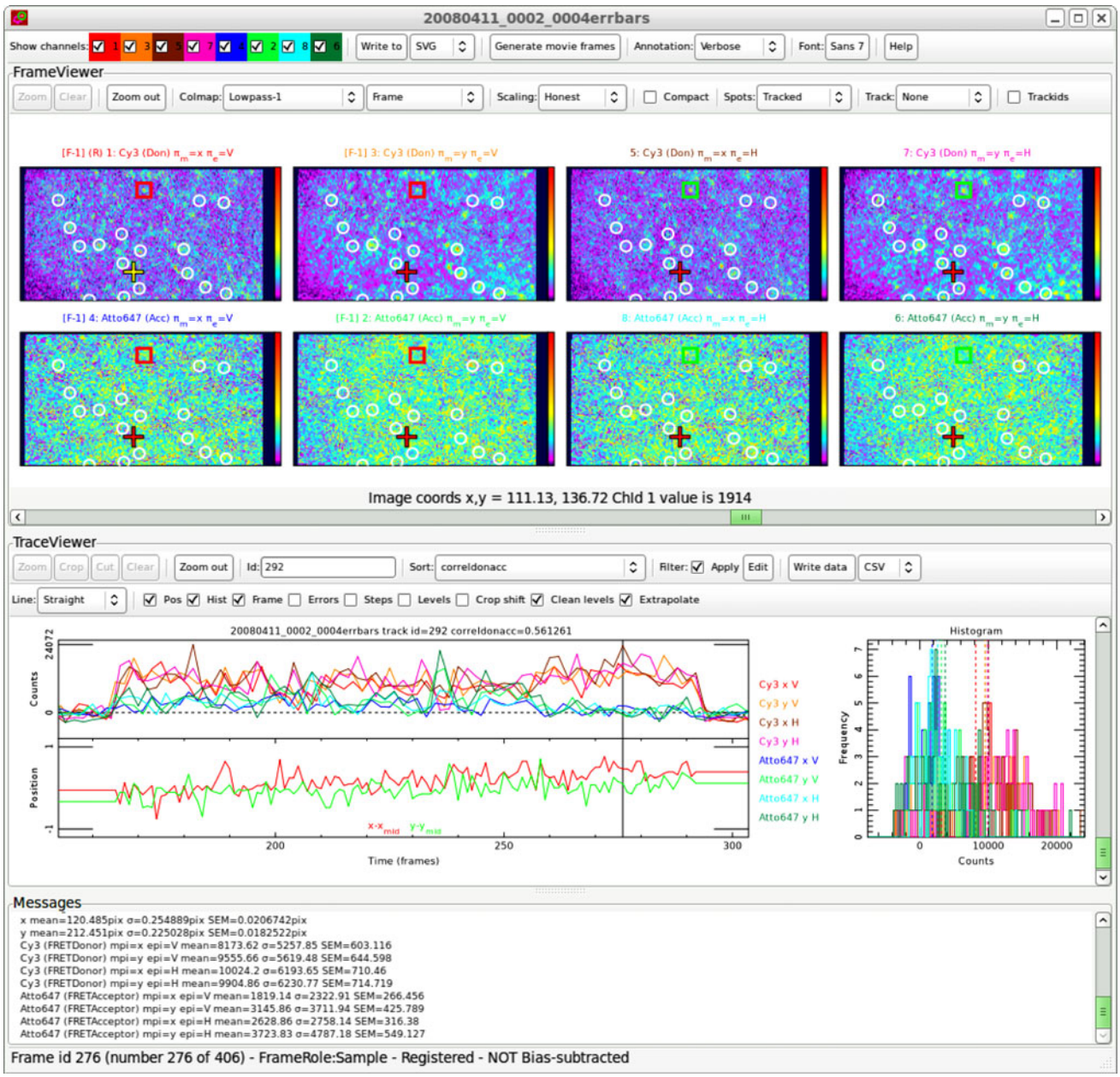

Fig. 6 Screen shot showing the graphical visualisation/analysis tool

data are analysed, and features with two or more intensity levels are identified for further analysis, in which a multiple fluorophore model is fitted to relevant regions of the original images to determine the locations and sub-pixel separations of features to high precision. The step and level identification methods already described help accelerate identification of likely candidate features for this post analysis. Additional filtering according to the statistics of the feature positions in each identified level is used to identify those features likely to have short receptor separations. Figures 4 and 8 show example traces.

EGFR tracking

Single-particle tracking is a useful tool to investigate the kinetics of heterogeneous systems such as the living cell. The method locates each particle and measures its individual movements, instead of an ensemble average. To illustrate our tracking algorithm we have chosen the much studied EGFR system for which kinetic data have been collected for nearly three decades, first using fluorescence recovery after photobleaching (FRAP) (Rees et al. 1984) and later using single molecule tracking. In this case we followed the

motion at the plasma membrane of EGFR molecules bound to a non-activating anti-EGFR affibody (Fig. 9a) and compared it with the motion of the same receptors bound to their activating ligand, EGF (Fig. 9b), both in live T47D cells. The plots of the mean square displacement (MSD) versus time (Fig. 9c) show that inactive EGFR molecules bound to fluorescently tagged anti-EGFR affibodies display confined diffusion as previously seen (Orr et al. 2005). Unlike the affibody (Nordberg et al. 2010), the binding of EGF induces receptor oligomerisation and the activation of the receptor. EGF receptors bound to fluorescently labelled EGF show reduced diffusion and increased confinement, consistent with oligomerisation-induced trapping (Kusumi et al. 2005). This behaviour is also expected in activated receptors that could be interacting with other intracellular proteins involved in ligand induced EGFR signalling and endocytosis (Chung et al. 2010; Xiao et al. 2008; Saffarian et al. 2007).

