



HAL
open science

Photobleaching Step Counting and Localization for Fluorescence Microscopy

Charles Truong, Thomas Bugea, Baptiste Bouhet, Guillaume Tresset,
François Marquier, Karen Perronet

► **To cite this version:**

Charles Truong, Thomas Bugea, Baptiste Bouhet, Guillaume Tresset, François Marquier, et al.. Photobleaching Step Counting and Localization for Fluorescence Microscopy. 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul 2024, Orlando (Floride), United States. pp.1-4, 10.1109/EMBC53108.2024.10782704 . hal-04869324

HAL Id: hal-04869324

<https://hal.science/hal-04869324v1>

Submitted on 7 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Photobleaching Step Counting and Localization for Fluorescence Microscopy

Charles Truong¹, Thomas Bugea^{2,3}, Baptiste Bouhet², Guillaume Tresset³, François Marquier², Karen Perronet²

¹ *Université Paris Saclay, Université Paris Cité, ENS Paris-Saclay, CNRS, SSA, INSERM, Centre Borelli*

² *Université Paris-Saclay, ENS Paris-Saclay, CNRS, CentraleSupélec, LuMin*

³ *Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides*

Abstract—Fluorescence microscopy is a fundamental tool for studying biological systems at the single-molecule level. In particular, photobleaching step patterns can provide valuable insights into molecule interactions. Several works have proposed an automated and robust way to estimate when photobleaching steps occur. However, most methods are challenging to calibrate by non-experts in statistics or signal processing and struggle with low signal-to-noise ratios. This work introduces a supervised approach to localize photobleaching steps in fluorescent microscopy traces. Our method does not require any calibration but instead needs labels, that is, a few traces where an expert has manually provided the step positions. Our algorithm uses this information to automatically tune a change-point detection method that can reproduce the expert’s annotations on new signals. We show on simulated data that our approach better copes with noise than existing ones. We then illustrate how our algorithm can be used to estimate the translation speed of a ribosome.

Index Terms—photobleaching step analysis, fluorescence microscopy, single-molecule microscopy, change-point detection

I. INTRODUCTION

Fluorescence microscopy is a popular tool for tracking the interactions between different biomolecules in space and time at the single-molecule level [4], [6]. The measurement involves recording fluorescence signals in the respective spectral ranges of the fluorescent entities used to label the usually non-fluorescent biomolecules. To ensure that labelling does not interfere with the activity of these biomolecules, fluorescent proteins or organic dyes, which may be small in size but are prone to photobleaching after a certain excitation time, are often used. A compromise must be made between signal-to-noise ratio and observation time, making fluorescence signals delicate to analyze [1], [5]. A single fluorescent molecule has a characteristic ON/OFF fluorescence signal [3]. Either the molecule emits photons at a constant rate, depending on its intrinsic properties and excitation power, or it has photobleached and can no longer emit photons. Only the blinking phenomenon (transient transition to a black state) can interfere with these characteristics. So, to detect interactions between an immobile biomolecule and another capable of diffusing in the medium, we can look locally for breaks in the fluorescence signal of the mobile molecule.

C. Truong is funded by the PhLAMES chair (ENS Paris-Saclay). T. Bugea is supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (Labex NanoSaclay, reference: ANR-10-LABX-0035).

