



HAL
open science

COLORME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation

Vasiliki Stergiopoulou, Luca Calatroni, José Henrique de M Goulart,
Sébastien Schaub, Laure Blanc-Féraud

► **To cite this version:**

Vasiliki Stergiopoulou, Luca Calatroni, José Henrique de M Goulart, Sébastien Schaub, Laure Blanc-Féraud. COLORME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation. *Biological Imaging*, 2022, 2, pp.e1. 10.1017/S2633903X22000010 . hal-03320950v2

HAL Id: hal-03320950

<https://hal.science/hal-03320950v2>


Submitted on 12 Mar 2022

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.

RESEARCH ARTICLE

COLORME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation

Vasiliki Stergiopoulou^{1*} , Luca Calatroni¹, José Henrique de Morais Goulart², Sébastien Schaub³ and Laure Blanc-Féraud¹

¹Université Côte d'Azur, CNRS, INRIA, I3S, Sophia Antipolis, 06900, France

²IRIT, Université de Toulouse, Toulouse INP, Toulouse, 31071, France

³Sorbonne Université, CNRS, LBDV, Villefranche-sur-Mer, 06230, France

*Corresponding author. Email: vasiliki.stergiopoulou@i3s.unice.fr

Received: 02 Aug 2021; **Revised:** 14 Dec 2021; **Accepted:** 27 Jan 2022

Keywords: Super-Resolution, Fluorescence microscopy, Sparse Optimization, SOFI method

Abstract

To overcome the physical barriers caused by light diffraction, super-resolution techniques are often applied in fluorescence microscopy. State-of-the-art approaches require specific and often demanding acquisition conditions to achieve adequate levels of both spatial and temporal resolution. Analyzing the stochastic fluctuations of the fluorescent molecules provides a solution to the aforementioned limitations, as sufficiently high spatio-temporal resolution for live-cell imaging can be achieved by using common microscopes and conventional fluorescent dyes. Based on this idea, we present COLORME, a method for COvariance-based ℓ_0 super-Resolution Microscopy with intensity Estimation, which achieves good spatio-temporal resolution by solving a sparse optimization problem in the covariance domain and discuss automatic parameter selection strategies. The method is composed of two steps: the former where both the emitters' independence and the sparse distribution of the fluorescent molecules are exploited to provide an accurate localization; the latter where real intensity values are estimated given the computed support. The paper is furnished with several numerical results both on synthetic and real fluorescence microscopy images and several comparisons with state-of-the-art approaches are provided. Our results show that COLORME outperforms competing methods exploiting analogously temporal fluctuations; in particular, it achieves better localization, reduces background artifacts and avoids fine parameter tuning.

Impact Statement

This research paper describes a super-resolution method improving the spatial resolution of images acquired by common fluorescence microscopes and conventional blinking/fluctuating fluorophores. The problem is formulated in terms of a sparse and convex/non-convex optimization problem in the covariance domain for which a well-detailed algorithmic and numerical description are provided. It is addressed to an audience working at the interface between applied mathematics and biological image analysis. The proposed approach is validated on several synthetic datasets and shows promising results also when applied to real data, thus paving the way for new future research directions.

1. Introduction

In the field of fluorescence (or, more generally, light) microscopy, the main factor characterizing the microscope resolution is the limit imposed by the diffraction of light: structures with size smaller than the diffraction barrier (typically around 250nm in the lateral direction) cannot be well distinguished nor localized. The need to investigate small sub-cellular entities thus led to the implementation of a plethora of super-resolution methods.

A large and powerful family of imaging techniques achieving nanometric resolution are the ones often known as Single Molecule Localization Microscopy (SMLM) techniques, see, e.g.^(1,2) for a review. Among them, methods such as Photo-Activated Localization Microscopy (PALM)⁽³⁾ and STochastic Optical Reconstruction Microscopy (STORM)⁽⁴⁾ are designed so as to create a super-resolved image (achieving around 20nm of resolution) by activating and precisely localizing only a few molecules in each of thousands of acquired frames at a time. For their use, these methods need specific photoactivatable, photoswitchable, and binding-activated fluorophores, among others⁽⁵⁾, as well as, a large number (typically thousands) of sparse acquired frames leading to a poor temporal resolution and large exposure times which can significantly damage the sample. A different technique improving spatial resolution is well-known under the name of STimulated Emission Depletion (STED) microscopy⁽⁶⁾. Similarly to SMLM, STED techniques are based on a time-consuming and possibly harmful acquisition procedure requiring special equipment. In STED microscopy, the size of the point spread function (PSF) is reduced as a depletion beam of light will induce stimulated emission from molecules outside the region of interest and thus switch them off. Structured Illumination Microscopy (SIM)⁽⁷⁾ methods use patterned illumination to excite the sample; differently from the aforementioned approaches, images here can be recovered with high temporal-resolution via high speed acquisitions that cause comparatively little damage to the sample, but at the cost of a relatively low spatial resolution and, more importantly, the requirement of a specific illumination setup. Note that in this paper we address grid-based super-resolution approaches, i.e. the ones that formalize the super-resolution problem as the task of retrieving a well-detailed image on a fine grid from coarse measurements. More recently, off-the-grid super-resolution approaches have started to be studied in the literature, such as the one of Candès *et al.*⁽⁸⁾, with applications to SMLM data in Denoyelle *et al.*⁽⁹⁾, as well as DAOSTORM⁽¹⁰⁾, a high-density super-resolution microscopy algorithm. The great advantage of the gridless approaches is that there are no limitations imposed by the size of the discrete grid considered. However, both the theoretical study of the problem and its numerical realization become very hard due to the infinite-dimensional and typically non-convex nature of the optimization.

During the last decade, a new approach taking advantage of the independent stochastic temporal fluctuations/blinking of conventional fluorescent emitters appeared in the literature. A stack of images is acquired at a high temporal rate, typically 20 – 100 images/s, by means of common microscopes (such as widefield, confocal or Total Internal Reflection Fluorescence (TIRF) ones) using standard fluorophores, and then their independent fluctuations/blinking are exploited so as to compute a super-resolved image. Note that no specific material is needed here, neither for the illumination setup nor for fluorophores. Several methods exploiting the sequence of images have been proposed over the last years. Due to standard acquisition settings, temporal resolution properties are drastically improved. To start with, Super-resolution Optical Fluctuation Imaging (SOFI)⁽¹¹⁾ is a powerful technique where second and/or higher-order statistical analysis is performed, leading to a significant reduction of the size of the PSF. An extension of SOFI that combines several cumulant orders and achieves better resolution levels than SOFI is the method bSOFI⁽¹²⁾. However, spatial resolution still cannot reach the same levels of PALM/STORM. Almost the same behavior has been noticed in Super-Resolution Radial Fluctuations (SRRF)⁽¹³⁾ microscopy, where super-resolution is achieved by calculating the degree of local symmetry at each frame. Despite its easy manipulation and broad applicability, SRRF creates

significant reconstruction artifacts which may limit its use in view of accurate analysis. Other methods which belong to the same category and are worth mentioning are: the method 3B⁽¹⁴⁾, which uses Bayesian analysis and takes advantage of the blinking and bleaching events of standard fluorescent molecules, the method Entropy-based Super-resolution Imaging (ESI)⁽¹⁵⁾ that computes entropy values pixel-by-pixel, weighted with higher order statistics and the method Spatial COvariance REconstructive (SCORE)⁽¹⁶⁾ that analyzes intensity statistics, similarly to SOFI, but further reduces noise and computational cost by computing only a few components that have a significant contribution to the intensity variances of the pixels. In addition, the approach SPARsity-based super-resolution COrrrelation Microscopy (SPARCOM)^(17,18) exploits, as SOFI, both the lack of correlation between distinct emitters as well as the sparse distribution of the fluorescent molecules via the use of an ℓ_1 regularization defined on the emitters' covariance matrix. Along the same lines, a deep-learning method exploiting algorithmic unfolding, called Learned SPARCOM (LSPARCOM)⁽¹⁹⁾, has recently been introduced. Differently from plain SPARCOM, the advantage of LSPARCOM is that neither previous knowledge of the PSF nor any heuristic choice of the regularization parameter for tuning the sparsity level is required. As far as the reconstruction quality is concerned, both SPARCOM and LSPARCOM create some artifacts under challenging imaging conditions, for example when the noise and/or background level are relatively high. Finally, without using higher order statistics, a constrained tensor modeling approach that estimates a map of local molecule densities and their overall intensities, as well as, a matrix-based formulation that promotes structure sparsity via an ℓ_0 type regularizer, are available in⁽²⁰⁾. These approaches can achieve excellent temporal resolution levels, but the spatial resolution is limited.

Contribution. In this paper, we propose a method for live-cell super-resolution imaging based on the sparse analysis of the stochastic fluctuations of molecule intensities. The proposed approach provides a good level of both temporal and spatial resolution, thus allowing for both precise molecule localization and intensity estimation at the same time, while relaxing the need for special equipment (microscope, fluorescent dyes) typically encountered in state-of-the-art super-resolution methods such as, e.g., SMLM. The proposed method is called COLORME, which stands for COvariance-based super-Resolution Microscopy with intensity Estimation. Similarly to SPARCOM⁽¹⁸⁾, COLORME enforces signal sparsity in the covariance domain by means of sparsity-promoting terms, either of convex (ℓ_1 , TV) or non-convex (ℓ_0 -based)-type. Differently from SPARCOM, COLORME allows also for an accurate estimation of the noise variance in the data and is complemented with an automatic selection strategy of the model hyperparameters. Furthermore, and more importantly, COLORME allows for the estimation of both signal and background intensity, which are relevant pieces of information for biological studies. By exploiting information on the estimated noise statistics, the parameter selection in this step is also made fully automatic, based on the standard discrepancy principle. We remark that an earlier version of COLORME has been already introduced by the authors in⁽²¹⁾. Here, we consider an extended formulation combined with automatic parameter selection strategies which allows for the analysis of more challenging data having, e.g., spatially varying background. The method is validated on simulated and tested on challenging real data. Our results show that COLORME outperforms competing methods in terms of localization precision, parameter tuning and removal of background artifacts.

2. Mathematical Modeling

For real scalars $T, M > 1$ and $t \in \{1, 2, \dots, T\}$, let $\mathbf{Y}_t \in \mathbb{R}^{M \times M}$ be the blurred, noisy and down-sampled image frame acquired at time t . We look for a high-resolution image $\mathbf{X} \in \mathbb{R}^{L \times L}$ being defined as $\mathbf{X} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t$ with $L = qM$ and defined on a q -times finer grid, with $q \in \mathbb{N}$. Note that in the following applications we typically set $q = 4$. The image formation model describing the acquisition process at each t can be written as:

$$\mathbf{Y}_t = \mathcal{M}_q(\mathcal{H}(\mathbf{X}_t)) + \mathbf{B} + \mathbf{N}_t, \quad (1)$$

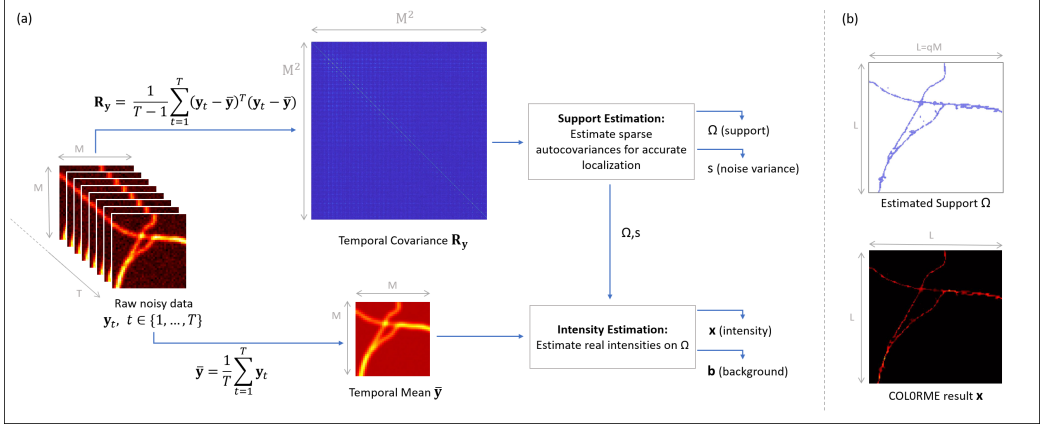


Figure 1. Principles of COLORME. (a) An overview of the two steps (Support Estimation and Intensity Estimation) by visualizing the inputs/outputs of each, as well as the interaction between them. (b) The two main outputs of COLORME are: the support $\Omega \subset \mathbb{R}^{L^2}$ containing the locations of the fine-grid pixels with at least one fluorescent molecule, and the intensity $\mathbf{x} \in \mathbb{R}^{L^2}$ whose non-null values are estimated only on Ω .

where $\mathcal{M}_q : \mathbb{R}^{L \times L} \rightarrow \mathbb{R}^{M \times M}$ is a down-sampling operator summing every q consecutive pixels in both dimensions, $\mathcal{H} : \mathbb{R}^{L \times L} \rightarrow \mathbb{R}^{L \times L}$ is a convolution operator defined by the PSF of the optical imaging system and $\mathbf{B} \in \mathbb{R}^{M \times M}$ models the background, which collects the contributions of the out-of-focus (and the ambient) fluorescent molecules. Motivated by experimental observations showing that the blinking/fluctuating behaviour of the out-of-focus molecules is not visible after convolution with wide de-focused PSFs, we assume that the background is temporally constant (\mathbf{B} does not depend on t), while we allow it to smoothly vary in space. Finally, $\mathbf{N}_t \in \mathbb{R}^{M \times M}$ describes the presence of noise modeled here as a matrix of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance $s \in \mathbb{R}^+$ taking into account both the underlying electronic noise and the noise bias induced by \mathbf{B} (see Remark 1 for more details on the approximation considered). We assume that the molecules are located at the center of each pixel and that there is no displacement of the specimen during the imaging period, which is a reasonable assumption whenever short time acquisitions are considered.