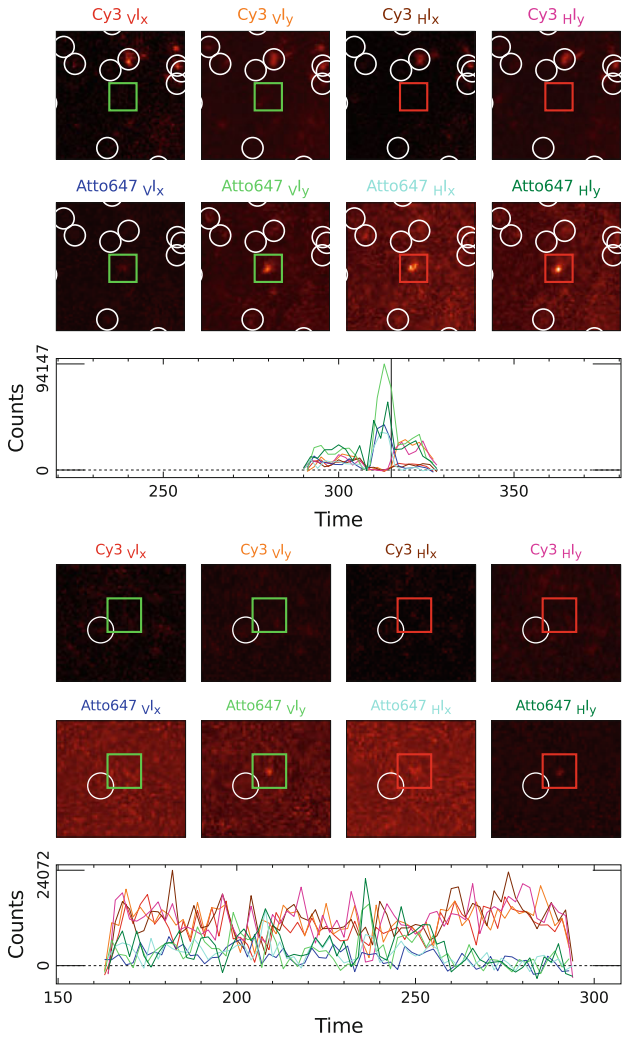

Fig. 7 Two intensity traces from multiplexed single-pair FRET and polarisation data in A431 cells (Webb et al. 2008). *Upper panels* show the region around the feature in a pair of consecutive time points (each time point containing 4 of the 8 channels). The selected feature is *highlighted* with the *red/green box*. Other tracked features are seen marked with *white circles*. The intensity trace for each channel is seen in the *lower panels*. *Colours* in the intensity trace correspond to the *colours* in the image titles above each image, indicating the channel to which each corresponds. The donor is Cy3, the acceptor Atto647N

Conclusions

This paper has presented a series of algorithms and methods that have been developed and implemented as part of a multidimensional single molecule microscopy analysis suite. This suite is the first complete pipeline of which we are aware aimed at analysis of data in an arbitrary number of channels, incorporating Bayesian segmentation feature detection methods and implemented for use by non-expert users. The approach is not specific to a single problem, but flexible to deal with multiple issues. The algorithms and software are in continual use and both minor and significant improvements to core aspects of the suite are planned and under development. These will be published in future papers, with further systematic testing and comparisons of the algorithms discussed.

Acknowledgments The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council through grant BB/G006911/1.

Appendix 1: Feature detection

Each feature is assumed to have the same shape, or profile, $t(x, y)$, which is normalised so that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x,y)\mathrm{d}x\mathrm{d}y = 1$. For sources of negligible size $t(x, y)$ will correspond to the point spread function of the microscope at the detector. The signal (excluding noise contributions) in the image at position (x, y) due to a feature at position $(x = X, y = Y)$ with total integrated intensity I is then

$$s(x, y) = It(x - X, y - Y).$$

Now consider a pixelised image with pixel centres at $x = x_i (i = 0, 1, 2, \ldots, N_x - 1)$ and $y = y_j (j = 0, 1, 2, \ldots, N_y - 1)$. The implementation described in this paper uses $x_i = i + 0.5$, $y_j = j + 0.5$, i.e. x and y are measured in units of pixel size with pixel centres at half-integer values. The signal in image pixel (i, j) due to a feature at (X, Y) is then

$$s_{ij} = I \int_{x=i}^{i+1} \int_{y=j}^{j+1} t(x - X, y - Y)\mathrm{d}x\mathrm{d}y = It_{ij}(X, Y),$$

where the profile has been integrated over the pixel area. s_{ij} is a component of a vector $\mathbf{s}$ of size $N_x N_y$, which represents

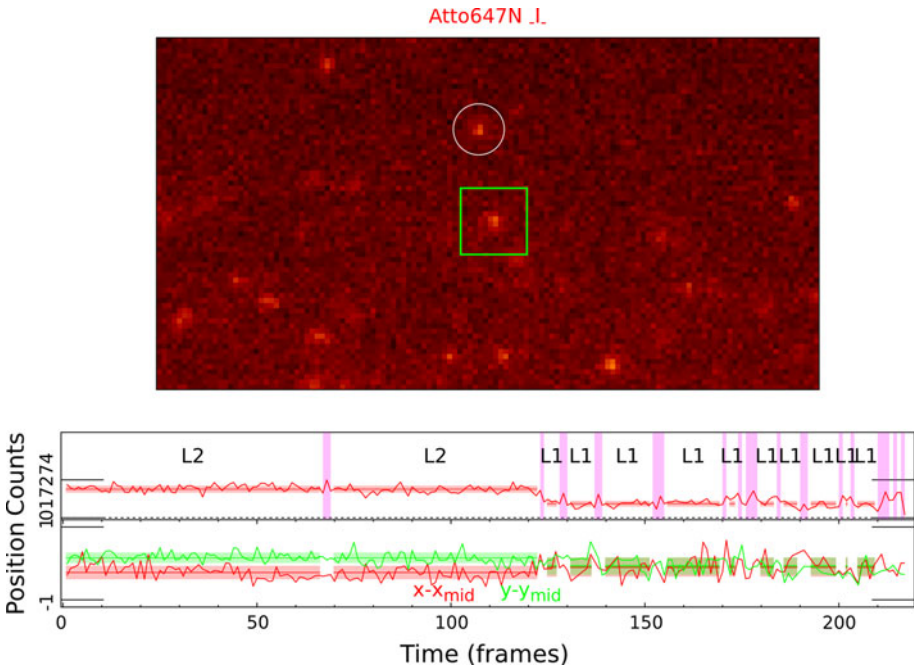

Fig. 8 Intensity and position trace of a feature in a cell corresponding to two molecules with sub-pixel separation. *Upper panel* shows image around feature (feature highlighted in *green*). The other highlighted feature also matched specified filter criteria. The *lower panels* show the intensity trace and identified levels (see Fig. 4), and the x and y position traces with *bars* showing the mean ± standard deviation of positions in each intensity level

signal in all pixels. Similarly vector $\mathbf{t}(X, Y)$ has components t_{ij}, and corresponds to the vector $\mathbf{t}$ in Hobson et al. (2009), except that in Hobson et al. (2009) $\mathbf{t}$ is normalised to unit peak. This simply means that for the definition here the amplitude A becomes the intensity I. Note that this integration of the profile over each pixel area correctly models the pixelation of the image, avoiding the "pixelation noise" effect discussed in Thompson et al. (2002).