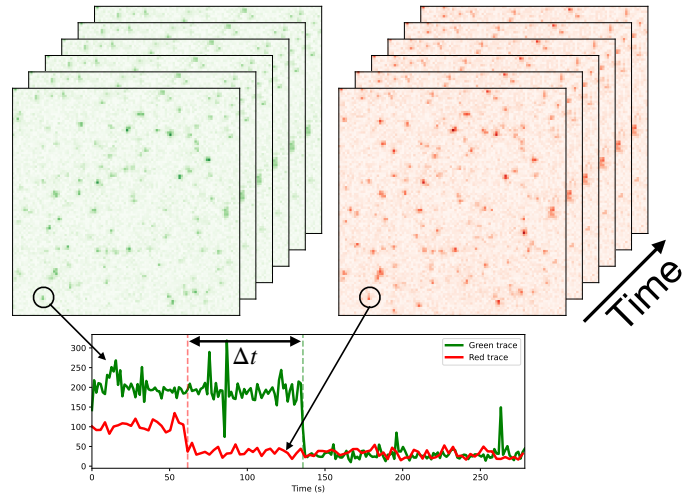


Fig. 1. We observe red and green fluorophores with a fluorescence microscope to estimate ribosomes’ translation rate. This yields two image stacks. A trace (or signal) is the intensity of a small region of interest (black circle) across time. An abrupt downward step indicates that the observed fluorescent marker has stopped emitting. The time difference Δt between the two steps (green and red) measures the ribosome’s translation time.

For example, we are using this method to study the translation speed of a single protein [2]. During this process, a macromolecule called a ribosome recognizes a messenger RNA (mRNA) and then reads its contents to translate it into a protein. The position of the ribosome cannot be tracked in real time. Fluorescent markers of different colours are attached to specific points on the mRNA to obtain this information. The ribosome will detach these during translation, losing their fluorescence signal. In this way, we have access to the instants at which the ribosome reaches these positions on the mRNA, and then, via statistical processing of the instants at which the markers disappear, we can trace the translation rate. A schematic illustration is shown in Figure 1.

Despite its broad applicability, few works have provided robust algorithms for automatic step analysis in photobleaching data. One of the most recent articles [3] describes a procedure called *QuickPBSA* that consists of a rough step estimation followed by a refinement. This method requires an estimate of the average step amplitude, which can be acquired by visually inspecting several traces. The authors show they can correctly estimate the step numbers on traces of around 3000 samples from DNA origami data. Compared to the setting of [3],

our data are shorter (maximum of 200 samples) and have a far lower signal-to-noise ratio. As shown in the experiments, this significantly degrades detection accuracy, motivating the introduction of a more robust approach.

Our contribution consists in an automatic and robust procedure to perform step analysis of photobleaching traces. Our approach uses a change-point detection algorithm and a supervised calibration procedure. The resulting algorithm, denoted *SuperSA* (Supervised Step Analysis), only requires a few annotated traces, meaning that an expert has manually provided the step localizations on several traces. We show in simulated experiments that *SuperSA* better copes with noise than *QuickPBSA*. On a real task, we demonstrate that our method only needs a dozen labels to find steps in hundreds of traces correctly.

II. METHOD

This section describes our step detection algorithm and a supervised strategy to learn the optimal algorithm’s parameters. Here, we assume several traces have been extracted from an image stack.

A. Step detection

Let $y = (y_1, \dots, y_T)$ be a noisy observed signal with T samples. Denote by $\bar{y}_{a..b}$ the empirical average of the sub-signal $(y_a, y_{a+1}, \dots, y_{b-1})$ for two indexes $a < b$, i.e. $\bar{y}_{a..b} = 1/(b-a) \sum_{t=a}^{b-1} y_t$.

We assume that the signal y is well approximated by a piecewise constant signal with few shifts in value. Our objective is to recover the true but unknown step localizations, the indexes where there is a shift. Note that the step number is also unknown. Let $\mathcal{T} = \{t_1, t_2, \dots, t_K\}$ be a set of K potential step indexes. Its cardinal is denoted $|\mathcal{T}|$, i.e. $|\mathcal{T}| = K$. A natural measure of the approximation quality of the t_k is

$$V(\mathcal{T}, y) = \sum_{k=0}^K \sum_{t=t_k}^{t_{k+1}-1} (y_t - \bar{y}_{t_k..t_{k+1}})^2. \quad (1)$$