Remark 1. A more appropriate model taking also into account the presence of signal-dependent Poisson noise in the data would be the following:

$$\mathbf{Y}_t = P(\mathcal{M}_q(\mathcal{H}(\mathbf{X}_t)) + \mathbf{B}) + \mathbf{E}_t = P(\mathcal{M}_q(\mathcal{H}(\mathbf{X}_t))) + P(\mathbf{B}) + \mathbf{E}_t, \quad \forall t = 1, 2, \dots, T, \quad (2)$$

where, for $\mathbf{W} \in \mathbb{R}^{M \times M}$, $P(\mathbf{W})$ represents the realization of a multivariate Poisson variable of parameter \mathbf{W} and $\mathbf{E}_t \in \mathbb{R}^{M \times M}$ models electronic noise with a matrix of i.i.d. Gaussian entries of zero mean and constant variance $\sigma^2 \in \mathbb{R}_+$. Note that the second equality in (2) holds due to the independence between $\mathcal{M}_q(\mathcal{H}(\mathbf{X}_t))$ and \mathbf{B} . Model (2) is indeed the one we used for the generation of the simulated data, see Section 6.1. However, to simplify the reconstruction process, we simplified (2) by assuming that \mathbf{B} has sufficiently large entries, so that $P(\mathbf{B})$ can be approximated as $P(\mathbf{B}) \approx \hat{\mathbf{B}}$ with $\hat{\mathbf{B}}_{i,j} \sim \mathcal{N}(\mathbf{B}_{i,j}, \mathbf{B}_{i,j})$, where $(i, j) \in \{1, \dots, M\}^2$, thus considering:

$$\mathbf{Y}_t = P(\mathcal{M}_q(\mathcal{H}(\mathbf{X}_t))) + \hat{\mathbf{B}} + \mathbf{E}_t, \quad \forall t = 1, 2, \dots, T. \quad (3)$$

By now further approximating the variance of $\hat{\mathbf{B}}$ with a constant $b \in \mathbb{R}_+$ to be interpreted as the average of \mathbf{B} , we have that by simple manipulations:

$$\hat{\mathbf{B}} + \mathbf{E}_t = \mathbf{B} + \mathbf{N}_t,$$

where the independence between $\hat{\mathbf{B}}$ and \mathbf{E}_t has been exploited. We can thus retrieve (1) from (3) by neglecting the Poisson noise dependence in $P(\mathcal{M}_q(\mathcal{H}(\mathbf{X}_t)))$ and that the variance of every entry of the random term \mathbf{N}_t is $s = \sigma^2 + b$. A more detailed and less approximated modelling taking into account the signal-dependent nature of the noise in the data could represent a very interesting area of future research.

In vectorized form, model (1) reads:

$$\mathbf{y}_t = \Psi \mathbf{x}_t + \mathbf{b} + \mathbf{n}_t, \quad (4)$$

where $\Psi \in \mathbb{R}^{M^2 \times L^2}$ is the matrix representing the composition $\mathcal{M}_q \circ \mathcal{H}$, while $\mathbf{y}_t \in \mathbb{R}^{M^2}$, $\mathbf{x}_t \in \mathbb{R}^{L^2}$, $\mathbf{b} \in \mathbb{R}^{M^2}$ and $\mathbf{n}_t \in \mathbb{R}^{M^2}$ are the column-wise vectorizations of \mathbf{Y}_t , \mathbf{X}_t , \mathbf{B} and \mathbf{N}_t in (1), respectively.

For all t and given Ψ and \mathbf{y}_t , the problem can thus be formulated as

$$\text{find } \mathbf{x} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \in \mathbb{R}^{L^2}, \mathbf{b} \in \mathbb{R}^{M^2} \text{ and } s > 0 \text{ s.t. } \mathbf{x}_t \text{ solves (4).}$$

In order to exploit the statistical behavior of the fluorescent emitters, we reformulate the model in the covariance domain. This idea was previously exploited by the SOFI approach⁽¹¹⁾ and was shown to significantly reduce the full-width-at-half-maximum (FWHM) of the PSF. In particular, the use of second-order statistics for a Gaussian PSF corresponds to a reduction factor of the FWHM of $\sqrt{2}$.

To formulate the model, we consider the frames $(\mathbf{y}_t)_{t \in \{1, \dots, T\}}$ as T realizations of a random variable \mathbf{y} with covariance matrix defined by:

$$\mathbf{R}_y = \mathbb{E}_y \{ (\mathbf{y} - \mathbb{E}_y \{ \mathbf{y} \}) (\mathbf{y} - \mathbb{E}_y \{ \mathbf{y} \})^\top \}, \quad (5)$$

where $\mathbb{E}_y \{ \cdot \}$ denotes the expected value computed w.r.t. to the unknown law of \mathbf{y} . We estimate \mathbf{R}_y by computing the empirical covariance matrix, i.e.:

$$\mathbf{R}_y \approx \frac{1}{T-1} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}}) (\mathbf{y}_t - \bar{\mathbf{y}})^\top,$$

where $\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t$ denotes the empirical temporal mean. From (4) and (5), we thus deduce the relation:

$$\mathbf{R}_y = \Psi \mathbf{R}_x \Psi^\top + \mathbf{R}_n, \quad (6)$$

where $\mathbf{R}_x \in \mathbb{R}^{L^2 \times L^2}$ and $\mathbf{R}_n \in \mathbb{R}^{M^2 \times M^2}$ are the covariance matrices of $(\mathbf{x}_t)_{t \in \{1, \dots, T\}}$ and $(\mathbf{n}_t)_{t \in \{1, \dots, T\}}$, respectively. As the background is stationary by assumption, the covariance matrix of \mathbf{b} is zero. Recalling now that the emitters are uncorrelated by assumption, we deduce that \mathbf{R}_x is diagonal. We thus set $\mathbf{r}_x := \text{diag}(\mathbf{R}_x) \in \mathbb{R}^{L^2}$. Furthermore, by the i.i.d. assumption on \mathbf{n}_t , we have that $\mathbf{R}_n = s \mathbf{I}_{M^2}$, where $s \in \mathbb{R}_+$ and \mathbf{I}_{M^2} is the identity matrix in $\mathbb{R}^{M^2 \times M^2}$. Note that the model in equation (6) is similar to the SPARCOM one presented in⁽¹⁸⁾, with the difference that here we consider also noise contributions by including in the model the diagonal covariance matrix \mathbf{R}_n . Finally, the vectorized form of the model in the covariance domain can thus be written as:

$$\mathbf{r}_y = (\Psi \odot \Psi) \mathbf{r}_x + s \mathbf{v}_1,$$

where \odot denotes the Khatri–Rao (column-wise Kronecker) product, $\mathbf{r}_y \in \mathbb{R}^{M^2}$ is the column-wise vectorization of \mathbf{R}_y and $\mathbf{v}_I = \text{vec}(\mathbf{I}_{M^2})$.

3. COLORME, step I: support estimation for precise molecule localization

Similarly to SPARCOM⁽¹⁸⁾, our approach makes use of the fact that the solution \mathbf{r}_x is sparse, while including further the estimation of $s > 0$ for dealing with more challenging scenarios. In order to compare specific regularity *a-priori* constraints on the solution, we make use of different regularization terms, whose importance is controlled by a regularization hyperparameter $\lambda > 0$. By further introducing some non-negativity constraints for both variables \mathbf{r}_x and s , we thus aim to solve:

$$\arg \min_{\mathbf{r}_x \geq 0, s \geq 0} \mathcal{F}(\mathbf{r}_x, s) + \mathcal{R}(\mathbf{r}_x; \lambda), \quad (7)$$

where the data fidelity term is defined by:

$$\mathcal{F}(\mathbf{r}_x, s) = \frac{1}{2} \|\mathbf{r}_y - (\Psi \odot \Psi)\mathbf{r}_x - s\mathbf{v}_I\|_2^2, \quad (8)$$

and $\mathcal{R}(\cdot; \lambda)$ is a sparsity-promoting penalty. Ideally, one would like to make use of the ℓ_0 norm to enforce sparsity. However, as it is well-known, solving the resulting non-continuous, non-convex and combinatorial minimization problem is an NP-hard problem. A way to circumvent this difficulty consists in using the continuous exact relaxation of the ℓ_0 norm (CEL0) proposed by Soubies *et al.* in⁽²²⁾. The CEL0 regularization is continuous, non-convex and preserves the global minima of the original $\ell_2 - \ell_0$ problem while removing some local ones. It is defined as follows:

$$\mathcal{R}(\mathbf{r}_x; \lambda) = \Phi_{\text{CEL0}}(\mathbf{r}_x; \lambda) = \sum_{i=1}^{L^2} \lambda - \frac{\|\mathbf{a}_i\|^2}{2} \left(|(\mathbf{r}_x)_i| - \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|} \right)^2 \mathbb{1}_{\{|(\mathbf{r}_x)_i| \leq \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|}\}}, \quad (9)$$

where $\mathbf{a}_i = (\Psi \odot \Psi)_i$ denotes the i -th column of the operator $\mathbf{A} := \Psi \odot \Psi$.

A different, convex way of favoring sparsity consists in taking as regularizer the ℓ_1 norm, that is:

$$\mathcal{R}(\mathbf{r}_x; \lambda) = \lambda \|\mathbf{r}_x\|_1. \quad (10)$$

Besides convexity and as it is well-known, the key difference between using the ℓ_0 and the ℓ_1 -norm is that the ℓ_0 provides a correct interpretation of sparsity by counting only the number of the non-zero coefficients, while the ℓ_1 depends also on the magnitude of the coefficients. However, its use as a sparsity-promoting regularizer is nowadays well-established (see, e.g.,⁽²³⁾) and also used effectively in other microscopy applications, such as SPARCOM⁽¹⁸⁾.

Finally, in order to model situations where piece-wise constant structures are considered, we consider a different regularization term favoring gradient-sparsity by using the Total Variation (TV) regularization defined in a discrete setting as follows:

$$\mathcal{R}(\mathbf{r}_x; \lambda) = \lambda TV(\mathbf{r}_x) = \lambda \sum_{i=1}^{L^2} \left(|(\mathbf{r}_x)_i - (\mathbf{r}_x)_{n_{i,1}}|^2 + |(\mathbf{r}_x)_i - (\mathbf{r}_x)_{n_{i,2}}|^2 \right)^{\frac{1}{2}}, \quad (11)$$

where $(n_{i,1}, n_{i,2}) \in \{1, \dots, L^2\}^2$ indicate the locations of the horizontal and vertical nearest neighbor pixels of pixel i , as shown in Figure 2. For the computation of the TV penalty, Neumann boundary conditions have been used.

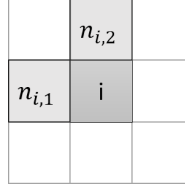


Figure 2. The one-sided nearest horizontal and vertical neighbors of the pixel i used to compute the gradient discretization in (11).

To solve (7) we use the Alternate Minimization algorithm between s and \mathbf{r}_x ⁽²⁴⁾, see the pseudo-code reported in Algorithm 1. Note that, at each $k \geq 1$, the update for the variable s can be efficiently computed through the following explicit expression:

$$s^{k+1} = \frac{1}{M^2} \mathbf{v}_1^\top (\mathbf{r}_y - (\Psi \odot \Psi) \mathbf{r}_x^k).$$

Concerning the update of \mathbf{r}_x , different algorithms were used depending on the choice of the regularization term in (9), (10) and (11). For the CEL0 penalty (9) we used the iteratively reweighted ℓ_1 algorithm (IRL1)⁽²⁵⁾, following Gazagnes et al.⁽²⁶⁾ with Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)⁽²⁷⁾ as inner solver. If the ℓ_1 norm (10) is chosen, FISTA is used. Finally, when the TV penalty (11) is employed, the Primal-Dual Splitting Method in⁽²⁸⁾ was considered.

Algorithm 1 COLORME, Step I: Support Estimation

Require: $\mathbf{r}_y \in \mathbb{R}^{M^4}$, $\mathbf{r}_x^0 \in \mathbb{R}^{L^2}$, $\lambda > 0$
repeat
 $s^{k+1} = \arg \min_{s \in \mathbb{R}_+} \mathcal{F}(\mathbf{r}_x^k, s)$
 $\mathbf{r}_x^{k+1} = \arg \min_{\mathbf{r}_x \in \mathbb{R}_+^{L^2}} \mathcal{F}(\mathbf{r}_x, s^{k+1}) + \mathcal{R}(\mathbf{r}_x; \lambda)$
until convergence
return Ω_x, s

Following the description provided by Attouch *et al.* in⁽²⁴⁾, convergence of Algorithm 1 can be guaranteed only if an additional quadratic term is introduced in the objective function of the second minimization sub-problem. Nonetheless, empirical convergence was observed also without such additional terms.

To evaluate the performance of the first step of the method COLORME using the different regularization penalties described above, we created two noisy simulated datasets, with low background (LB) and high background (HB), respectively and used them to apply COLORME and estimate the desired sample support. More details on the two datasets are available in the following sub-section 6.1. The results obtained by using the three different regularizers are reported in Figure 3. In this example we chose the regularization parameter λ heuristically, while more details about the selection of the parameter are given in the subsection 5.1.

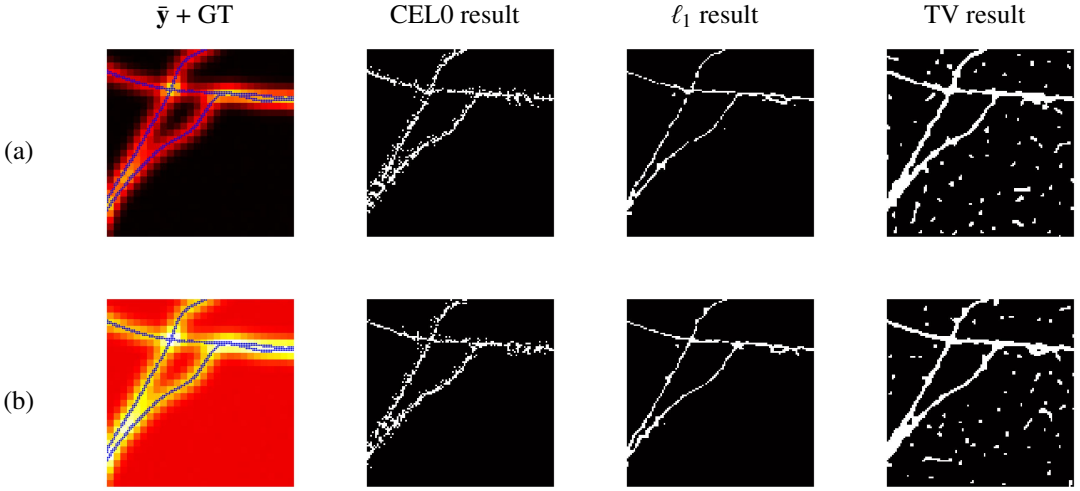


Figure 3. (a) Noisy simulated dataset with low-background (LB) and stack size: $T = 500$ frames, (b) Noisy simulated high-background (HB) dataset, with $T = 500$ frames. From left to right: Superimposed diffraction limited image (temporal mean of the stack) with 4x zoom on ground truth support (blue), CEL0 reconstruction, ℓ_1 reconstruction and TV reconstruction.