If $t(x, y)$ is a Gaussian profile with standard deviation σ_t

$$t(x-X, y-Y) = \frac{1}{2\pi\sigma_t^2}\exp\left(-\frac{(x-X)^2+(y-Y)^2)}{2\sigma_t^2}\right)$$

and

$$t_{ij}(X,Y) = \frac{1}{4}\left(\mathrm{erf}\left(\frac{i+1-X}{\sigma_t\sqrt{2}}\right) - \mathrm{erf}\left(\frac{i-X}{\sigma_t\sqrt{2}}\right)\right) \times \left(\mathrm{erf}\left(\frac{j+1-Y}{\sigma_t\sqrt{2}}\right) - \mathrm{erf}\left(\frac{j-Y}{\sigma_t\sqrt{2}}\right)\right).$$

Bayes's theorem relates the *posterior probability*, $\Pr(M|d, K)$ that a particular hypothesis (or model), M, is true given some data, d, and background knowledge, K, to the probability, $\Pr(d|M, K)$, that the data would have been measured if the hypothesis were true, given the same background information. $\Pr(d|M, k)$ is called the *likelihood*, and is generally a lot easier to specify directly than the posterior. Bayes's theorem states that

$$\Pr(M|d,K) = \frac{\Pr(d|M,K)\Pr(M|K)}{\Pr(d|K)}$$

The hypothesis M may represent a particular model proposed to explain a system, or a set of parameters for such a model, the data set d some observations or measurements of the system, and the background knowledge K is any information about the system known prior to making the observations in question, for example known constraints on particular properties of the system. $\Pr(M|K)$ is the *prior probability* for the hypothesis, encoding the assumed probability of the hypothesis based on background information/bias only, before making the observations. If M represents a set of parameters, then the *prior probability* would be set to 0 for parameter values known to be impossible. $\Pr(d|K)$ is commonly called the *evidence*. For parameter estimation, where M represents a set of parameters of an assumed model, and the most probable range of parameter values consistent with the data is sought, the evidence is just a normalising factor and can be ignored. In the case of model selection, however, the evidence is crucial.

We wish to compare the posterior probability of hypotheses H_1 and H_0 at each point in the image. The ratio of the posterior probabilities, ρ, is given by

$$\rho = \frac{\Pr(H_1|\mathbf{d})}{\Pr(H_0|\mathbf{d})}$$

where $\mathbf{d}$ is the image data (Hobson et al. 2009). From Bayes's theorem this is

$$\rho = \frac{\Pr(\mathbf{d}|H_1)\Pr(H_1)}{\Pr(\mathbf{d}|H_0)\Pr(H_0)},$$

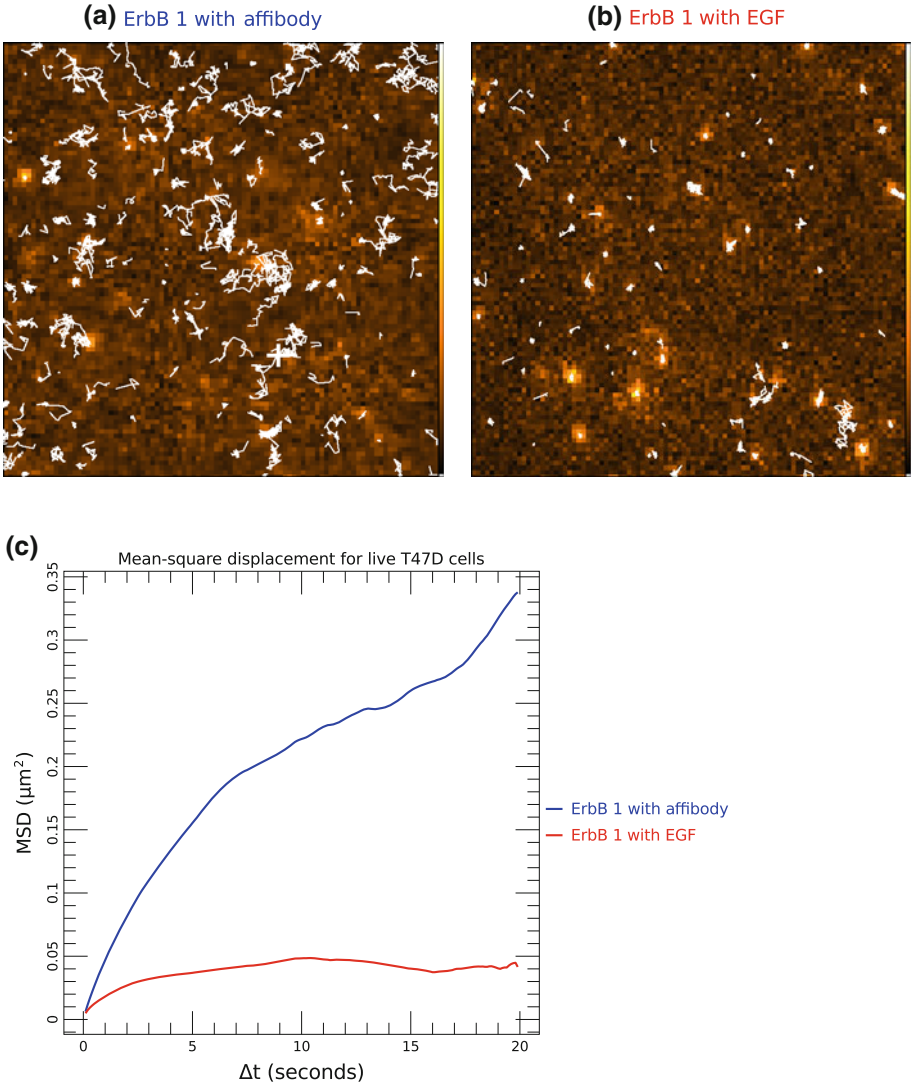

Fig. 9 Tracking of EGF receptors with fluorescently tagged anti-EGFR affibodies or bound to fluorescently labelled EGF in live T47D cells. **a** and **b** show an example single molecule image and track position traces for each treatment. **c** shows the measured mean square displacement versus time for each treatment