(By convention, $t_0 = 1$ and $t_{K+1} = T + 1$ are dummy indices.) This quantity is the squared error of the best piecewise constant approximation of y that only has steps at the positions t_k . Without constraint on the step number, one would choose the trivial approximation with T steps to have an error equal to 0. For this reason, we do not minimize $V(\mathcal{T})$ but rather a trade-off between data fidelity –measured by the squared error $V(\mathcal{T})$ – and segmentation complexity –measured by the number of steps $|\mathcal{T}|$. Formally, this yields the following optimization problem:

$$\begin{aligned} & \text{minimize } V_\beta(\mathcal{T}, y) \\ & \text{over all step sets } \mathcal{T} = \{t_1, t_2, \dots\} \end{aligned} \quad (2)$$

where $V_\beta(\mathcal{T}, y) = V(\mathcal{T}, y) + \beta|\mathcal{T}|$ and $\beta > 0$ is a user-defined parameter that controls the trade-off between data fidelity and segmentation complexity. The parameter β is often referred to as a penalty. Too large a β will produce an approximation with very few steps and potentially miss true ones; the signal

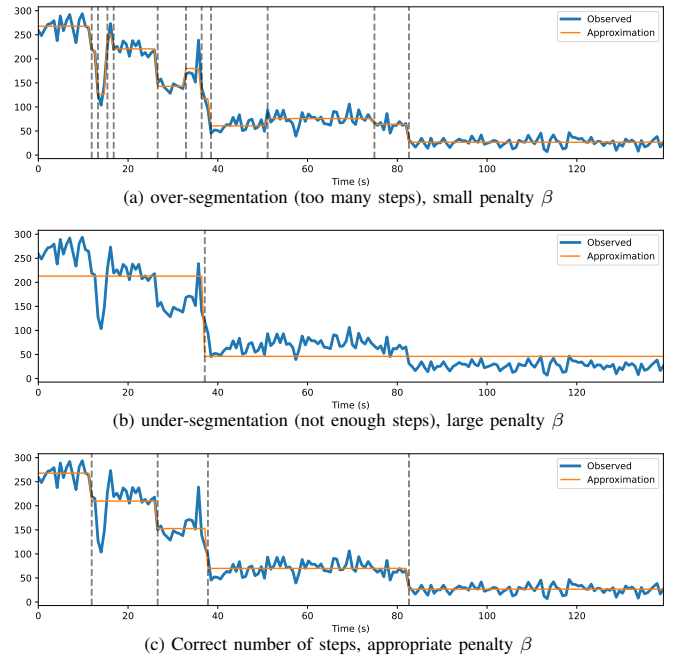


Fig. 2. Influence of the penalty value on the step detection procedure. All three plots have the same observed signal (in arbitrary units). Vertical dashed lines indicate the step localizations.

y is under-segmented. Conversely, too small a β will result in many detected steps, including false ones; the signal y is said to be over-segmented. The influence of the penalty β is illustrated in Fig. 5. This parameter is crucial to filter out noise and only keep meaningful steps. A procedure to calibrate β is described in the next section.

For a fixed β , there remains the question of solving the optimization problem (2). Such problems have been studied extensively in statistics and signal processing and are part of a more general setting called change-point detection or signal segmentation [8]. In this work, we use the efficient implementation from the *ruptures* package¹. The inputs of this algorithm are a signal y and a penalty $\beta > 0$; it outputs the solution of (2), that is the step set $\hat{\mathcal{T}} = \{\hat{t}_1, \hat{t}_2, \dots\}$ which minimizes the objective function. The number of steps is $\hat{K} = |\hat{\mathcal{T}}|$. Note that the estimated step number and localizations depend on the signal y and the penalty β .

B. Finding the optimal penalty

Finding a reasonable penalty β is an open problem that receives a lot of attention [8]. In most situations, when the data do not fit a well-known model, this parameter must be calibrated manually by trial and error. We aim to replace this fastidious and cumbersome process with a fully automatic procedure. To that end, we adopt a supervised approach, meaning that we assume that an expert has provided L signals $y^{(l)}$ ($l = 1, \dots, L$) along with their step localizations $\mathcal{T}^{(l)} = \{t_1^{(l)}, t_2^{(l)}, \dots\}$. The set of annotated signals $y^{(l)}$ and their annotations $\mathcal{T}^{(l)}$ is called a training set. Obtaining a