Despite its continuous and smooth reconstruction, we observe that the reconstruction obtained by the TV regularizer does not provide precise localization results. For example, the separation of the two filaments on the top-right corner is not visible and while the junction of the other two filaments on the bottom-left should appear further down, we clearly see that those filaments are erroneously glued together. Nonetheless, the choice of an appropriate regularizer tailored to favor fine structures as the ones observed in the GT image constitutes a challenging problem that should be addressed in future research.

The Jaccard indices (JI) of both the results obtained when using the CEL0 and ℓ_1 regularizer, that allow for more precise localization, have been computed. The Jaccard index, is a quantity in the range $[0, 1]$ computed as the ratio between correct detections (CD) and the sum of correct detections, false positives (FP) and false negatives (FN), that is $JI := CD / (CD + FN + FP)$, up to a tolerance $\delta > 0$, measure in nm. A correct detection occurs when one pixel at most δ nm away from a ground truth pixel is added to the support. In order to match the pixels from the estimated support to the ones from the ground truth, we employ the standard Gale–Shapley algorithm⁽²⁹⁾. Once the matching has been performed, we can simply count the number of ground truth pixels which have not been detected (false negatives) and also the number of pixels in the estimated support which have not been matched to any ground truth pixel (false positives).

The Figure 4 reports the average Jaccard index computed from 20 different noise realizations, as well as, an error bar (vertical lines) that represent the standard deviation, for several stack sizes. According to the figure, a slightly better Jaccard index is obtained when the CEL0 regularizer is being used, while an increase in the number of frames, when both regularizers being used, leads to better Jaccard index, hence better localization. As the reader may notice, such quantitative assessment could look inconsistent with the visual results reported in Figure 3. By definition, the JI tends to assume higher values whenever more CD are found even in presence of more FP (as it happens for the CEL0 reconstruction), while it gets more penalized when FN happen, as they affect the computation "twice", reducing the numerator and increasing the denominator.

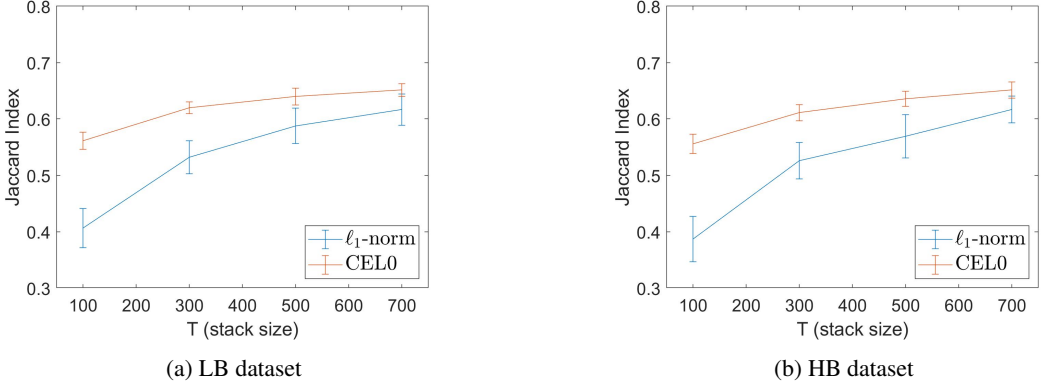


Figure 4. Jaccard Index values with tolerance $\delta = 40\text{nm}$ for the low-background (LB) and high-background (HB) dataset, for different stack sizes and regularization penalty choices. The tolerance, $\delta = 40\text{ nm}$, is set so that we allow the correct detections, that needed to be counted for the computation of the Jaccard Index, to be found not only in the same pixel but also to any of the 8-neighbouring pixels.

3.1. Accurate noise variance estimation

Along with the estimations of the emitter’s temporal sparse covariance matrix, the estimation of the noise variance in the joint model (7) allows for much more precise results even in challenging acquisition conditions. In Figure 5 we show the relative error between the computed noise variance s and the constant variance of the electronic noise σ^2 used to produce simulated low-background (LB) and high-background (HB) data. The relative error is higher in the case of the HB dataset, something that is expected, as in our noise variance estimation s there is a bias coming from the background (see Remark 1). In the case of the LB dataset, as the background is low, the bias is sufficiently small so that it is barely visible in the error graph. In our experiments, a Gaussian noise with a corresponding SNR of approximately 16 dB is being used, while the value of σ^2 is in average equal to 7.11×10^5 for the LB dataset and 7.13×10^5 for the HB dataset. Note that, in general, the estimation of the noise variance s obtained by COLORME is very precise.

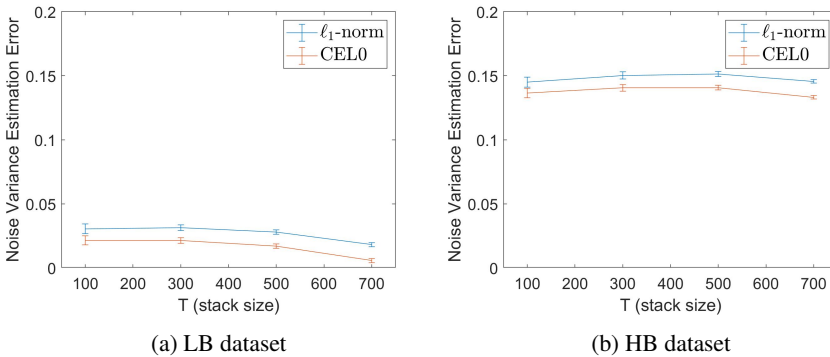


Figure 5. The relative error in noise variance estimation, defined as: $\text{Error} = \frac{|s - \sigma^2|}{|\sigma^2|}$, where σ^2 is the constant variance of the electronic noise. The Error is computed for 20 different noise realizations, presenting in the graph the mean and the standard deviation (error bars).

4. COLORME, step II: Intensity estimation

From the previous step, we obtain a sparse estimation of $\mathbf{r}_\mathbf{x} \in \mathbb{R}^{L^2}$. Its support, i.e. the location of non-zero variances, can thus be deduced. This is denoted in the following by $\Omega := \{i : (\mathbf{r}_\mathbf{x})_i \neq 0\} \subset \{1, \dots, L^2\}$. Note that this set corresponds indeed to the support of the desired \mathbf{x} , hence in the following we will use the same notation to denote both sets.

We are now interested in enriching COLORME with an additional step where intensity information of the signal \mathbf{x} can be retrieved in correspondence with the estimated support Ω . To do so, we thus propose an intensity estimation procedure for \mathbf{x} restricted only to the pixels of interest. Under this modeling assumption, it is thus reasonable to consider a regularization term favoring smooth intensities on Ω , in agreement to the intensity typically found in real images.

In order to take into account the modeling of blurry and out-of-focus fluorescent molecules, we further include in our model (4) a regularization term for smooth background estimation. We can thus consider the following joint minimization problem:

$$\arg \min_{\mathbf{x} \in \mathbb{R}_+^{|\Omega|}, \mathbf{b} \in \mathbb{R}_+^{M^2}} \frac{1}{2} \|\Psi_\Omega \mathbf{x} - (\bar{\mathbf{y}} - \mathbf{b})\|_2^2 + \frac{\mu}{2} \|\nabla_\Omega \mathbf{x}\|_2^2 + \frac{\beta}{2} \|\nabla \mathbf{b}\|_2^2, \quad (12)$$

where the data term models the presence of Gaussian noise, $\mu, \beta > 0$ are regularization parameters and the operator $\Psi_\Omega \in \mathbb{R}^{M^2 \times |\Omega|}$ is a matrix whose i -th column is extracted from Ψ for all indexes $i \in \Omega$. Finally, the regularization term on \mathbf{x} is the squared norm of the discrete gradient restricted to Ω , i.e.:

$$\|\nabla_\Omega \mathbf{x}\|_2^2 := \sum_{i \in \Omega} \sum_{j \in \mathcal{N}(i) \cap \Omega} (x_i - x_j)^2,$$

where $\mathcal{N}(i)$ denotes the 8-pixel neighborhood of $i \in \Omega$. Note that, according to this definition, $\nabla_\Omega \mathbf{x}$ denotes a (redundant) isotropic discretization of the gradient of \mathbf{x} evaluated for each pixel in the support Ω . Note that this definition coincides with the standard one for $\nabla \mathbf{x}$ restricted to points in the support Ω .

The non-negativity constraints on \mathbf{x} and \mathbf{b} as well as the one restricting the estimation of \mathbf{x} on Ω can be relaxed by using suitable smooth penalty terms, so that, finally, the following optimization problem can be addressed:

$$\arg \min_{\mathbf{x} \in \mathbb{R}^{L^2}, \mathbf{b} \in \mathbb{R}^{M^2}} \frac{1}{2} \|\Psi \mathbf{x} - (\bar{\mathbf{y}} - \mathbf{b})\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x}\|_2^2 + \frac{\beta}{2} \|\nabla \mathbf{b}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \mathbf{x}\|_2^2 + \sum_{i=1}^{L^2} [\phi(x_i)]^2 + \sum_{i=1}^{M^2} [\phi(\mathbf{b}_i)]^2 \right), \quad (13)$$

where the parameter $\alpha \gg 1$ can be chosen arbitrarily high to enforce the constraints, \mathbf{I}_Ω is a diagonal matrix acting as characteristic function of Ω , i.e. defined as:

$$\mathbf{I}_\Omega(i, i) = \begin{cases} 0 & \text{if } i \in \Omega, \\ 1 & \text{if } i \notin \Omega \end{cases}, \quad \forall i \in \{1, \dots, L^2\},$$

and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is used to penalize negative entries, being defined as:

$$\phi(z) := \begin{cases} 0 & \text{if } z \geq 0, \\ z & \text{if } z < 0 \end{cases}, \quad \forall z \in \mathbb{R}. \quad (14)$$

We anticipate here that considering the unconstrained problem (13) instead of the original, constrained, one (12), will come in handy for the design of an automatic parameter selection strategy, as we further detail in Section 5.2.

To solve the joint-minimization problem (13) we use the Alternate Minimization algorithm, see Algorithm 2. In the following subsections, we provide more details on the solution of the two minimization sub-problems.

Algorithm 2 COLORME, Step II: Intensity Estimation

Require: $\bar{\mathbf{y}} \in \mathbb{R}^{M^2}$, $\mathbf{x}^0 \in \mathbb{R}^{L^2}$, $\mathbf{b}^0 \in \mathbb{R}^{M^2}$, $\mu, \beta > 0$, $\alpha \gg 1$

repeat

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^{L^2}} \frac{1}{2} \|\Psi \mathbf{x} - (\bar{\mathbf{y}} - \mathbf{b}^k)\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I} \Omega \mathbf{x}\|_2^2 + \sum_{i=1}^{L^2} [\phi(\mathbf{x}_i)]^2 \right)$$

$$\mathbf{b}^{k+1} = \arg \min_{\mathbf{b} \in \mathbb{R}^{M^2}} \frac{1}{2} \|\mathbf{b} - (\bar{\mathbf{y}} - \Psi \mathbf{x}^{k+1})\|_2^2 + \frac{\beta}{2} \|\nabla \mathbf{b}\|_2^2 + \frac{\alpha}{2} \sum_{i=1}^{M^2} [\phi(\mathbf{b}_i)]^2$$

until convergence

return \mathbf{x}, \mathbf{b}

4.1. First sub-problem: update of \mathbf{x}

In order to find at each $k \geq 1$ the optimal solution $\mathbf{x}^{k+1} \in \mathbb{R}^{L^2}$ for the first sub-problem, we need to solve a minimization problem of the form:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^{L^2}} g(\mathbf{x}; \mathbf{b}^k) + h(\mathbf{x}), \quad (15)$$

where, for $\mathbf{b}^k \in \mathbb{R}^{M^2}$ being fixed at each iteration $k \geq 1$, $g(\cdot; \mathbf{b}^k) : \mathbb{R}^{M^2} \rightarrow \mathbb{R}_+$ is a proper and convex function with Lipschitz gradient, defined as:

$$g(\mathbf{x}; \mathbf{b}^k) := \frac{1}{2} \|\Psi \mathbf{x} - (\bar{\mathbf{y}} - \mathbf{b}^k)\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x}\|_2^2, \quad (16)$$

and where the function $h : \mathbb{R}^{L^2} \rightarrow \mathbb{R}$ encodes the penalty terms:

$$h(\mathbf{x}) = \frac{\alpha}{2} \left(\|\mathbf{I} \Omega \mathbf{x}\|_2^2 + \sum_{i=1}^{L^2} [\phi(\mathbf{x}_i)]^2 \right). \quad (17)$$

Solution of (15) can be obtained iteratively, using, for instance, the proximal gradient descent algorithm, whose iteration can be defined as follows :

$$\mathbf{x}^{n+1} = \mathbf{prox}_{h, \tau}(\mathbf{x}^n - \tau \nabla g(\mathbf{x}^n)), \quad n = 1, 2, \dots, \quad (18)$$

where $\nabla g(\cdot)$ denotes the gradient of g , $\tau \in (0, \frac{1}{L_g}]$ is the algorithmic step-size chosen inside a range depending on the Lipschitz constant of ∇g , here denoted by L_g , to guarantee convergence. The proximal update in (18) can be computed explicitly using the computations reported in Appendix A. One can show in fact that, for each $\mathbf{w} \in \mathbb{R}^{L^2}$ there holds element-wise:

$$(\mathbf{prox}_{h, \tau}(\mathbf{w}))_i = \mathbf{prox}_{h, \tau}(\mathbf{w}_i) = \begin{cases} \frac{\mathbf{w}_i}{1 + \alpha \tau \mathbf{I} \Omega(i, i)} & \text{if } \mathbf{w}_i \geq 0, \\ \frac{\mathbf{w}_i}{1 + \alpha \tau (\mathbf{I} \Omega(i, i) + 1)} & \text{if } \mathbf{w}_i < 0. \end{cases} \quad (19)$$

Remark 2. As the reader may have noted, we consider the proximal gradient descent algorithm (18) for solving (15), even though both functions g and h in (16) and (17) respectively, are smooth and convex, hence, in principle, (accelerated) gradient descent algorithms could be used. Note, however, that the presence of the large penalty parameter $\alpha \gg 1$ would significantly slow down convergence speed in such case as the step size τ in this case would be constrained to the smaller range $(0, \frac{1}{L_g + \alpha}]$. By considering the penalty contributions in terms of their proximal operators, this limitation doesn't affect the range of τ and convergence is still guaranteed⁽³⁰⁾ in a computationally fast way through the update (19).