where $\Pr(H_1)$ and $\Pr(H_0)$ are the prior probabilities of the two hypotheses, and $\Pr(\mathbf{d}|H_1)$ and $\Pr(\mathbf{d}|H_0)$ are evidences for model H_1 and H_0 respectively. The evidences are calculated from the likelihoods and prior probabilities for all the possible parameters for each model,

$$\Pr(\mathbf{d}|H_1) = \int\int \Pr(\mathbf{d}|I, B_1, H_1)\Pr(I, B_1|H_1)\mathrm{d}I\mathrm{d}B_1$$

and

$$\Pr(\mathbf{d}|H_0) = \int \Pr(\mathbf{d}|B_0, H_0)\Pr(B_0|H_0)\mathrm{d}B_0.$$

Writing $\rho_0 = \frac{\Pr(H_1)}{\Pr(H_0)} = \langle n \rangle$ (where $\langle n \rangle$ is the expected number of features per pixel in the image; Hobson et al. (2009)) gives the evidence ratio

$$R = \frac{\rho}{\rho_0} = \frac{\Pr(\mathbf{d}|H_1)}{\Pr(\mathbf{d}|H_0)}.$$

R is calculated for each pixel in the image with the feature profile for model H_1 centred at the pixel centre. Pixels in this *evidence map* with $R > R_{\min}$ ($R_{\min} = 1/\langle n \rangle$) can be identified as features, i.e. the locations of the fluorophores, since in these pixels model H_1 is more probable than model H_0. In practise the threshold on R, $R_{\min}$, can be decreased/increased to increase/decrease the feature detection rate (and false discovery rate). Since the number of single point emitting sources expected is not necessarily easy to determine, and there is some freedom to choose priors on I and B (e.g. change $I_{\max}$ and $B_{\max}$, see below), which will also change the evidence, such tuning of the threshold is generally going to be necessary.

Single evidence ratio peaks are identified by segmenting the evidence map above the threshold into local maximum regions using a watershed-like algorithm, as follows.

1. Find the pixel with the highest value in the evidence map above the threshold, $R_{\min}$, which has not already

been identified as associated with a source. If there is no such pixel the segmentation is complete.
2. Identify this pixel as the location of a new source. This becomes the current pixel and its value of R is called R_{current}.
3. Grow the set of pixels that are associated with this source by adding any of the eight pixels directly or diagonally adjacent to the current pixel that satisfy $R_{\text{min}} < R \leq R_{\text{current}}$. For each such added pixel repeat this step using that pixel as the new current pixel and its R as R_{current}, recursively continuing until no further pixels can be added.
4. Go to step 1.

This method will find all peaks above R_{min} and will correctly identify multiple peaks within single regions of the the map above R_{min}. For each such segmented region, the pixel with the maximum value of R is identified as the location of the source. See Fig. 10 for an example.

Constraints on the feature intensity and background are imposed by choosing appropriate prior probability distributions. We assume priors $\Pr(I,B_1|H_1) = \pi_I(I)\pi_B(B_1)$ and $\Pr(B_0|H_0) = \pi_B(B_0)$ with $\pi_I(I)$ and $\pi_B(B)$ uniform for I and B in a finite range, i.e.

$$\pi_I(I) = \begin{cases} 1/I_{\text{max}} & \text{if } 0 \leq I \leq I_{\text{max}} \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_B(B) = \begin{cases} 1/B_{\text{max}} & \text{if } 0 \leq B \leq B_{\text{max}} \\ 0 & \text{otherwise} \end{cases}.$$

For each detected feature location the most probable intensity and background level, $\hat{I}$ and $\hat{B}_1$, can be determined by Bayesian parameter estimation (Hobson et al. 2009). Maps of $\hat{I}$ can be efficiently calculated for each pixel at the same time as R, making estimates of the feature intensities available as soon as the feature locations have been identified.

The covariance matrix of $P_1 = \Pr(I, B|\mathbf{d}, H_1)$, $\mathbf{C} = (-\mathbf{H})^{-1}$, can be used to determine the root mean square uncertainties in $\hat{I}$ and $\hat{B}_1$, $\sigma(\hat{I}) = \sqrt{\mathbf{C}_{11}}$ and $\sigma(\hat{B}_1) = \sqrt{\mathbf{C}_{22}}$, where $\mathbf{H}$ is the Hessian matrix,

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \ln P_1}{\partial I^2} & \frac{\partial^2 \ln P_1}{\partial I \partial B} \\ \frac{\partial^2 \ln P_1}{\partial B \partial I} & \frac{\partial^2 \ln P_1}{\partial B^2} \end{pmatrix}.$$

evaluated at $(\hat{I}, \hat{B}_1)$. Note that these uncertainties do not account for any error in the feature FHWM or position, and as such are likely to underestimate the error to some extent.