¹github.com/deepcharles/ruptures

small training set is no more burdensome than estimating the average step height, as in QuickPBSA. Our approach SuperSA aims at finding a penalty β that can reproduce the segmentation strategy of the expert. Intuitively, for an ideal β , the step set $\hat{\mathcal{T}}^{(l)}$ estimated on signal $y^{(l)}$ is close to the labelled step set $\mathcal{T}^{(l)}$, i.e. $\mathcal{T}^{(l)} \approx \hat{\mathcal{T}}^{(l)}$ for all l . Since there is no easy method to find such a penalty β , we relax our objective: instead, we want our estimation to match the same data fidelity/segmentation complexity as the labels, namely $V_\beta(\mathcal{T}^{(l)}, y^{(l)}) \approx V_\beta(\hat{\mathcal{T}}^{(l)}, y^{(l)})$. This leads to an optimization problem:

$$\begin{aligned} & \text{minimize} \quad \sum_{l=1}^L |V_\beta(\mathcal{T}^{(l)}, y^{(l)}) - V_\beta(\hat{\mathcal{T}}^{(l)}, y^{(l)})| \\ & \text{over } \beta > 0 \end{aligned} \quad (3)$$

Note that, by design, the objective function in (3) is non-negative and equal to 0 if there exists a penalty β such that $\hat{\mathcal{T}}_\beta = \mathcal{T}^{(l)}$ for all l . A method to solve this optimization problem has first been described in [7]. The solution to (3) is a penalty value $\hat{\beta}_{\text{opt}}$ that can be used to detect steps on new signals outside of the training set.

III. EXPERIMENTAL RESULTS

We study the performance of our approach SuperSA on simulated data and image stacks from a real experiment. Our algorithm is compared to the most recent and state-of-the-art automatic method for step analysis, QuickPBSA [3].

A. Synthetic simulations

Setting. We generate fluorescence intensity curves over 1500 images, representative of the emission of a few single fluorophores, as follows. Each fluorophore emits a signal with an average value μ_f that follows a Gaussian distribution $\mathcal{N}(0.15, \sigma_f)$. The fluorescence signal takes on a step-wise pattern depending on the number of emitters fluorescing simultaneously. The positions of these steps are random, following a Poisson distribution, resulting in signals that typically feature an average of 15 steps. Finally, Gaussian noise is incorporated into the signal with a standard deviation σ_n to take into account the photon noise. Four data sets are generated, each having different values for σ_f and σ_n (see Table I). We use precision, recall and F-score to measure the performance of our step detection analysis. A step is considered localized if an estimated step is at the same location. Precision is the ratio of the number of correctly localized steps (true positives) and the number of predicted steps; recall is the ratio of the number of true positives and the number of true steps; F-score is the geometric mean of precision and recall.

Results. The performance of SuperSA and QuickPBSA are shown in Table II. Two observations can be made:

- Both methods have very good scores on Data Set 1, which has the lowest noise variance and amplitude variance.

TABLE I
PARAMETERS OF THE SYNTHETIC DATA SETS: σ_f IS THE STEP AMPLITUDE VARIANCE AND σ_n IS THE NOISE VARIANCE.

Data set name	σ_f	σ_n
DATA SET 1	0.05	0.03
DATA SET 2	0.05	0.05
DATA SET 3	0.10	0.05
DATA SET 4	0.07	0.07

TABLE II
PERFORMANCE ON THE SYNTHETIC DATA SETS (BEST SCORES IN BOLD)

Data set	Algorithm	Recall (%)	Precision (%)	F-score (%)
DATA SET 1	SuperSA	90.1	94.8	92.4
	QuickPBSA	84.0	78.5	81.2
DATA SET 2	SuperSA	72.2	81.9	76.7
	QuickPBSA	70.3	62.3	66.1
DATA SET 3	SuperSA	67.4	84.1	74.8
	QuickPBSA	64.8	62.2	63.5
DATA SET 4	SuperSA	52.9	69.9	60.2
	QuickPBSA	53.8	51.8	52.8

The F-score deteriorates when noise increases, and step amplitude varies more.