4.2. Second sub-problem: update of \mathbf{b}

As far as the estimation of the background is concerned, the minimization problem we aim to solve at each $k \geq 1$ takes the form:

$$\mathbf{b}^{k+1} = \arg \min_{\mathbf{b} \in \mathbb{R}^{M^2}} r(\mathbf{b}; \mathbf{x}^{k+1}) + q(\mathbf{b}), \quad (20)$$

where:

$$r(\mathbf{b}; \mathbf{x}^{k+1}) := \frac{1}{2} \|\mathbf{b} - (\bar{\mathbf{y}} - \Psi \mathbf{x}^{k+1})\|_2^2 + \frac{\beta}{2} \|\nabla \mathbf{b}\|_2^2, \quad q(\mathbf{b}) := \frac{\alpha}{2} \sum_{i=1}^{M^2} [\phi(\mathbf{b}_i)]^2.$$

Note that $r(\cdot; \mathbf{x}^{k+1}) : \mathbb{R}^{M^2} \rightarrow \mathbb{R}_+$ is a convex function with L_r -Lipschitz gradient and $q : \mathbb{R}^{M^2} \rightarrow \mathbb{R}_+$ encodes (large, depending on $\alpha \gg 1$) penalty contributions. Recalling Remark 2, we thus use again the proximal gradient descent algorithm for solving (20). The desired solution $\hat{\mathbf{b}}$ at each $k \geq 1$ can thus be found by iterating:

$$\mathbf{b}^{n+1} = \mathbf{prox}_{q, \delta}(\mathbf{b}^n - \delta \nabla r(\mathbf{b}^n)), \quad n = 1, 2, \dots, \quad (21)$$

for $\delta \in (0, \frac{1}{L_r}]$. The proximal operator $\mathbf{prox}_{q, \delta}(\cdot)$, has an explicit expression and it is defined element-wise for $i = 1, \dots, M^2$ as:

$$\left(\mathbf{prox}_{q, \delta}(\mathbf{d})\right)_i = \mathbf{prox}_{q, \delta}(\mathbf{d}_i) = \begin{cases} \mathbf{d}_i & \text{if } \mathbf{d}_i \geq 0, \\ \frac{\mathbf{d}_i}{1 + \alpha \delta} & \text{if } \mathbf{d}_i < 0. \end{cases} \quad (22)$$

4.3. Intensity and Background estimation results

Intensity estimation results can be found in Figure 6 where (13) is used for intensity/background estimation on the supports $\Omega_{\mathcal{R}}$ estimated from the first step of COLORME using $\mathcal{R} = \text{CEL0}$, $\mathcal{R} = \ell_1$ and $\mathcal{R} = \text{TV}$. We are referring to them as COLORME-CEL0, COLORME- ℓ_1 and COLORME-TV, respectively. The colormap ranges are different for the coarse-grid and fine-grid representations, as explained in section 6.1 The result on Ω_{TV} , even after the second step does not allow for the observation of a few significant details (e.g. the separation of the two filament on the bottom left corner) and that is why it will not further discussed.

A quantitative assessment for the other two regularization penalty choices, Ω_{CEL0} and Ω_{ℓ_1} , is available in Figure 7. More precisely we compute the Peak-Signal-to-Noise-Ratio (PSNR), given the following formula:

$$\text{PSNR}_{\text{dB}} = 10 \log_{10} \left(\frac{\text{MAX}_{\mathbf{R}}^2}{\text{MSE}} \right), \quad \text{MSE} = \frac{1}{L^2} \sum_{i=1}^{L^2} (\mathbf{R}_i - \mathbf{K}_i)^2, \quad (23)$$

where $\mathbf{R} \in \mathbb{R}^{L^2}$ is the reference image, $\mathbf{K} \in \mathbb{R}^{L^2}$ the image we want to evaluate using the PSNR metric and $\text{MAX}_{\mathbf{R}}$ the maximum value of the image \mathbf{R} . In our case, the reference image is the ground truth intensity image: $\mathbf{x}^{GT} \in \mathbb{R}^{L^2}$. The higher the PSNR, the better the quality of the reconstructed image.

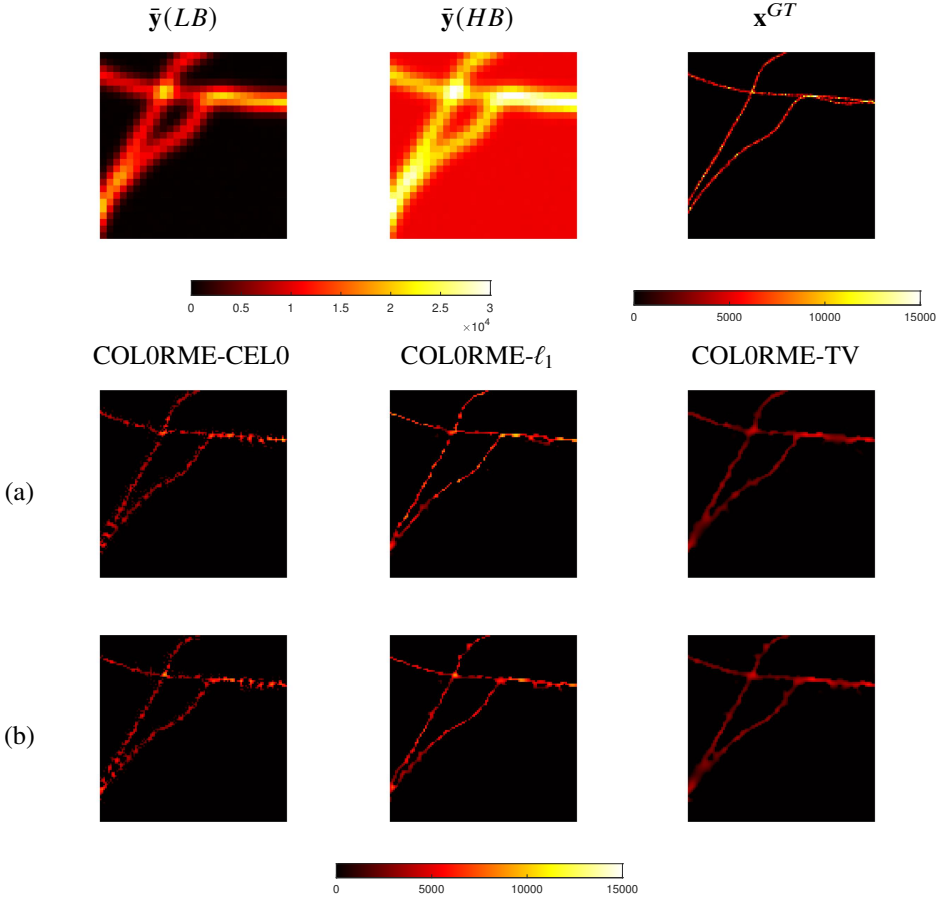


Figure 6. On top: Diffraction limited image $\bar{y} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t$, with $T=500$, (4x zoom) for the low-background (LB) dataset and for the high-background (HB) dataset, Ground truth (GT) intensity image. (a) Reconstructions for the noisy simulated dataset with low-background (LB), (b) Reconstruction for the noisy simulated dataset with high-background (HB). From left to right: intensity estimation result on estimated support using CEL0 regularization, ℓ_1 regularization and TV regularization. For all COL0RME intensity estimations, the same colorbar, presented at the bottom of the figure, has been used.

According to Figures 6 and 7, when only a few frames are considered (eg. $T = 100$ frames, high temporal resolution), the method performs better by using the CEL0 penalty for the support estimation. However, when longer temporal sequences are available (e.g. $T = 500$ or $T = 700$ frames) the method performs better by using the ℓ_1 -norm instead. In addition to this, for both penalizations, PSNR improves as the number of temporal frames increases.

Background estimation results are available in Figure 8 where (13) is used for intensity/background estimation on the supports $\Omega_{\mathcal{R}}$, with $\mathcal{R} = \text{CEL0}$ and $\mathcal{R} = \ell_1$, that have been already estimated in the first step. In the figure there is also the constant background generated by the SOFI Simulation Tool⁽³¹⁾, the software we used to generate our simulated data (more details in Section 6.1). Although the results look different due to the considered space-variant regularisation on \mathbf{b} , the variations are very little. The estimated background is smooth, as expected, while higher values are estimated near the simulated filaments and values closer to the true background are found away from them.

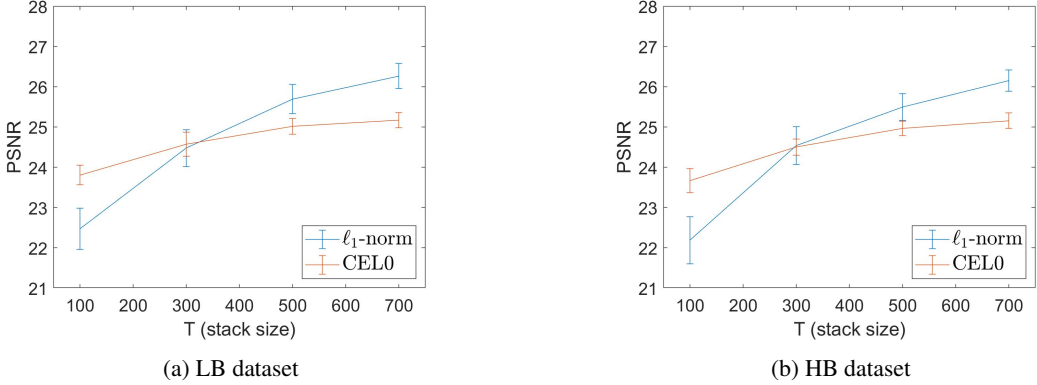


Figure 7. COLORME PSNR values for two different datasets (low-background and high-background dataset), stack sizes and regularization penalty choices. The mean and the standard deviation of 20 different noise realizations are presented.

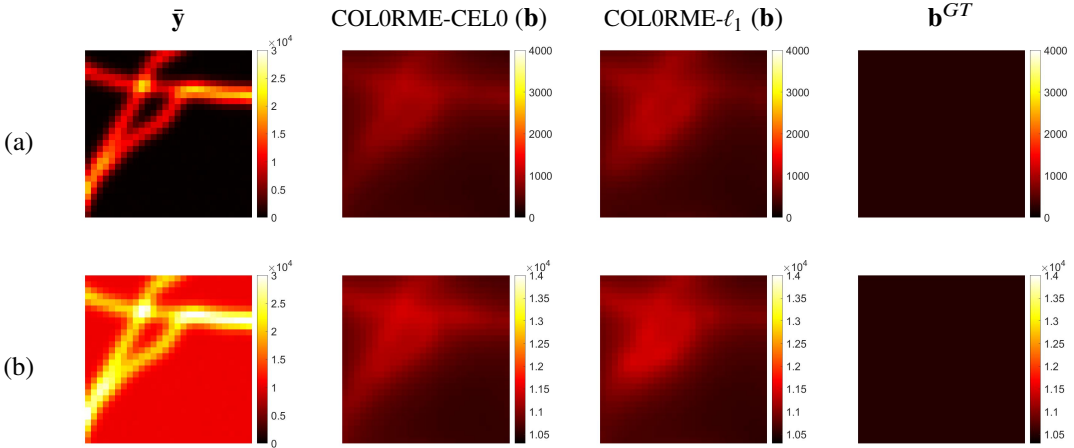


Figure 8. (a) Low-background (LB) dataset: Diffraction limited image $\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t$ with $T=500$ (4x zoom), Background estimation result on estimated support using CEL0 and ℓ_1 regularization, Ground truth (GT) background image. (b) High-background (HB) dataset: Diffraction limited image $\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t$ with $T=500$ (4x zoom), Background estimation result on estimated support using CEL0 and ℓ_1 regularization, Ground truth (GT) background image. Please note the different scales between the diffraction limited and background images for a better visualization of the results.

5. Automatic selection of regularization parameters

We describe in this section two parameter selection strategies addressing the problem of estimating the regularization parameters λ and μ appearing in the COLORME support estimation problem (7) and intensity estimation one (12), respectively. The other two regularization parameters β and α do not need fine tuning. They are both chosen arbitrary high, so as with large enough β to allow for a very smooth background and with very high α to respect the required constraints (positivity for both intensity and background and restriction to the predefined support only for the intensity estimation).

5.1. Estimation of support regularization parameter λ

The selection of the regularization parameter value λ in (7) is critical, as it determines the sparsity level of the support of the emitters. For its estimation, we start by computing a reference value λ_{max} , defined as the smallest regularization parameter for which the identically zero solution is found. It is indeed possible to compute such a λ_{max} for both regularization terms CEL0 and ℓ_1 (see⁽³²⁾ and⁽³³⁾). Once such values are known, we thus need to find a fraction $\gamma \in (0, 1)$ of λ_{max} corresponding to the choice $\lambda = \gamma\lambda_{max}$. For the CEL0 regularizer the expression for λ_{max} (see Proposition 10.9 in⁽³²⁾) is:

$$\lambda_{max}^{CEL0} := \max_{1 \leq i \leq L^2} \frac{\langle \mathbf{a}_i, \mathbf{r}_y \rangle^2}{2\|\mathbf{a}_i\|^2}, \quad (24)$$

where $\mathbf{a}_i = (\Psi \odot \Psi)_i$ denotes the i -th column of the operator $\mathbf{A} := \Psi \odot \Psi$. Regarding the ℓ_1 -norm regularization penalty, λ_{max} is given as follows:

$$\lambda_{max}^{\ell_1} := \|\mathbf{A}^\top \mathbf{r}_y\|_\infty = \max_{1 \leq i \leq L^2} \langle \mathbf{a}_i, \mathbf{r}_y \rangle. \quad (25)$$

As far as ℓ_1 is used as regularization term in (7), we report in Figure 9 a graph showing how the PSNR value of the final estimated intensity image (i.e. after the application of the second step of COLORME) varies for the two datasets considered depending on λ . It can be observed that for a large range of values λ , the final PSNR remains almost the same. Although this may look a bit surprising at a first sight, we remark that such a robust result is due, essentially, to the second step of the algorithm where false localizations related to an underestimation of λ can be corrected through the intensity estimation step. Note, however, that in the case of an overestimation of λ , points contained in the original support are definitively lost so no benefit is obtained from the intensity estimation step, hence the overall PSNR decreases.