The feature detection was tested by simulating multiple data sets (as described in "Simulations" and "Appendix 3"), each with 100 frames of 100 constant intensity non-moving features of FWHM 3 pixels separated by at least 9 pixels. Data sets were simulated with feature intensities varying from 10 to 2,000 photons per frame (integrated over the profile), and with constant uniform background intensity 0, 1, 5 or 10 photons per pixel per frame. Poisson noise, a linear gain of 200 and readout noise of standard deviation 5 were added. Each simulation was performed using a Gaussian profile and also using an Airy disc profile. In the latter case, the Airy profile contribution to a particular pixel was determined by integrating an oversampled Airy profile over the pixel. The feature detection and measurement were performed on each simulated data set. The number of features correctly detected and the number of false detections (detections not corresponding to true features) were determined. Figure 11 shows the results as plots of the true-positive and false discovery rates as

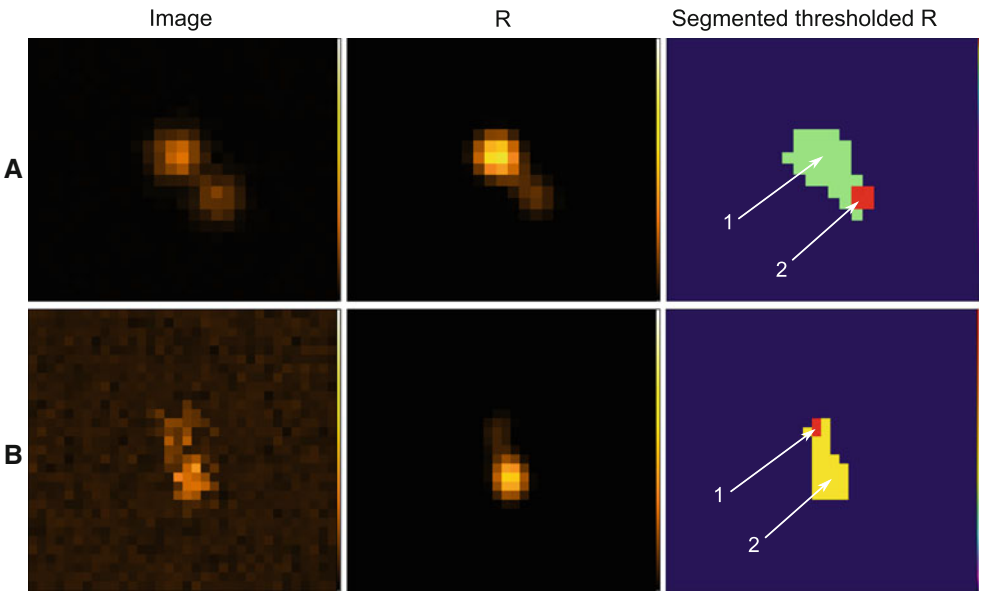

Fig. 10 Feature detection. *Three panels* showing the image data, the evidence map and the segmented evidence map. The segmented map shows the identification of two features within one region of the map above R_{min}. The pixel with the highest value of R in each segmented region is the corresponding source location (identified by *arrows 1* and *2*)

functions of photons in a feature and signal-to-noise at the peak of a feature. For a signal-to-noise in the peak pixel above about 1 the detection rate is better than 98 percent and the false discovery rate less than a few percent. There is little difference in these figures whether the model profile uses and Airy disc or a Gaussian profile, confirming that the Gaussian model in the analysis is justified for typical single molecule data. For the simulations with zero background emission but low signal-to-noise, the false discovery rate is a little higher with an Airy disc than with a Gaussian. This appears to be occasional false detections of features from emission in the Airy rings due to the approximate image noise estimation and Gaussian profile assumption. However this should not affect single molecule data from cells in which there is always some background intensity. There is no difference in the true positive rate (detection rate) between Airy and Gaussian profile simulations. This will be revisited when the full physical noise model has been implemented to see if it affects this observation.

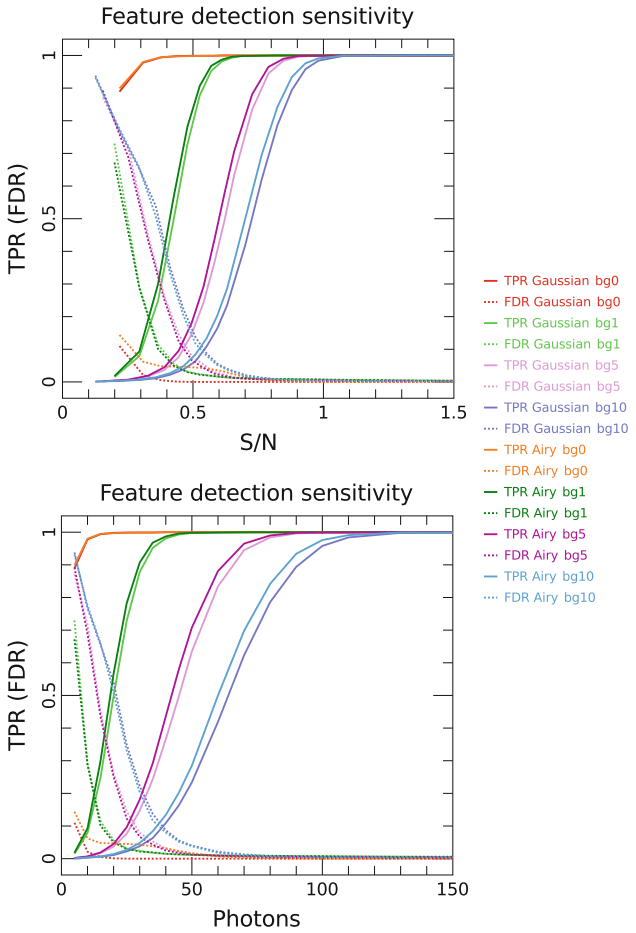

Fig. 11 Feature detection metrics measured from various simulated data sets (see text). The true-positive rate and false discovery rate as functions of the signal-to-noise at the peak of a feature (*upper panel*) and number of photons in a feature (*lower panel*). The true-positive rate, *TPR*, is the fraction of true features that were detected, i.e. *TPR* = number of true features detected/total number of true features. The false discovery rate, *FDR*, is the fraction of feature detections that do not correspond to a true feature, *FPR* = number of false positives/total number of detections (true and false). The signal-to-noise, *S/N*, is defined here as *S/N* = model intensity at peak of feature/SD of intensity at peak

Appendix 2: Registration transformation

The function to transform a position $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ in the non-reference channel to the corresponding point $\mathbf{r}_{\text{ref}} = \begin{pmatrix} x_{\text{ref}} \\ y_{\text{ref}} \end{pmatrix}$ in the reference channel is

$$\begin{aligned}\mathbf{r}_{\text{ref}} &= \mathbf{c}_{\text{ref}} + \mathbf{Rot}(-\theta)\mathbf{Rot}(\theta_S)\mathbf{S}\,\mathbf{Rot}(-\theta_S)(\mathbf{r} - \mathbf{c} + \mathbf{D}) \\ &= \mathbf{A}(\mathbf{r}, \mathbf{T}).\end{aligned}$$

$\mathbf{c}$ and $\mathbf{c}_{\text{ref}}$ are the midpoints of the channel image region for the non-reference and reference channel respectively, and are adopted as sensible origins for the transformations. $\mathbf{Rot}(\phi)$ is the matrix $\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$, which rotates a vector clockwise through angle ϕ. $\mathbf{D} = \begin{pmatrix} D_x \\ D_y \end{pmatrix}$ is the translation to be applied, $\mathbf{S} = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$ specifies the orthogonal scaling factors, S_1 and S_2. θ_S is the angle specifying the orientation of the scaling axes and θ is the angle of the rotation. The parameters to be determined are $\mathbf{T} = [D_x, D_y, S_1, S_2, \theta, \theta_S]$.