- SuperSA has a better F-score than QuickPBSA for all data scenarios. For Data Set 4, which has high noise variance and high amplitude variance, QuickPBSA has a slightly higher recall but worse precision. On all other data sets, SuperSA has better precision and recall.

B. Study of the translation speed of a single protein

We use a setting similar to [2], where a reporter assay is developed to time an assembly of single ribosomes during translation. As mentioned in the introduction, the speed of the ribosome is deduced from the measurement of the time at which it detaches different fluorescent oligonucleotides hybridized at specific locations on the messenger RNA. To estimate those specific moments, we detect abrupt steps in the mean value of the fluorescence signals with our procedure SuperSA.

Experimental setup. It consists of a total internal reflection microscope (TIRFM): 2 laser beams (resp. 561 nm and 640 nm) are expanded and combined. They are focused off-optical axis in the back focal plane of an oil-immersion large numerical-aperture microscope objective, and totally reflected

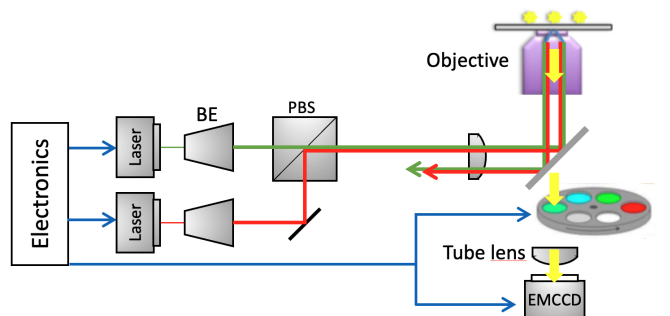


Fig. 3. Experimental total-internal-reflection fluorescence microscopy setup for studying the translation speed of a single protein, see Section III-B. BE is a beam expander, and PBS is a polarizing beam splitter.

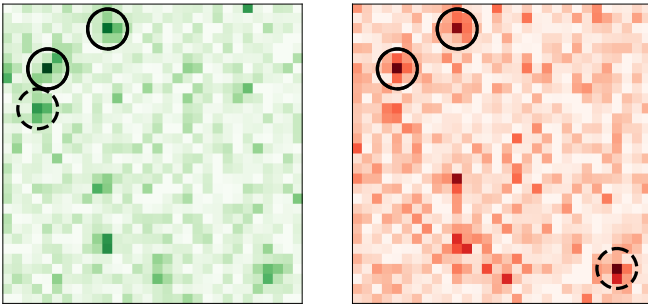


Fig. 4. Illustration of the preprocessing procedure. Peaks are detected on each frame (green and red) with a rough algorithm (peak detection step). Only those who appear in both frames are kept (solid black circles); the others (dashed black circles) are discarded (colocalization step).

at the glass-water interface of the coverslip. Two fluorescent signals are collected (one in green, one in red) by the objective and imaged by the tube lens on an EMCCD sensor after passing through the appropriate emission filter. Home-made electronics synchronise the different elements for data acquisition. A schematic view is shown in Figure 3.

Preprocessing. The background is removed on each frame using a rolling ball algorithm (radius: 50 pixels). Then we detect bright spot positions as follows: we pick one frame, smooth it with a Gaussian filter (width: 1 pixel) and find all local maxima above the background noise level. This crude peak detection procedure is refined by a second step where we filter out spots that only appear in the red video or the green video. We allow a distance of 2 pixels between matched spots. See Fig. 4 for an illustration. For each spot, the signal is the average of the surrounding pixels (3x3 square) across all frames. After this preprocessing, 1524 traces are extracted in each colour (2 × 1524 traces in total), and Fifty traces have been manually labelled in each colour to learn the penalty value.

Results. Using SuperSA, we learn two penalty values, one for each colour and then perform step analysis. For our application, we only keep traces with a single downward step in both colours. This allows us to compute the time difference between the green and red steps, as in Figure 1. The histogram of the time differences is shown in Figure 5. From this, we can estimate that a ribosome takes 74 seconds on average to go from the red marker to the green marker. Note that this is a rough estimate of the translation time, as a more elaborate statistical modelization is needed to account for particles that spontaneously stop emitting without any intervention of the ribosome.