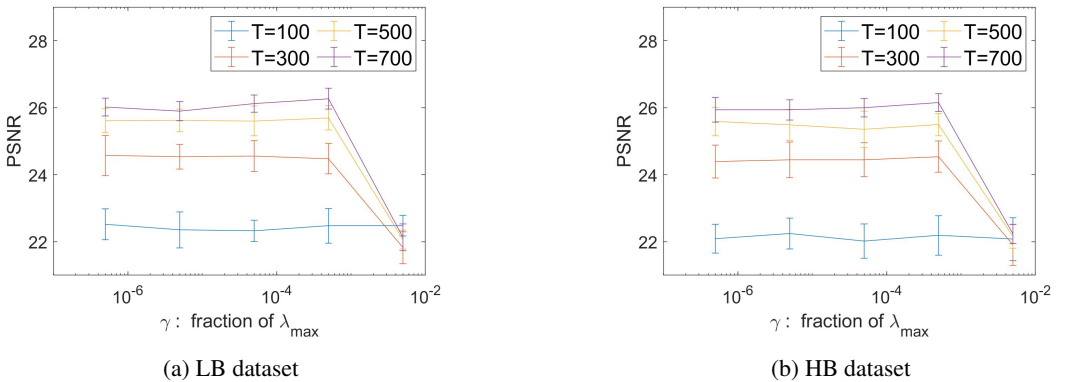


Figure 9. The PSNR value of the final COLORME image, using the ℓ_1 -norm regularizer for support estimation, for different γ values, evaluating in both the low-background (LB) and high-background (HB) dataset. The mean and the standard deviation of 20 different noise realization are presented.

When the CEL0 penalty is used for support estimation, a heuristic parameter selection strategy can be used to improve the localization results but also to avoid the fine parameter tuning. More specifically, the non-convexity of the model can be used by considering an algorithmic restarting approach to improve the support reconstruction quality. In short, a value of λ can be fixed, typically $\lambda = \gamma\lambda_{max}^{CEL0}$ with $\gamma \approx 5 \times 10^{-4}$, so as to achieve a very sparse reconstruction. Then, the support estimation algorithm

can be run and iteratively repeated with a new initialization (that is, restarted) several times. While keeping λ fixed along this procedure, a wise choice of the initialization depending, but not being equal to the previous output can be used to enrich the support, see Appendix C for more details. Non-convexity is here exploited by changing, for a fixed λ , the initialization at each algorithmic restart, so that new local minimizers (corresponding to possible support points) can be computed. The final support image can thus be computed as the superposition of the different solutions computed at each restarting. In such a way, a good result for a not-finely-tuned value of λ can be computed.

5.2. Estimation of intensity regularization parameter μ by discrepancy principle

In this section we provide some details on the estimation of the parameter μ in (12), which is crucial for an accurate intensity estimation. Recall that the problem we are looking at in this second step is

$$\text{find } \mathbf{x} \in \mathbb{R}^{L^2} \quad \text{s.t.} \quad \bar{\mathbf{y}} = \Psi \mathbf{x} + \mathbf{b} + \bar{\mathbf{n}}, \quad (26)$$

where the quantities correspond to the temporal averages of the vectorized model in (4), so that $\bar{\mathbf{n}} = \frac{1}{T} \sum_{t=1}^T \mathbf{n}_t$. The temporal realizations \mathbf{n}_t of the random vector \mathbf{n} follow a normal distribution with zero mean and covariance matrix $s\mathbf{I}_{M^2}$, where s has been estimated in the first step of the algorithm, see Section 3.1. Consequently, the vector $\bar{\mathbf{n}}$ follows also a normal distribution with zero mean and covariance matrix equal to $\frac{s}{T}\mathbf{I}_{M^2}$. As both s and T are known, we can use the discrepancy principle, a well-known a-posteriori parameter-choice strategy (see, e.g., ^(34,35)), to efficiently estimate the hyperparameter μ . To detail how the procedure is applied to our problem, we write \mathbf{x}_μ in the following to highlight the dependence of \mathbf{x} on μ . According to the discrepancy principle strategy, the regularization parameter μ is chosen so that the residual norm of the regularized solution satisfies:

$$\|\bar{\mathbf{y}} - \Psi \hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2 = \nu_{DP}^2 \|\bar{\mathbf{n}}\|_2^2, \quad (27)$$

where $\hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{M^2}$ are the solutions of (12). The expected value of $\|\bar{\mathbf{n}}\|_2^2$ is:

$$\mathbb{E}\{\|\bar{\mathbf{n}}\|_2^2\} = M^2 \frac{s}{T}, \quad (28)$$

which can be used as an approximation of $\|\bar{\mathbf{n}}\|_2^2$ for M^2 big enough. The scalar value $\nu_{DP} \approx 1$ is a 'safety factor' that plays an important role in the case when a good estimate of $\|\bar{\mathbf{n}}\|_2$ is not available. In such situations a value ν_{DP} closer to 2 is used. As detailed in Section 3.1, the estimation of s is rather precise in this case, hence we fix $\nu_{DP} = 1$ in the following.

We can now define the function $f(\mu) : \mathbb{R}_+ \rightarrow \mathbb{R}$ as:

$$f(\mu) = \frac{1}{2} \|\bar{\mathbf{y}} - \Psi \hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2 - \frac{\nu_{DP}^2}{2} \|\bar{\mathbf{n}}\|_2^2. \quad (29)$$

We want to find the value $\hat{\mu}$ such that $f(\hat{\mu}) = 0$. This can be done iteratively, using the Newton's method whose iterations read:

$$\mu_{n+1} = \mu_n - \frac{f(\mu_n)}{f'(\mu_n)}, \quad n = 1, 2, \dots \quad (30)$$

In order to be able to compute easily the values $f(\mu)$ and $f'(\mu)$, the values $\hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$, $\hat{\mathbf{b}} \in \mathbb{R}^{M^2}$ and $\hat{\mathbf{x}}'_\mu = \frac{\partial}{\partial \mu} \hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$ need to be computed, as it can be easily noticed by writing the expression of $f'(\mu)$ which reads:

$$f'(\mu) = \frac{\partial}{\partial \mu} \left\{ \frac{1}{2} \|\bar{\mathbf{y}} - \Psi \hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2 \right\} = (\hat{\mathbf{x}}'_\mu)^\top \Psi^\top (\bar{\mathbf{y}} - \Psi \hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}). \quad (31)$$

The values $\hat{\mathbf{x}}_\mu$ and $\hat{\mathbf{b}}$ can be found by solving the minimization problem (12). As far as $\hat{\mathbf{x}}'_\mu$ is concerned, we report in Appendix B the steps necessary for its computation. We note here, however, that in order to compute such a quantity, the relaxation of the support/non-negativity constraints by means of the smooth quadratic terms discussed above is fundamental. One can show that $\hat{\mathbf{x}}'_\mu$ is the solution of the following minimization problem:

$$\hat{\mathbf{x}}'_\mu = \arg \min_{\mathbf{x} \in \mathbb{R}^{L^2}} \frac{1}{2} \|\Psi \mathbf{x}\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x} + \mathbf{c}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu} \mathbf{x}\|_2^2 \right), \quad (32)$$

where \mathbf{c} is a known quantity defined by $\mathbf{c} = \frac{1}{\mu} \nabla \hat{\mathbf{x}}_\mu$, and the diagonal matrix $\mathbf{I}_{\hat{\mathbf{x}}_\mu} \in \mathbb{R}^{L^2 \times L^2}$ identifies the support of $\hat{\mathbf{x}}_\mu$ by:

$$\mathbf{I}_{\hat{\mathbf{x}}_\mu}(i, i) = \begin{cases} 0 & \text{if } (\hat{\mathbf{x}}_\mu)_i \geq 0, \\ 1 & \text{if } (\hat{\mathbf{x}}_\mu)_i < 0. \end{cases}$$

We can find $\hat{\mathbf{x}}'_\mu$ by iterating

$$\mathbf{x}'_{\mu^{n+1}} = \mathbf{prox}_{\bar{h}, \tau}(\mathbf{x}'_{\mu^n} - \tau \nabla \bar{g}(\mathbf{x}'_{\mu^n})), \quad n = 1, 2, \dots, \quad (33)$$

where

$$\bar{g}(\mathbf{x}) := \frac{1}{2} \|\Psi \mathbf{x}\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x} + \mathbf{c}\|_2^2, \quad \bar{h}(\mathbf{x}) := \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu} \mathbf{x}\|_2^2 \right). \quad (34)$$

For $\mathbf{z} \in \mathbb{R}^{L^2}$, the proximal operator $\mathbf{prox}_{\bar{h}, \tau}(\mathbf{z})$ can be obtained following the computations in Appendix A:

$$(\mathbf{prox}_{\bar{h}, \tau}(\mathbf{z}))_i = \mathbf{prox}_{\bar{h}, \tau}(\mathbf{z}_i) = \frac{\mathbf{z}_i}{1 + \alpha \tau \left(\mathbf{I}_\Omega(i, i) + \mathbf{I}_{\hat{\mathbf{x}}_\mu}(i, i) \right)}, \quad (35)$$

while

$$\nabla \bar{g}(\mathbf{x}') = (\Psi^\top \Psi + \mu \nabla^\top \nabla) \mathbf{x}' + \nabla^\top \nabla \hat{\mathbf{x}}_\mu, \quad (36)$$

and the step $\tau \in (0, \frac{1}{L_{\bar{g}}}]$, with $L_{\bar{g}} = \|\Psi^\top \Psi + \mu \nabla^\top \nabla\|_2$ the Lipschitz constant of $\nabla \bar{g}$. A pseudo-code explaining the procedure we follow to find the optimal $\hat{\mu}$ can be found in Algorithm 3. Finally, in Figure 10, a numerical example is available to show the good estimation of the parameter $\hat{\mu}$.

Algorithm 3 Discrepancy Principle

Require: $\bar{\mathbf{y}} \in \mathbb{R}^{M^2}$, $\mathbf{x}^0 \in \mathbb{R}^{L^2}$, $\mathbf{b}^0 \in \mathbb{R}^{M^2}$, $\mu_0, \beta > 0$, $\alpha \gg 1$

repeat

 Find $\hat{\mathbf{x}}_{\mu_n}, \hat{\mathbf{b}}$ using Algorithm 2

 Find $\hat{\mathbf{x}}'_{\mu_n}$ solving (32)

 Compute $f(\mu_n), f'(\mu_n)$ from (29) and (31)

$\mu_{n+1} \leftarrow \mu_n - \frac{f(\mu_n)}{f'(\mu_n)}$

until convergence

return $\hat{\mu}$

6. Results

In this section we compare the method COLORME with state-of-the-art methods that exploit the temporal fluctuations/blinking of fluorophores, while applying them to simulated and real data. More precisely we compare: COLORME-CEL0 (using the CEL0 regularization in the support estimation),

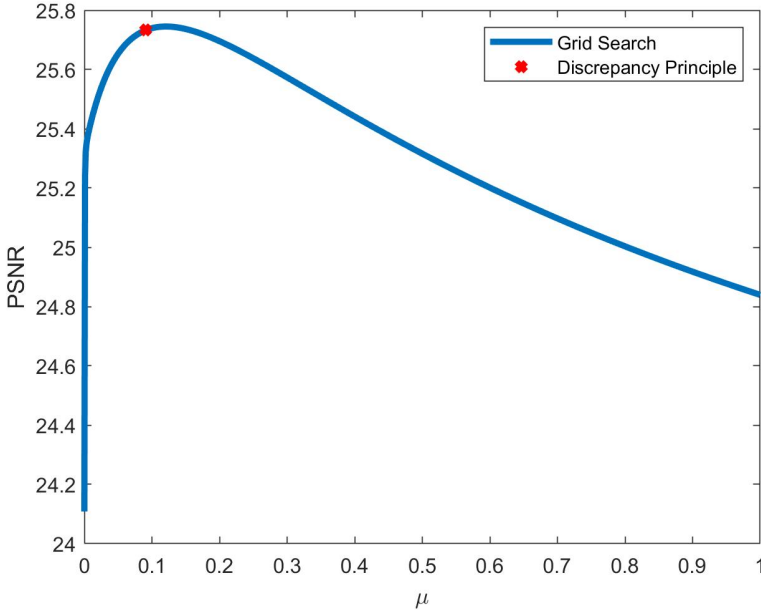


Figure 10. The solid blue line shows the PSNR values computed by solving (13) for several values of μ within a specific range. The data used are the HB dataset with $T = 500$ frames (Figure 12c) and the ℓ_1 -norm regularization penalty. The red cross shows the PSNR value $\hat{\mu}$ obtained by applying the Discrepancy Principle. We note that such value is very close to one maximizing the PSNR metric.

COLORME- ℓ_1 (using the ℓ_1 -norm regularization in the support estimation), SRRF⁽¹³⁾, SPARCOM⁽¹⁸⁾ and LSPARCOM⁽¹⁹⁾. We further performed preliminary comparisons also with the ESI, 3B and bSOFI approaches using available codes provided by the authors on the web¹, but we did not successfully obtain satisfactory results, so we omit them in the following.

6.1. Simulated Data

To evaluate the method COLORME we choose images of tubular structures that simulate standard microscope acquisitions with standard fluorescent dyes. In particular, the spatial pattern (see Figure 12a) is taken from the MT0 microtubules training dataset uploaded for the SMLM Challenge of 2016². The temporal fluctuations are obtained by using the SOFI Simulation Tool⁽³¹⁾. This simulation software, implemented in MATLAB, generates realistic stacks of images, similar to the ones obtained from real microscopes, as it makes use of parameters of the microscope setup and some of the sample's main properties. However, differently from the fluctuating³ microscopic data presented in section 6.2, the blinking generated by the SOFI Simulation Tool have a more distinctive "on-off" behaviour.

For the experiments presented in this paper, we generate initially a video of 700 frames, however we evaluate the methods using the first $T = 100$, $T = 300$, $T = 500$ and $T = 700$ frames, so as to examine further the trade-off between temporal and spatial resolution. The frame rate is fixed at 100 frames per second (fps) and the pixel size is 100 nm. Regarding the optical parameters, we set the numerical aperture equal to 1.4 and the emission wavelength to 525 nm, while the FWHM of the PSF is equal

¹ESI: <https://github.com/biophotonics-bielefeld/ESI>, 3B: <http://www.coxphysics.com/3b>, bSOFI implemented in SOFI Simulation Tool software package: <https://github.com/lob-epfl/sofitool>

²<http://bigwww.epfl.ch/smlm/datasets/index.html>

³the emission of a single fluorophore over time can be described by a Poisson distribution

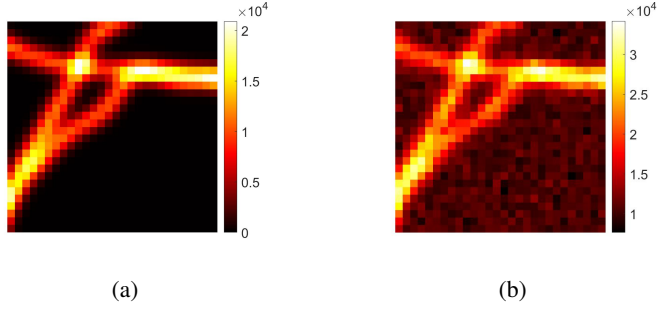


Figure 11. One frame of the HB dataset, before and after the addition of background and the simulated noise degradation. (a) A convoluted and down-sampled image $\Psi \mathbf{x}_t^{GT}$ obtained from a ground truth frame \mathbf{x}_t^{GT} , (b) A frame of the final noisy sequence: \mathbf{y}_t . Note the different colormaps to better capture the presence of noise and background.

to 228.75nm. The fluorophore parameters are set as follows: 20ms for on-state average lifetime, 40ms for off-state average lifetime and 20s for average time until bleaching. The emitter density is equal to 10.7 emitters/pixel/frame, while 500 photons are emitted, on average, by a single fluorescent molecule in every frame.