The Scott and Longuet-Higgins algorithm (Scott and Longuet-Higgins 1991) is used to match beads in each non-reference channel with their counterparts in the reference channel. If the positions of the $N_{\text{bead,ref}}$ beads identified in the reference channel are $\mathbf{r}_{\text{bead,ref,i}}$ $(i = 1 \ldots N_{\text{bead,ref}})$ and in the non-reference channel $\mathbf{r}_{\text{bead},j}$ $(j = 1 \ldots N_{\text{bead}})$, a proximity matrix, $\mathbf{G}$ is calculated with elements $G_{ij} = \exp\left(-\frac{\left|\mathbf{r}_{\text{bead,ref},i} - \mathbf{r}_{\text{bead},j}\right|^2}{\sigma^2_{\text{slh},ij}}\right)$ where σ_{slh} controls the scale on which features in different channels may be associated and influences the effectiveness of the algorithm. A value $\sigma_{\text{slh}} = 10$ pixels was used. The algorithm exploits the properties of singular value decomposition of $\mathbf{G}$ to produce a modified proximity matrix that can be used to identify which bead in the reference channel corresponds to which in the non-reference channel. For the small rotations (typically $<1°$) and scalings (typically $<4\%$) found, and provided the translations are not too large (approximately <10 pixels) this algorithm works very well. In some cases where the offset, $\mathbf{D}$, gets larger than about 10 pixels, the

algorithm fails. Such cases are easily overcome by trying offsets to $\mathbf{r}_{\text{bead},j}$, $\mathbf{r}_{off} = \begin{pmatrix} p \times 10\text{pix} \\ q \times 10\text{pix} \end{pmatrix}$ with integers p and q both in the range $-3\ldots3$, before calculating the proximity matrix.

With corresponding beads in each channel identified, determination of the transformation parameters simply requires finding the transformations for which non-reference channel beads are best aligned with their reference channel counterparts. It is not necessary to consider all possible combinations of beads in one channel with beads in the other. If the corresponding pairs of the N_{matched} matched beads have positions $\mathbf{r}_{\text{bead},\,k}$ and $\mathbf{r}_{\text{bead,ref},k}$ $(k = 1\ldots N_{\text{matched}})$, define

$$\Delta r_{\text{bead},k}(\mathbf{T}) = \left|\mathbf{A}(\mathbf{r}_{\text{bead},k}, \mathbf{T}) - \mathbf{r}_{\text{bead,ref},k}\right|,$$

the difference between the reference bead position and the position of the non-reference bead registered with the transformation $\mathbf{T}$. One approach to determining the optimal transformation parameters would be find the $\mathbf{T}$ that minimises the sum of the squared $\Delta r_{\text{bead},k}(\mathbf{T})$s, i.e. minimise $\sum_{k=1}^{N_{\text{matched}}} \Delta r_{\text{bead},k}(\mathbf{T})^2$, however this method would allow any mismatched beads to influence the determined transformation. Although the Scott and Longuet-Higgins method does a good job of matching beads, there are often a small number of mismatched beads (Fig. 1c). An alternative objective function, $P(\mathbf{T})$, is therefore used, with

$$P(\mathbf{T}) = \sum_{k=1}^{N_{\text{matched}}} \exp\left(-\frac{\Delta r_{\text{bead},k}(\mathbf{T})^2}{2\sigma_P^2}\right),$$

which must be maximised, where σ_P controls the sensitivity to mismatch. In this objective function any beads that are significantly mismatched compared to σ_P will produce small local peaks in $P(\mathbf{T})$ away from the main maximum, so as long as the optimiser finds the main maximum these mismatched beads will not influence the final parameters. Two passes of optimisation are performed, first with $\sigma_P = 10$ pixels, which broadens the maximum ensuring the global maximum is identified approximately, then a second pass beginning at the maximum of $P(\mathbf{T})$ from the first pass but with σ_P set to the standard deviation of the feature profiles, σ_t. The second pass thus homes in on the peak more precisely, with $P(\mathbf{T})$ approximating a correlation coefficient of the profile of each bead with its counterpart. Optimisation of $P(\mathbf{T})$ is performed using the Downhill Simplex method (Press et al. 1992).

To assess the quality of the determined transformation solution, three scores are calculated. Bead positions are registered, and those that are within $\sigma_t/2$ of their corresponding reference channel bead are identified as *matched*. The scores are

- f_{matched}, the ratio of matched features to maximum possible matches, where the number of maximum possible matches is the sum of $min(N_{\text{bead}}, N_{\text{bead,ref}})$ over images of beads,
- Δr_{69} and Δr_{95}, the largest separation of the best matched 69%/95% of the matched beads.

$f_{\text{matched}} = 0.7$ is used as a threshold below which registration is assumed to have failed. Δr_{69} and Δr_{95} are useful diagnostics giving an idea of the typical agreement of registered features, but they will be a combination of the localisation uncertainty in the measurement of the beads and the error in the registration. As such they can be used as estimates of the upper limit of the registration error, where Δr_{95} is the more conservative estimate.

Appendix 3: Simulations

This appendix details the model used for creating simulated data sets that may then be used to test the single molecule analysis methods.