As mentioned, this result is obtained using 100 annotated traces (50 in each colour). To show that we can cope with fewer labels, we randomly draw 20 annotated traces (10 in each colour) and execute SuperSA a second time, using only the smaller training set. This operation is repeated 100 times. When we compare the steps obtained with the complete training set and the ones obtained with the smaller set, 94% of the time, the same number of steps is detected, and when that happens, 97% of step localizations are the same for both

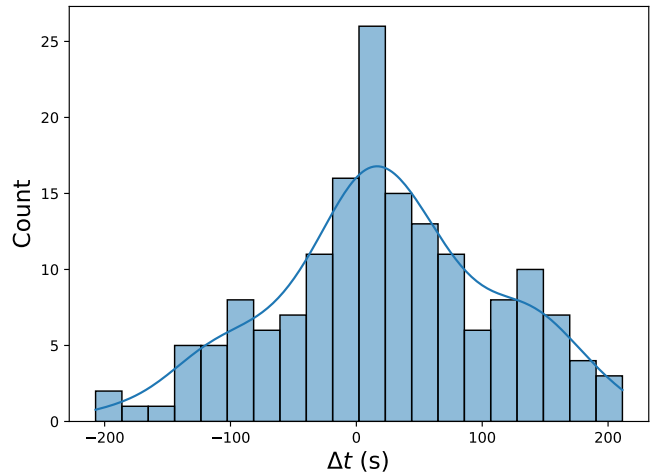


Fig. 5. Density estimation of green step times minus the red step times training sets. This supports the fact that even with limited annotations, SuperSA still produces the same results.

On a final note, we point out that the proposed pipeline takes only 14.5 seconds to process a stack of 300 512×512 images on a standard computer (Intel(R) Core(TM) i5-1030NG7 CPU @ 1.10GHz); most of the computation time (13 seconds) is devoted to extracting the traces and the remaining time (1.5 seconds) is spent on SuperSA. This low computational cost makes our approach suitable for interactive use.

IV. CONCLUSION

In this work, we have introduced an automatic and calibration-free photobleaching step detection procedure for fluorescent microscopy. It is supervised, meaning that it requires annotated traces from an expert. Experiments on a real application have shown that the number of labelled signals can be in the dozens to detect steps in hundreds of new traces. Furthermore, our method is fast and can be used interactively.

REFERENCES

- [1] J. S. Bryan IV, I. Sgouralis, and S. Pressé. Diffraction-limited molecular cluster quantification with Bayesian nonparametrics. *Nature Computational Science*, 2(2):102–111, 2022.
- [2] O. Bugaud, N. Barbier, H. Chommy, N. Fiszman, A. Le Gall, D. Dulin, M. Saguy, N. Westbrook, K. Perronet, and O. Namy. Kinetics of CrPV and HCV IRES-mediated eukaryotic translation using single-molecule fluorescence microscopy. *RNA*, 23(11):1626–1635, 2017.
- [3] J. Hummert, K. Yserentant, T. Fink, J. Euchner, Y. X. Ho, S. A. Tashev, and D. P. Herten. Photobleaching step analysis for robust determination of protein complex stoichiometries. *Molecular Biology of the Cell (MBoC)*, 32(21):1–12, 2021.
- [4] R. Kolar, L. Chmelikova, T. Vicar, and I. Provaznik. Increasing the Image Contrast via Fast Fluorescence Photobleaching. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2741–2744, Virtual Conference, 2021.
- [5] J. N. Milstein, D. F. Nino, X. Zhou, and C. C. Gradinaru. Single-molecule counting applied to the study of GPCR oligomerization. *Biophysical Journal*, 121(17):3175–3187, 2022.
- [6] Sviatlana Shashkova and Marc C. Leake. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience Reports*, 37(4):BSR20170031, 2017.
- [7] C. Truong, L. Oudre, and N. Vayatis. Penalty learning for changepoint detection. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pages 1569–1573, Kos, Greece, 2017.
- [8] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline changepoint detection methods. *Signal Processing*, 167, 2020.