We create two datasets with the main difference between them being the background level, as in real scenarios the background is usually present. More precisely we create: the low-Background (LB) dataset, where the background is equal to 50 photons/pixel/frame and, the most realistic of the two, the high-Background (HB) dataset, where the background is equal to 2500 photons/pixel/frame. In both datasets, we proceed as follows: initially, Poisson noise is added to simulate the photon noise (see (2)); subsequently, the number of photons recorded by each camera pixel is converted into an electric charge in accordance with the quantum efficiency and gain of the camera that have been set to 0.7 and 6 respectively (thus resulting in an overall gain of 4.2); finally, Gaussian noise is added. In order to give a visual inspection of the background and noise, in Figure 11, one frame of the HB dataset is presented before and after the background/noise addition. As we want, also, to provide a quantitative assessment, we measure the quality of the reconstruction of the final sequence of T frames ($\mathbf{y}_t, t = 1, 2, \dots, T$) using the Signal-to-Noise-Ration (SNR) metric, given by the following formula:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{\frac{1}{TM^2} \sum_{i=1}^{TM^2} (\mathbf{R}_i)^2}{\frac{1}{TM^2} \sum_{i=1}^{TM^2} (\mathbf{R}_i - \mathbf{K}_i)^2} \right), \quad (37)$$

where $\mathbf{R} \in \mathbb{R}^{TM^2}$ is the reference image and $\mathbf{K} \in \mathbb{R}^{TM^2}$ the image we want to evaluate, both of them in a vectorized form. As reference, we choose the sequence of convoluted and down-sampled ground truth frames (see one frame of the reference sequence in Figure 11a). The SNR values for a sequence of $T = 500$ frames for the LB and HB dataset are 15.57dB and -6.07 dB, respectively. A negative value is computed for the HB dataset due to the very high background used in this case.

The diffraction limited image (the average image of the stack) of each dataset as well as the ground truth intensity image are available in Figure 12. In the LB dataset, due to the high signal values, the background is not visible. Further, as the observed microscopic images and the reconstructed ones belong to different grids, coarse and fine grid respectively, their intensity values are not comparable and we can not use the same colorbar to represent them. The intensity of one pixel in the coarse grid is the summation of the intensities of $q \times q$ pixels in the fine grid, where q is the super-resolution factor. For this reason, we use two different colorbars.

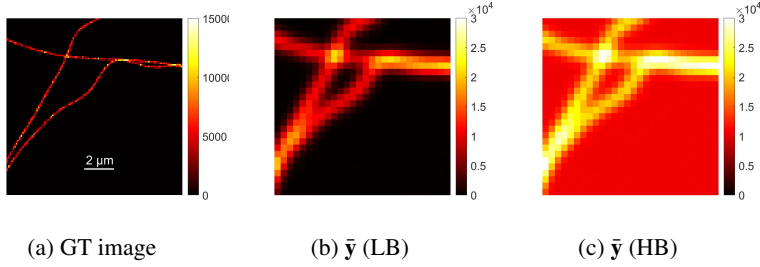


Figure 12. The Ground truth (GT) intensity image, as well as, the diffraction limited images $\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t$ for the two datasets with a 4x zoom, for a sequence of $T=500$ frames.

The comparison of the method COLORME with other state-of-the-art methods that take advantage of the blinking fluorophores is available bellow. Regarding the method COLORME-CEL0 and COLORME- ℓ_1 , a regularization parameter equal to $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{CEL0}$ and $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{\ell_1}$, respectively, is used in the support estimation. The hyper-parameters α and β are set as follows: $\alpha = 10^6$, $\beta = 20$. For the method COLORME-CEL0 the algorithmic restarting approach is used for a better support estimation. It stops when there are not additional pixels added to the estimated support or if a maximum number of 10 restarts is reached. Such number was empirically determined by preliminary simulations. For the method SRRF we are using the NanoJ SRRF plugin for ImageJ⁴. Concerning the method SPARCOM, we make use of the MATLAB code available online⁵. As regularization penalty we choose the ℓ_1 -norm with a regularization parameter equal to 10^{-10} and we avoid the post-processing step (the convolution with a small Gaussian function) for most precise localization. Finally we test the method LSPARCOM, using the code that is available online⁶ and the tubulin (TU) training set that is provided.

In Figure 13 we compare the reconstructions of the methods COLORME-CEL0, COLORME- ℓ_1 , SRRF, SPARCOM and LSPARCOM for the LB dataset and in Figure 14 for the HB dataset, for a sequence of $T = 500$ frames. Results for different stack sizes, are available in the Supplementary Figures S1, S2 and S3. Quantitative metrics like the Jaccard Index (JI) for the localization precision and the Peak-Signal-to-Noise-ration (PSNR) for the evaluation of the estimated intensities, are only available for the methods COLORME-CEL0 and COLORME- ℓ_1 (see Figures 4, 7). For the rest of the methods, the JI values are very small due to background and noise artifacts in the reconstructions that lead to the appearance of many false positives, while the PSNR metric is not possible to be computed as the methods SRRF, SPARCOM and LSPARCOM do not reconstruct the intensity level. In both datasets, LB and HB dataset, and for a sequence of $T = 500$ frames, the better reconstruction, visually, is the one of the method COLORME- ℓ_1 , as it is able to achieve a more clear separation of the filaments in the critical regions (yellow and green zoom boxes). The method COLORME-CEL0 achieves also a good result, eventhough the separation of the filaments, that are magnified in the green box, is not so obvious. The same happens also when the method SPARCOM is being used. Finally, the reconstruction of the methods SRRF and LSPARCOM, is slightly misleading.

6.2. Real Data

To show the effectiveness of our method for handling real-world data, we apply COLORME to an image sequence acquired from a Total Internal Reflection Fluorescence (TIRF) microscope. The TIRF microscope offers a good observation of the activities happening next to the cell membrane, as it uses an evanescent wave to illuminate and excite fluorescent molecules only in this restricted region of

⁴<https://github.com/HenriquesLab/NanoJ-SRRF>

⁵<https://github.com/KrakenLeaf/SPARCOM>

⁶<https://github.com/gilidar/LSPARCOM>

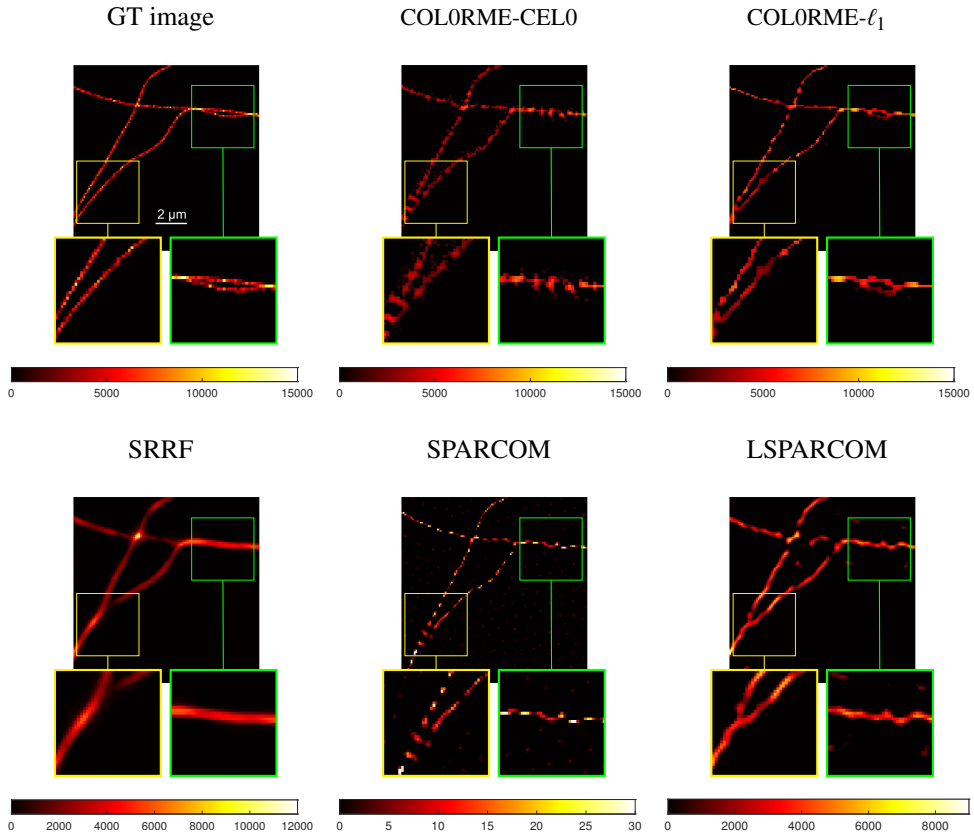


Figure 13. Results for the low-background (LB) dataset with $T = 500$. Note that the methods SRRF, SPARCOM and LSPARCOM do not estimate real intensity values. Between the compared methods only COLORME is capable of estimating them, while the other methods estimate the mean of a radiality image sequence (SRRF) and normalized autocovariances (SPARCOM, LSPARCOM).

the specimen⁽³⁶⁾. Further, the TIRF microscope does not require specific fluorescent dyes, allows live cell imaging by using a low illumination laser, with really low out-of-focus contribution and produces images with a relatively good, in comparison with other fluorescence microscopy techniques, SNR. To enhance the resolution of the images acquired from a TIRF microscope, super-resolution approaches that exploit the temporal fluctuations of blinking/fluctuating fluorophores, like COLORME, can be applied.

The data we are using have been obtained from a Multi-Angle TIRF microscope, with a fixed angle close to the critical one. A sequence of 500 frames has been acquired, with an acquisition time equal to 25s. Tubulins in endothelial cells are being imaged, while they are colored with the Alexa Fluor 488. The variance of fluctuations over time for a typical pixel is measured and is belonging to the range $5 \times 10^5 - 7 \times 10^5$. The diffraction limited image, or with other words the mean stack image $\bar{\mathbf{y}}$ is shown in Figure 15, together with one frame \mathbf{y}_t extracted from the entire stack. The FWHM of the PSF has been measured experimentally and is equal to 292.03nm, while the CCD camera has a pixel of size 106nm.

The results of the method COLORME-CELO and COLORME- ℓ_1 and more precisely the intensity and the background estimation, can be found in Figure 15. Experiments using different stack sizes have been done showing that the more frames we use (up to a point that we do not have many molecules bleached), the more continuous filaments we find. However, by acquiring only 500 frames we have

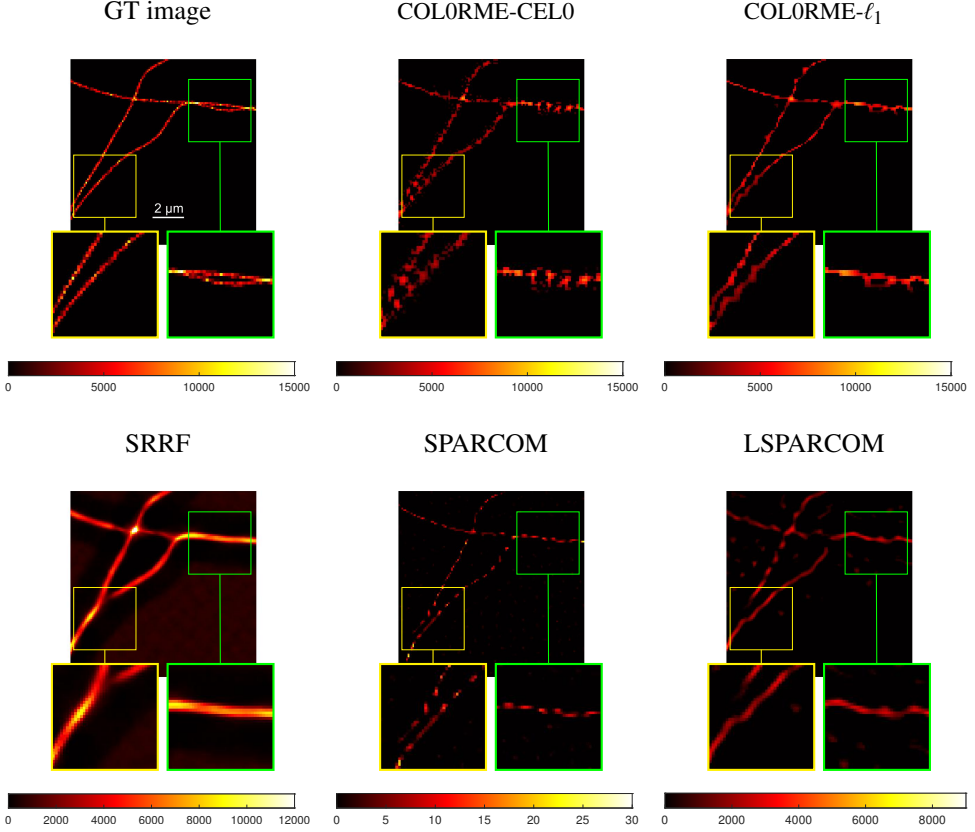


Figure 14. Results for the high-background(HB) dataset with $T = 500$.

a good balance between temporal and spatial resolution. For this reason we present here only results using a stack of 500 frames. For the method COLORME-CELO the regularization parameter λ is equal to $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{CELO}$ and the algorithmic restarting approach has been used (stopping criteria: when, in a certain restarting, there are not additional pixels added to the global support, but with maximum 10 restarts). Regarding the method COLORME- ℓ_1 the regularization parameter λ is equal to $\lambda = 5 \times 10^{-6} \times \lambda_{max}^{\ell_1}$, a relatively small value so as to be sure that we will include all the pixels that contain fluorescent molecules. Even if we underestimate λ and find more false positives in the support estimation, after the second step of the algorithm, the final reconstruction is corrected, as explained in 5.1. The hyper-parameters α and β are equal to: $\alpha = 10^6$, $\beta = 20$. Using any of the two regularizers the spatial resolution is enhanced, as it can be also observed from the yellow zoom boxes. However, the reconstruction obtained by both COLORME-CELO and COLORME- ℓ_1 is to some degree punctuated due to mainly limitations arising from experimental difficulties to get a staining sufficiently homogeneous for this imaging resolution. Furthermore, there are a few filaments that do not seem to be well reconstructed, especially using the COLORME-CELO method, e.g. the one inside the green box.