- Registration transformations for each non-reference channel (as described in the "Registration transformation" section) are generated. The parameters $[D_x, D_y, S_1, S_2, \theta, \theta_S]$ for each channel are sampled from probability distributions that approximate the typical range of values seen in real data sets.
- A number N_{tracks} of tracks is modelled.
  - Each track consists of a group of N_{fluor} co-located fluorophores moving together and individually fluorescing/blinking/bleaching.
  - N_{fluor} is sampled from a Poisson distribution of specified mean (currently 2) and $N_{\text{fluor}} = 0$ is avoided.
  - Each fluorophore is either off (not yet fluorescing), on (fluorescing), blinking or photo-bleached.
  - Fluorophores are individually switched on, may then blink, recover and photo-bleach all at times determined randomly according to rates specified for each event.
  - Each track has an initial position, speed and direction of motion chosen at random, the speed chosen from a specified normal distribution. At each time step the direction and speed are modified by applying changes sampled from normal distributions with specified standard deviations.
  - At each time point the track intensity is $ev_{\text{fac}} I_{\text{fluor}}$ times the number of fluorescing fluorophores, where I_{fluor} is the individual fluorophore intensity and $ev_{\text{fac}} = \exp -d/d_{\text{ev}}$. d is the depth of the fluorophores in the evanescent field, and is constant and

randomly chosen for each track between 0 and twice the evanescent field depth, d_{ev}.
- At present the intensity of fluorophores is the same in each channel. Models for different types of experiment will be implemented to improve on this, e.g. FRET and co-localisation experiments.

- Single molecule images are then modelled at each time point, considering each fluorophore group as a Gaussian spot of common full width half maximum (FWHM). Real images contain background structures on a variety of spatial scales, for example autofluorescence, which will challenge feature detection algorithms. To simulate background emission with spatial structure on a range of spatial scales, a single image consisting of Perlin noise is generated and used as the background signal in all channels at every time point. This Perlin noise image consists of random spatial fluctuations in intensity, on spatial scales L ranging from the model feature profile FWHM to the image size, with amplitudes proportional to $1/L$. This is not necessarily an accurate reflection of the background structure, but is a simple parametrised way to add challenging background emission that can be improved upon when the background properties are better characterised. The intensity incident on each pixel is then the contribution from fluorophores and the background structure, from which the number of photons in the pixel is Poisson sampled. The quantum efficiency of the detector is modelled by sampling from a binomial distribution (which, convolved with the initial Poisson distribution, gives another Poisson) and a linear gain is then applied followed by the addition of Gaussian noise and an offset to model the bias and readout noise. This is a simplified model for a CCD response. Corresponding images are produced for each channel, but with the background image, fluorophore positions and FWHMs transformed according to the registration vectors before sampling onto the model CCD/pixel grid.
- Bead images are also simulated, where bead samples are modelled as a random spatially uniform distribution of Gaussian point sources, sampled on the model detector using the same detector model as above.

References

Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313(5793):1642–1645. doi:10.1126/science.1127344, http://www.ncbi.nlm.nih.gov/pubmed/16902090

Bräuchle C, Lamb DC, Michaelis J (2010) Single particle tracking and single molecule energy transfer. Wiley, London

Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells. Nature 464(7289):783–787. doi:10.1038/nature08827, http://dx.doi.org/10.1038/nature08827

Citri A, Yarden Y (2006) EGF-ERBB signalling: towards the systems level. Nature Rev Mol Cell Biol 7(7):505–516. doi:10.1038/nrm1962, http://www.ncbi.nlm.nih.gov/pubmed/16829981

Cronin B, de Wet B, Wallace MI (2009) Lucky imaging: improved localization accuracy for single molecule imaging. Biophys J 96(7):2912–2917. doi:10.1016/j.bpj.2008.12.3945

Das SK, Darshi M, Cheley S, Wallace MI, Bayley H (2007) Membrane protein stoichiometry determined from the step-wise photobleaching of dye-labelled subunits. Chembiochem: a European journal of chemical biology 8(9):994–999. doi:10.1002/cbic.200600474, http://www.ncbi.nlm.nih.gov/pubmed/17503420

Dunne PD, Fernandes RA, McColl J, Yoon JW, James JR, Davis SJ, Klenerman D (2009) DySCo: quantitating associations of membrane proteins using two-color single-molecule tracking. Biophys J 97(4):L5–L7. doi:10.1016/j.bpj.2009.05.046, http://www.ncbi.nlm.nih.gov/pubmed/19686638

Fili N, Mashanov GI, Toseland CP, Batters C, Wallace MI, Yeeles JTP, Dillingham MS, Webb MR, Molloy JE (2010) Visualizing helicases unwinding DNA at the single molecule level. Nucleic Acids Res 38(13):4448–4457. doi:10.1093/nar/gkq173

Hern Ja, Baig AH, Mashanov GI, Birdsall B, Corrie JET, Lazareno S, Molloy JE, Birdsall NJM (2010) Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules. Proc Natl Acad Sci USA 107(6):2693–2698. doi:10.1073/pnas.0907915107

Hobson MP, McLachlan C (2003) A Bayesian approach to discrete object detection in astronomical data sets. Monthly Notices R Astron Soc 338(3):765–784. doi:10.1046/j.1365-8711.2003.06094.x, http://cdsads.u-strasbg.fr/abs/2003MNRAS.338..765H

Hobson MP, Rocha G, Savage RS (2009) Bayesian source extraction, chap 7.5.5. In: Hobson MP, Jaffe AH, Liddle AR, Mukherjee P, Parkinson D (eds) Bayesian methods in cosmology. Cambridge University Press, Cambridge, p 187

Holden SJ, Uphoff S, Kapanidis AN (2011) DAOSTORM: an algorithm for high-density super-resolution microscopy. Nature Methods 8(4):279–280. doi:10.1038/nmeth0411-279, http://www.ncbi.nlm.nih.gov/pubmed/21451515

Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, Danuser G (2008) Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods 5(8):695–702. doi:10.1038/nmeth.1237, http://dx.doi.org/10.1038/nmeth.1237

Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science 314(5802):1144–1147. doi:10.1126/science.1131399

Kusumi A, Ike H, Nakada C, Murase K, Fujiwara T (2005) Single-molecule tracking of membrane molecules: plasma membrane compartmentalization and dynamic assembly of raft-philic signaling molecules. Sem Immunol 17(1):3–21. doi:10.1016/j.smim.2004.09.004, http://www.ncbi.nlm.nih.gov/pubmed/15582485

Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, Armitage JP (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. Nature 443(7109):355–358. doi:10.1038/nature05135, http://www.ncbi.nlm.nih.gov/pubmed/16971952

Mashanov GI, Molloy JE (2007) Automatic detection of single fluorophores in live cells. Biophys J 92(6):2199–2211. doi:10.1529/biophysj.106.081117

Mortensen KI, Churchman LS, Spudich Ja, Flyvbjerg H (2010) Optimized localization analysis for single-molecule tracking and

super-resolution microscopy. Nature Methods 7(5). doi:10.1038/nmeth.1447, http://www.ncbi.nlm.nih.gov/pubmed/20364147

Nordberg E, Ekerljung L, Sahlberg SH, Carlsson J (2010) Effects of an EGFR-binding affibody molecule on intracellular signaling pathways, pp 967–972, doi:10.3892/ijo

Orr G, Hu D, Ozçelik S, Opresko LK, Wiley HS, Colson SD (2005) Cholesterol dictates the freedom of EGF receptors and HER2 in the plane of the membrane. Biophys J 89(2):1362–1373. doi:10.1529/biophysj.104.056192

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes. Cambridge University Press, Cambridge. http://library.lanl.gov/numerical/bookcpdf.html

Rees AR, Gregoriou M, Johnson P, Garland PB (1984) High affinity epidermal growth factor receptors of A431 cells have restricted lateral diffusion the surface. EMBO J 3(8):1843–1847

Roy R, Kozlov AG, Lohman TM, Ha T (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. Nature 461(7267):1092–1097. doi:10.1038/nature08442

Rust M, Bates M, Zhuang X (2006) Stochastic optical reconstruction microscopy (STORM) provides sub-diffraction-limit image resolution. Nature Methods 3(10):793. doi:10.1038/nmeth929.Stochastic, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2700296/

Saffarian S, Li Y, Elson EL, Pike LJ (2007) Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. Biophys J. doi:10.1529/biophysj.107.105494

Savage RS, Oliver S (2007) Bayesian methods of astronomical source extraction. Astrophys J 661(2):1339–1346. doi:10.1086/515393, http://cdsads.u-strasbg.fr/abs/2007ApJ...661.1339S

Schaaf MJM, Koopmans WJa, Meckel T, van Noort J, Snaar-Jagalska BE, Schmidt TS, Spaink HP (2009) Single-molecule microscopy reveals membrane microdomain organization of cells in a living vertebrate. Biophys J 97(4):1206–1214. doi:10.1016/j.bpj.2009.05.044, http://www.ncbi.nlm.nih.gov/pubmed/19686669

Scott GL, Longuet-Higgins HC (1991) An algorithm for associating the features of two images. Proc R Soc B Biol Sci 244(1309):21–26. http://www.jstor.org/stable/76644

Selvin P, Ha T (2007) Single molecule techniques: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Semrau S, Schmidt T (2007) Particle image correlation spectroscopy (PICS): retrieving nanometer-scale correlations from high-density single-molecule position data. Biophys J 92(2):613–621. doi:10.1529/biophysj.106.092577, http://www.ncbi.nlm.nih.gov/pubmed/17085496

Sergé A, Bertaux N, Rigneault H, Marguet D (2008) Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nature Methods 5(8):687–694. doi:10.1038/nmeth.1233, http://dx.doi.org/10.1038/nmeth.1233

Stryer L, Haugland RP (1967) Energy transfer: a spectroscopic ruler. Proc Natl Acad Sci USA 58(2):719–726. http://www.hubmed.org/display.cgi?uids=5233469

Teramura Y, Ichinose J, Nishida K, Yanagida T, Sako Y (2006) Single-molecule analysis of epidermal growth factor binding on the surface of living cells. EMBO J 1–8. doi:10.1038/sj.emboj.7601308

Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82(5):2775–83. doi:10.1016/S0006-3495(02)75618-X

Vizcay-Barrena G, Webb SED, Martin-Fernandez ML, Wilson ZA (2011) Subcellular and single-molecule imaging of plant fluorescent proteins using total internal reflection fluorescence microscopy (TIRFM). Wilson J Exp Bot. doi:10.1093/jxb/err212

Webb SED, Needham SR, Roberts SK, Martin-Fernandez ML (2006) Multidimensional single-molecule imaging in live cells using total-internal-reflection fluorescence microscopy. Optics Lett 31(14):2157–2159. http://www.ncbi.nlm.nih.gov/pubmed/16794711

Webb SED, Rolfe DJ, Needham SR, Roberts SK, Clarke DT, McLachlan CI, Hobson MP, Martin-Fernandez ML (2008) Simultaneous widefield single molecule orientation and FRET microscopy in cells. Optics Exp 16(25):20258–20265. http://www.ncbi.nlm.nih.gov/pubmed/19065164

Weiss S (1999) Fluorescence spectroscopy of single biomolecules. Science 283(5408):1676–1683. doi:10.1126/science.283.5408.1676, http://www.sciencemag.org/cgi/doi/10.1126/science.283.5408.1676

Xiao Z, Zhang W, Yang Y, Xu L, Fang X (2008) Single-molecule diffusion study of activated EGFR implicates its endocytic pathway. Biochem Biophys Res Commun 369(2):730–734. doi:10.1016/j.bbrc.2008.02.084, http://www.ncbi.nlm.nih.gov/pubmed/18313398

Yanagida T, Ishii Y (2009) Single molecule dynamics in life sciences. Wiley, London

Yoon JW, Bruckbauer A, Fitzgerald WJ, Klenerman D (2008) Bayesian inference for improved single molecule fluorescence tracking. Biophys J 94(12):4932–4947. doi:10.1529/biophysj.107.116285, http://www.ncbi.nlm.nih.gov/pubmed/18339757

Zhang W, Jiang Y, Wang Q, Ma X, Xiao Z, Zuo W, Fang X, Chen YG (2009) Single-molecule imaging reveals transforming growth factor-beta-induced type II receptor dimerization. Proc Nat Acad Sci USA 106(37):15679–15683. doi:10.1073/pnas.0908279106, http://www.ncbi.nlm.nih.gov/pubmed/19720988