Finally, the comparison of the methods COLORME-CELO and COLORME- ℓ_1 with the other state-of-the-art methods, is available in Figure 16. The parameters used for the methods SRRF, SPARCOM and LSPARCOM, are explained in the section 6.1. Here, we further use the post-processing step (convolution with a small Gaussian function) in the method SPARCOM, as the result was dotted. The methods COLORME-CELO and COLORME- ℓ_1 seem to have the most precise localization, by reconstructing thin

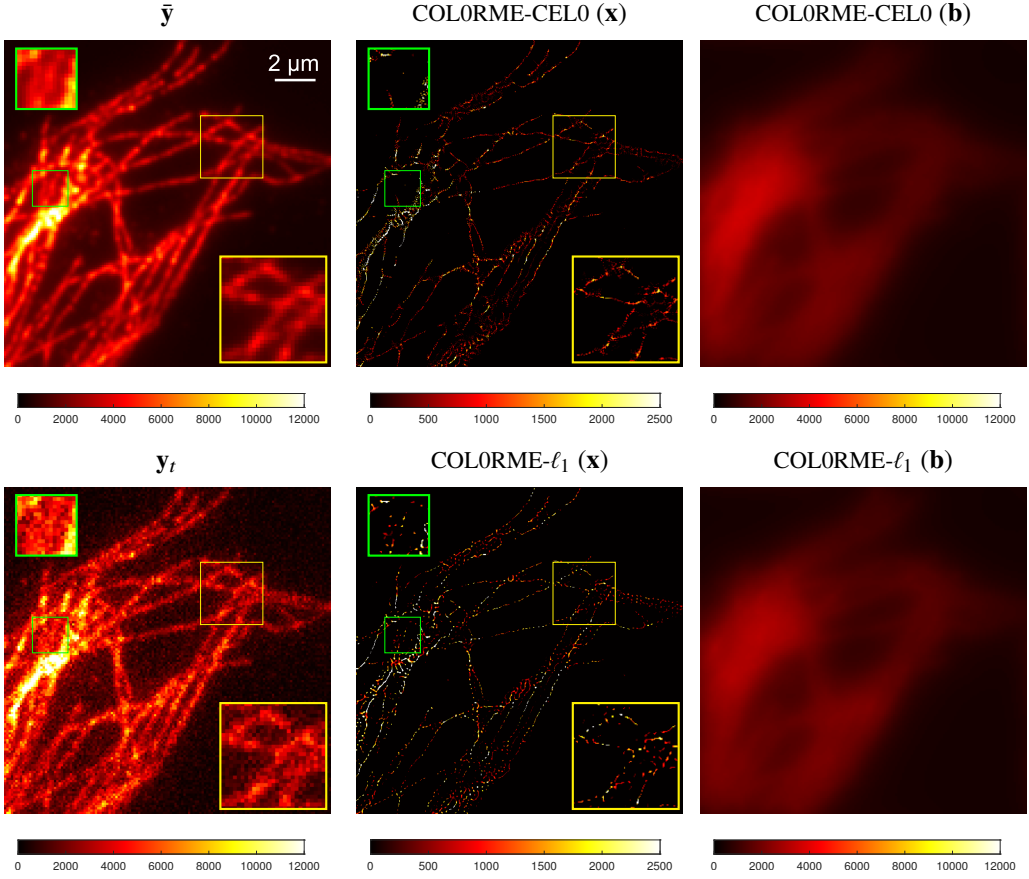


Figure 15. Real TIRF data, $T = 500$ frames. Diffraction limited image or the mean of the stack \bar{y} (4x zoom), A frame y_t from the stack (4x zoom), The intensity and background estimation of the methods COLORME-CELO and COLORME- ℓ_1 .

filaments, as shown in the cross-section plotted in Figure 16, though a bit punctuated. The most appealing visually is the result of the method SRRF, where the filaments have a more continuous structure, however from the cross-section, we can see that the resolution is not so much improved compared to the other methods. SPARCOM and LSPARCOM do not perform very well in this real image sequence due to, mainly, background artifacts.

7. Discussion and Conclusions

In this paper, we propose and discuss the model and the performance of COLORME, a method for super-resolution microscopy imaging based on the sparse analysis of the stochastic fluctuations of molecules' intensities. Similarly to other methods exploiting temporal fluctuations, COLORME relaxes all the requirements for special equipment (microscope and fluorescent dyes) and allows for live-cell imaging, due to the good temporal resolution and the low power lasers employed. In comparison with competing methods, COLORME achieves higher spatial resolution than other methods exploiting fluctuations while having a sufficient temporal resolution. COLORME is based on two different steps: a former one where accurate molecule localization and noise estimation are achieved by solving non-smooth

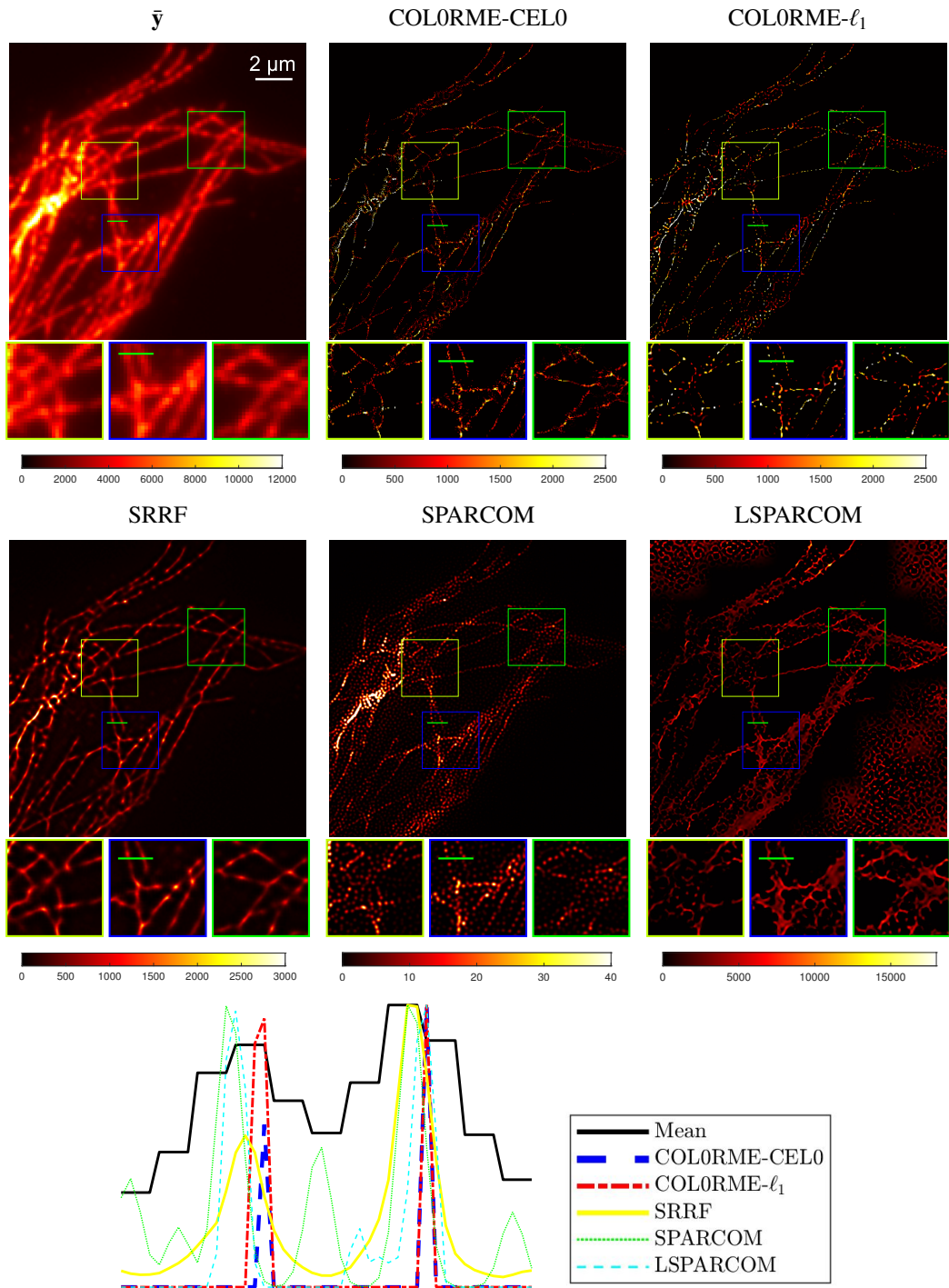


Figure 16. Real TIRF data, $T = 500$ frames. Diffraction limited image \bar{y} (4x zoom), Comparisons between the method that exploit the temporal fluctuations, Normalized cross-section along the green line presented in the diffraction limited and reconstructed images, but also in the blue zoom-boxes. Discription of colorbars: real intensity values for \bar{y} and COLORME in two different grids, mean of the radially image sequence for SRRF, normalized autocovariances for SPARCOM and LSPARCOM.

convex/non-convex optimization problems in the covariance domain and the latter where intensity information is retrieved in correspondence with the estimated support only. Our numerical results show that COLORME outperforms competing approaches in terms of localization precision. To the best of our knowledge, COLORME is the only super-resolution method exploiting temporal fluctuations which is capable of retrieving intensity-type information, signal and spatially-varying background, which are of fundamental interest in biological data analysis. For both steps, automatic parameter selection strategies are detailed. Let us remark that such strategy of intensity estimation could be applied to the other competing super-resolution methods in the literature. Several results obtained on both simulated and real data are discussed, showing the superior performance of COLORME in comparison with analogous methods such as SPARCOM, LSPARCOM and SRRF. Possible extensions of this work shall address the use of intensity information estimated by COLORME for 3D reconstruction in, e.g., MA-TIRF acquisitions. Furthermore, a systematic study to assess quantitatively the spatial resolution achieved by COLORME under different scenarios (different background levels, different PSNRs, number of frames) is envisaged.

A. Appendix. Proximal computations

Given the function $h : \mathbb{R}^{L^2} \rightarrow \mathbb{R}$, defined in (17), the proximal mapping of h is an operator given by:

$$\begin{aligned} \mathbf{prox}_{h,\tau}(\mathbf{w}) &= \arg \min_{\mathbf{u}} \left(\frac{1}{2\tau} \|\mathbf{u} - \mathbf{w}\|_2^2 + h(\mathbf{w}) \right) \\ &= \arg \min_{\mathbf{u}} \left(\frac{1}{2\tau} \|\mathbf{u} - \mathbf{w}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_{\Omega} \mathbf{u}\|_2^2 + \sum_{i=1}^{L^2} [\phi(\mathbf{u}_i)]^2 \right) \right). \end{aligned} \quad (\text{A.1})$$

The optimal solution $\hat{\mathbf{u}}$ ($\hat{\mathbf{u}} = \mathbf{prox}_{h,\tau}(\mathbf{w})$), as the problem (A.1) is convex, is attained when:

$$\begin{aligned} \mathbf{0} &\in \nabla \left(\frac{1}{2\tau} \|\hat{\mathbf{u}} - \mathbf{w}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_{\Omega} \hat{\mathbf{u}}\|_2^2 + \sum_{i=1}^{L^2} [\phi(\hat{\mathbf{u}}_i)]^2 \right) \right), \\ \mathbf{0} &\in \frac{1}{\tau} (\hat{\mathbf{u}} - \mathbf{w}) + \alpha \left(\mathbf{I}_{\Omega} \hat{\mathbf{u}} + [\phi(\hat{\mathbf{u}}_i) \phi'(\hat{\mathbf{u}}_i)]_{\{i=1, \dots, L^2\}} \right). \end{aligned} \quad (\text{A.2})$$

Starting from (14) we can compute $\phi' : \mathbb{R} \rightarrow \mathbb{R}_+$, as:

$$\phi'(z) := \begin{cases} 0 & \text{if } z \geq 0, \\ 1 & \text{if } z < 0, \end{cases} \quad \forall z \in \mathbb{R}. \quad (\text{A.3})$$

Given (A.3), we can write:

$$\mathbf{0} \in \frac{1}{\tau} (\hat{\mathbf{u}} - \mathbf{w}) + \alpha \left(\mathbf{I}_{\Omega} \hat{\mathbf{u}} + [\phi(\hat{\mathbf{u}}_i)]_{\{i=1, \dots, L^2\}} \right). \quad (\text{A.4})$$

Exploiting component-wise, as problem (A.1) is separable with respect to both \mathbf{x} and \mathbf{w} , and assuming $\hat{\mathbf{u}}_i \geq 0$, the derivative computed at (A.4) vanishes for:

$$\hat{\mathbf{u}}_i = \frac{1}{1 + \alpha\tau \mathbf{I}_{\Omega}(i, i)} \mathbf{w}_i, \quad (\text{A.5})$$

and it holds for $\mathbf{w}_i \geq 0$. Similarly, for the case $\hat{\mathbf{u}}_i < 0$, this analysis yields:

$$\hat{\mathbf{u}}_i = \frac{1}{1 + \alpha\tau(\mathbf{I}_\Omega(i, i) + 1)} \mathbf{w}_i, \quad (\text{A.6})$$

for $\mathbf{w}_i < 0$.

So finally, the proximal operator is given by:

$$(\mathbf{prox}_{h, \tau}(\mathbf{w}))_i = \mathbf{prox}_{h, \tau}(\mathbf{w}_i) = \begin{cases} \frac{\mathbf{w}_i}{1 + \alpha\tau\mathbf{I}_\Omega(i, i)} & \text{if } \mathbf{w}_i \geq 0, \\ \frac{\mathbf{w}_i}{1 + \alpha\tau(\mathbf{I}_\Omega(i, i) + 1)} & \text{if } \mathbf{w}_i < 0. \end{cases} \quad (\text{A.7})$$

In a similar way, we compute the proximal mapping of the function $\bar{h} : \mathbb{R}^{L^2} \rightarrow \mathbb{R}$, defined in (34), as follows:

$$\begin{aligned} \mathbf{prox}_{\bar{h}, \tau}(\mathbf{z}) &= \arg \min_{\mathbf{u}} \left(\frac{1}{2\tau} \|\mathbf{u} - \mathbf{z}\|_2^2 + \bar{h}(\mathbf{u}) \right) \\ &= \arg \min_{\mathbf{u}} \left(\frac{1}{2\tau} \|\mathbf{u} - \mathbf{z}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \mathbf{u}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu} \mathbf{u}\|_2^2 \right) \right). \end{aligned} \quad (\text{A.8})$$

The optimal solution $\hat{\mathbf{u}}$ of (A.8) ($\hat{\mathbf{u}} = \mathbf{prox}_{\bar{h}, \tau}(\mathbf{z})$) is attained when:

$$\begin{aligned} \mathbf{0} &\in \nabla \left(\frac{1}{2\tau} \|\hat{\mathbf{u}} - \mathbf{z}\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \hat{\mathbf{u}}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu} \hat{\mathbf{u}}\|_2^2 \right) \right), \\ \mathbf{0} &\in \frac{1}{\tau} (\hat{\mathbf{u}} - \mathbf{z}) + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{u}} + \mathbf{I}_{\hat{\mathbf{x}}_\mu} \hat{\mathbf{u}} \right). \end{aligned} \quad (\text{A.9})$$

By eliminating $\hat{\mathbf{u}}$ in the expression (A.9), we compute element-wise the proximal operator:

$$(\mathbf{prox}_{\bar{h}, \tau}(\mathbf{z}))_i = \mathbf{prox}_{\bar{h}, \tau}(\mathbf{z}_i) = \frac{\mathbf{z}_i}{1 + \alpha\tau \left(\mathbf{I}_\Omega(i, i) + \mathbf{I}_{\hat{\mathbf{x}}_\mu}(i, i) \right)}. \quad (\text{A.10})$$

B. Appendix. The minimization problem to estimate $\hat{\mathbf{x}}'_\mu$

Starting from the penalized optimization problem (13) and having \mathbf{b} fixed, we aim to find a relation that contains the optimal $\hat{\mathbf{x}}_\mu$. While there are only quadratic terms, we proceed as following :

$$\begin{aligned} \mathbf{0} &\in \nabla \left(\frac{1}{2} \|\Psi \hat{\mathbf{x}}_\mu - (\bar{\mathbf{y}} - \mathbf{b})\|_2^2 + \frac{\mu}{2} \|\nabla \hat{\mathbf{x}}_\mu\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \hat{\mathbf{x}}_\mu\|_2^2 + \sum_{i=1}^{L^2} [\phi((\hat{\mathbf{x}}_\mu)_i)]^2 \right) \right), \\ \mathbf{0} &\in \Psi^\top (\Psi \hat{\mathbf{x}}_\mu - (\bar{\mathbf{y}} - \mathbf{b})) + \mu \nabla^\top \nabla \hat{\mathbf{x}}_\mu + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)]'_{\{i=1, \dots, L^2\}} \right). \end{aligned} \quad (\text{A.11})$$

Given (A.3) we can write:

$$\mathbf{0} \in \Psi^\top (\Psi \hat{\mathbf{x}}_\mu - \bar{\mathbf{y}} - \mathbf{b}) + \mu \nabla^\top \nabla \hat{\mathbf{x}}_\mu + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)]'_{\{i=1, \dots, L^2\}} \right). \quad (\text{A.12})$$

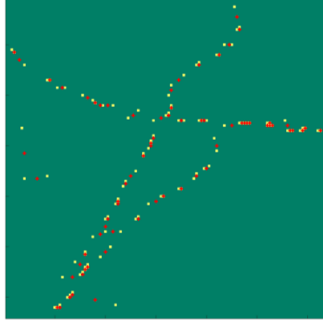


Figure 17. The yellow pixels belong to the support estimated in the previous restarting, while the red pixels belong to the initialization that is used in the current restarting.

Our goal is to compute $\hat{\mathbf{x}}'_\mu$, the partial derivative of $\hat{\mathbf{x}}_\mu$ w.r.t. μ . So, we derive as follows:

$$\begin{aligned} \frac{\partial}{\partial \mu} \left(\mathbf{0} \in \Psi^\top (\Psi \hat{\mathbf{x}}_\mu - \bar{\mathbf{y}} - \mathbf{b}) + \mu \nabla^\top \nabla \hat{\mathbf{x}}_\mu + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)]_{\{i=1, \dots, L^2\}} \right) \right), \\ \mathbf{0} \in \Psi^\top \Psi \hat{\mathbf{x}}'_\mu + \mu \nabla^\top \nabla \hat{\mathbf{x}}'_\mu + \nabla^\top \nabla \hat{\mathbf{x}}_\mu + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{x}}'_\mu + [\phi'((\hat{\mathbf{x}}_\mu)_i)(\hat{\mathbf{x}}'_\mu)_i]_{\{i=1, \dots, L^2\}} \right). \end{aligned} \quad (\text{A.13})$$

We define the matrix $\mathbf{I}_{\hat{\mathbf{x}}_\mu}$ such as:

$$\mathbf{I}_{\hat{\mathbf{x}}_\mu}(i, i) = \begin{cases} 0 & \text{if } (\hat{\mathbf{x}}_\mu)_i \geq 0, \\ 1 & \text{if } (\hat{\mathbf{x}}_\mu)_i < 0. \end{cases}$$

Now the vector $[\phi'((\hat{\mathbf{x}}_\mu)_i)(\hat{\mathbf{x}}'_\mu)_i]_{\{i=1, \dots, L^2\}}$, using further the equation (A.3), can be simply written as: $\mathbf{I}_{\hat{\mathbf{x}}_\mu} \hat{\mathbf{x}}'_\mu$ and then (A.13) becomes:

$$\mathbf{0} \in \Psi^\top \Psi \hat{\mathbf{x}}'_\mu + \mu \nabla^\top \nabla \hat{\mathbf{x}}'_\mu + \nabla^\top \nabla \hat{\mathbf{x}}_\mu + \alpha \left(\mathbf{I}_\Omega \hat{\mathbf{x}}'_\mu + \mathbf{I}_{\hat{\mathbf{x}}_\mu} \hat{\mathbf{x}}'_\mu \right).$$

The minimization problem we should solve in order to find $\hat{\mathbf{x}}'_\mu$ thus is:

$$\hat{\mathbf{x}}'_\mu = \arg \min_{\mathbf{x} \in \mathbb{R}^{L^2}} \frac{1}{2} \|\Psi \mathbf{x}\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x}\|_2^2 + \frac{1}{\mu} \|\nabla \hat{\mathbf{x}}_\mu\|_2^2 + \frac{\alpha}{2} \left(\|\mathbf{I}_\Omega \mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu} \mathbf{x}\|_2^2 \right). \quad (\text{A.14})$$

C. Appendix. Algorithmic restart.

Every initialization is based on the solution obtained at the previous restarting. There are many ways to choose the new initialization, deterministic and stochastic ones. In this paper we chose a deterministic way based on the following idea: for every pixel belonging to the solution of the previous restarting we find its closest neighbor. Then, we define the middle point between the two and we include it in the initialization of the current restarting. A small example is given in the Figure 17. The yellow points belong to the support estimation of the previous restarting. Starting from them we define the red points, used for the initialization of the current restarting.

Acknowledgements. The authors would like to thank E. van Obberghen-Shilling and D.Grall from the Institut de Biologie Valrose (iBV) who kindly prepared and provided the experimental samples. Furthermore, we would like to thank the anonymous reviewers for their valuable comments and suggestions.

Funding Statement. The work of VS and LBF has been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. LC acknowledges the support received by the academy on complex systems of UCA JEDI, the one received by the EU H2020 RISE program NoMADS, GA 777826, and the one received by the GdR ISIS grant SPLIN. The work of JHG was supported by the French Agence Nationale de la Recherche in the context of the project Investissements d'Avenir UCAJEDI (ANR-15-IDEX-01). Support for development of the microscope was received from IBiSA (Infrastructures en Biologie Santé et Agronomie) to the MICA microscopy platform.

Competing Interests. None

Data Availability Statement. Replication data and code can be found in: <https://github.com/VStergiop/COLORME>.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions. VS, LC, JHG and LBF conceived and designed the study. SS conducted data gathering. VS and JHG implemented the software. VS carried out the experiments. LC and LBF supervised the work. VS, LC, JHG and LBF wrote the article. All authors approved the final submission.

Supplementary Material. A supplementary document intended for publication has been provided with the submission.

D. References

1. Sage D, Kirshner H, Pengo T, Stuurman N, Min J, Manley S, et al. Quantitative Evaluation of Software Packages for Single-Molecule Localization Microscopy. *Nature methods*, 12. 2015 06.
2. Sage D, Pham TA, Babcock H, Lukes T, Pengo T, Chao J, et al. Super-resolution fight club: Assessment of 2D & 3D single-molecule localization microscopy software. *Nature Methods*. 2019 05;16.
3. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*. 2006;313(5793):1642-5. Available from: <https://science.sciencemag.org/content/313/5793/1642>.
4. Rust M, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*. 2006 11;3:793-5.
5. Li H, Vaughan J. Switchable Fluorophores for Single-Molecule Localization Microscopy. *Chemical Reviews*. 2018 09;118.
6. Hell SW, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett*, 19 (11). 1994:780-2. Available from: <http://ol.osa.org/abstract.cfm?URI=ol-19-11-780>.
7. Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*, 198 (2). 2000.
8. Candès E, Fernandez-Granda C. Super-Resolution from Noisy Data. *Journal of Fourier Analysis and Applications*. 2012;19:1229-54.
9. Denoyelle Q, Duval V, Peyré G, Soubies E. The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*. 2019 dec;36(1):014001. Available from: <https://doi.org/10.1088/1361-6420/ab2a29>.
10. Holden S, Uphoff S, Kapanidis A. DAOSTORM: An algorithm for high-density super-resolution microscopy. *Nature methods*. 2011 04;8:279-80.
11. Dertinger T, Colyer R, Iyer G, Weiss S, Enderlein J. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences*, 106 (52). 2009:22287-92. Available from: <https://www.pnas.org/content/106/52/22287>.
12. Geissbuehler S, Bocchio N, Dellagiacomma C, Berclaz C, Leutenegger M, Lasser T. Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI). *Optical Nanoscopy*. 2012 01;1.
13. Gustafsson N, Culley S, Ashdown G, Owen DM, Pereira PM, Henriques R. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature communications*, 7(1). 2016:12471-1.
14. Cox S, Rosten E, Monypenny J, Jovanovic-Talisman T, Burnette D, Lippincott-Schwartz J, et al. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature methods*. 2011 12;9:195-200.
15. Yahiatene I, Hennig S, Müller M, Huser T. Entropy-based Super-resolution Imaging(ESI): From Disorder to Fine Detail. *ACS Photonics*. 2015 07;2.
16. Deng Y, Sun M, Lin P, Ma J, Shaevitz J. Spatial Covariance Reconstructive (SCORE) Super-Resolution Fluorescence Microscopy. *PLoS one*. 2014 04;9:e94807.

17. Solomon O, Mutzafi M, Segev M, Eldar YC. Sparsity-based super-resolution microscopy from correlation information. *Opt Express*. 2018 Jul;26(14):18238-69. Available from: <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-14-18238>.
18. Solomon O, Eldar YC, Mutzafi M, Segev M. SPARCOM: Sparsity Based Super-resolution Correlation Microscopy. *SIAM Journal on Imaging Sciences*, 12 (1). 2019:392-419.
19. Dardikman-Yoffe G, Eldar YC. Learned SPARCOM: unfolded deep super-resolution microscopy. *Opt Express*. 2020 Sep;28(19):27736-63. Available from: <http://www.opticsexpress.org/abstract.cfm?URI=oe-28-19-27736>.
20. de Morais Goulart JH, Blanc-Féraud L, Debreuve E, Schaub S. A study on tensor and matrix models for super-resolution fluorescence microscopy. In: CAMSAP 2019 - IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing. Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing. Le Gosier, Guadeloupe; 2019. Available from: <https://hal.archives-ouvertes.fr/hal-02309659>.
21. Stergiopoulou V, de Morais Goulart JH, Schaub S, Calatroni L, Blanc-Féraud L. COLORME: Covariance-Based 10 Super-Resolution Microscopy with Intensity Estimation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021. p. 349-52.
22. Soubies E, Blanc-Féraud L, Aubert G. A Continuous Exact ℓ_0 Penalty (CEL0) for Least Squares Regularized Problem. *SIAM Journal on Imaging Sciences*, 8 (3). 2015:1607-39. Available from: <https://doi.org/10.1137/151003714>.
23. Candès E, Wakin M, Boyd S. Enhancing Sparsity by Reweighted L1 Minimization. *Journal of Fourier Analysis and Applications*. 2007 11;14:877-905.
24. Attouch H, Bolte J, Redont P, Soubeyran A. Proximal Alternating Minimization and Projection Methods for Non-convex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Mathematics of Operations Research*. 2010;35(2):438-57. Available from: <http://www.jstor.org/stable/40801236>.
25. Ochs P, Dosovitskiy A, Brox T, Pock T. On Iteratively Reweighted Algorithms for Nonsmooth Nonconvex Optimization in Computer Vision. *SIAM Journal on Imaging Sciences*. 2015;8(1):331-72. Available from: <https://doi.org/10.1137/140971518>.
26. Gazagnes S, Soubies E, Blanc-Féraud L. High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation. In: ISBI 2017; 2017. p. 28-31.
27. Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*. 2009;2(1):183-202.
28. Condat L. A Primal-Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. *Journal of Optimization Theory and Applications*. 2013 08;158.
29. Gale D, Shapley LS. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*. 1962;69(1):9-15. Available from: <http://www.jstor.org/stable/2312726>.
30. Combettes PL, Wajs VR. Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Model Simul*. 2005;4:1168-200.
31. Girsault A, Lukes T, Sharipov A, Geissbuehler S, Leutenegger M, Vandenberg W, et al. SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging. *PLOS ONE*, 11 (9). 2016;(9):1-13. Available from: <https://doi.org/10.1371/journal.pone.0161602>.
32. Soubies E. Sur quelques problèmes de reconstruction en imagerie MA-TIRF et en optimisation parcimonieuse par relaxation continue exacte de critères pénalisés en norme- ℓ_0 [Theses]. Université Côte d'Azur; 2016. Available from: <https://tel.archives-ouvertes.fr/tel-01479054>.
33. Koulouri A, Heins P, Burger M. Adaptive Superresolution in Deconvolution of Sparse Peaks. *IEEE Transactions on Signal Processing*. 2021;69:165-78.
34. Hansen PC. *Discrete Inverse Problems: Insight and Algorithms*. USA: Society for Industrial and Applied Mathematics; 2010.
35. Gfrerer H. An A Posteriori Parameter Choice for Ordinary and Iterated Tikhonov Regularization of Ill-Posed Problems Leading to Optimal Convergence Rates. *Mathematics of Computation*. 1987;49(180):507-22. Available from: <http://www.jstor.org/stable/2008325>.
36. Axelrod D. Total internal reflection fluorescence microscopy in cell biology. *Traffic*. 2001;2(11):764-74.