BENCHMARKING IMAGE-PROCESSING ALGORITHMS FOR BIOMICROSCOPY: REFERENCE DATASETS AND PERSPECTIVES

Daniel Sage, Hagai Kirshner, Cédric Vonesch, Stamatios Lefkimmiatis, Michael Unser

Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT

As the field of bioimage informatics matures, the issue of the validation of image reconstruction algorithms and the definition of proper performance criteria becomes more pressing. In this work, we discuss benchmarking aspects of fluorescence microscopy quantitative tools. We point out the importance of generating realistic datasets and describe our approach to this task. We rely on our experience and present arguments in favor of the use of 3D continuous-domain models of biological structures for simulating bioimaging datasets. We also present physically-realistic models of image formation that that are reasonably efficiently to implement.

Index Terms— Bioimage informatics, fluorescence microscopy, benchmarking, simulated datastets

1. INTRODUCTION

Advanced fluorescence microscopy for high-resolution imaging has been increasingly relying on quantitative tools. The ongoing effort in developing new and improved imaging methods is complemented by new mathematical approaches to image reconstruction and the emergence of fast and accurate software packages [1]. The net effect is an increasing intertwining between imaging modality and numerical reconstruction methods. The quality of the algorithmic components is decisive for extracting meaningful biological information and for leading to reproducible results.

Bioimaging introduces a plethora of imaging modalities, and the available software modules are proportionally diverse. Every software tool is developed for a particular biological need and made available to the research community at later stages. The performance of such algorithms, their applicability to similar imaging modalities, as well as their usability require further attention. Another important issue is software design and maintenance [2], which have immediate impact on reproducible research.

Grand software challenges have become an important framework for addressing such situations. They have been providing a common ground for evaluating algorithms in many other fields of engineering such as medical image processing [3] (presented to MICCAI), audio processing and video. Comparative evaluations are common in the field of biometric signal processing; e.g., the Face Recognition Grand Challenge [4]. For some applications, such comparative evaluations are a prerequisite for testing new algorithms. This is the case for stereo correspondence of two frames [5]. Such contests allow users to quickly know the state-of-art of a scientific field while developers of computational tools have a standardized protocol for comparing their algorithms. Yet, such challenges are not common practice for biological imaging applications. One possible reason is the lack of clear and agreed quantitative criteria for the image analysis task.

Following two recent biomicroscopy challenges—*particle tracking* and *segmentation of neuronal structures in EM stacks*—we introduce and analyze in this work the key aspects that make biomicroscopy challenges unique. We report on our experience in organizing two such challenges— *localization microscopy*[1] and *deconvolution microscopy*[2]—and point out the obstacles that need to be overcome in order to be successful. Localization microscopy and deconvolution microscopy are two computational approaches for improving the resolution of fluorescence micrographs. The final reconstructed images are highly dependent on the applied algorithm and on appropriate choice of parameters. In addition, a good characterization of the forward model and the acquisition device is required.

In the sequel, we describe our approach to datasets generation and to algorithm assessment and provide concrete examples. We argue that specifying a 3D continuous-domain model is more advantageous for simulating biological structures than considering discrete-domain phantoms [6-9], and that a realistic forward model does not necessarily impose computational limitations.

2. REFERENCE DATASETS

We point out two difficulties in establishing benchmark datasets in the field of biomicroscopy. The first arises from the plethora of imaging modalities, which does not allow the

[1] http://bigwww.epfl.ch/smlm/challenge/
[2] http://bigwww.epfl.ch/deconvolution/challenge/

community to rely on a small number of datasets. These imaging modalities correspond to several physical quantities (number of photons, phase difference, life time) and result in multi-dimensional datasets that include spatial, temporal and spectral information. Another difficulty arises from the lack of ground-truth information, which is a key component for algorithm validation and evaluation.

Biologists and microscopy practitioners prefer real data. But from an algorithm developer point of view, simulated data is a necessary step in the development and the assessment of software. The great potential of simulated datasets has already been demonstrated for the task of segmentation in Drelie et al. [10]. Since there is also considerable value in handling real imagery, we suggest having the following two types of datasets:

- real data with a-priori geometrical information about the biological structure (diameter, length, rotation, distances, symmetry, intersections);
- realistic simulated data of continuously-defined objects.

Within the context of simulated data, we bring up three aspects that are common to all imaging modalities: 1) continuous-domain description of the biological structure, 2) forward model characterization that take the optics and the noise sources into account, and 3) practical considerations about the data.

2.1. Continuous-domain description of biological structures

Sampling is an integral part of biomicroscopy, and we suggest describing the biological structure over the continuum. Doing so provides the ability to simulate different marker densities, to consider different magnification factors and CCD pixel sizes, to consider different lifetime models of the fluorescent molecules, and to conveniently include a drift model. Geometrical structures, such filaments, radially symmetric objects, and amorphous objects, are conveniently defined by splines; and we use results from spline theory for describing these 3D structures in a computationally efficient manner. The biological structure can then be generated at any scale, and visualized using several rendering methods (Fig. 1).

2.2. Forward model characterization

The forward model is very important in image reconstruction and inverse problem formulations, and biomicroscopy is no exception in this regard. A realistic point-spread function (PSF) model and noise sources characterization will result in realistic simulated data. As in many other bio-imaging algorithms (denoising, deconvolution and particle localization), the application of the forward model requires repeated calculations of the PSF, as well. This, in turn, introduces a computational complexity aspect of the datasets generation task.

The Gaussian function provides a reasonable approximation of the main lobe of the Airy pattern while introducing relatively low computational complexity.

Such approximation, however, discards the side-lobes of the PSF, which are particularly important in 3-D PSF modeling. The trade-off between choosing realistic and simplified PSF models is execution time, and we suggest to overcome that by using the scalar Gibson and Lanni PSF model. It accounts for the refractive index mismatch that is often present in biomicroscopy imaging and it based on parameters that are readily available for the microscopy practitioner. A computationally effective implementation of the Gibson and Lanni model was recently introduced in [13] within the context of localization microscopy, and we suggest to utilize it for reference datasets, as well. In order to further accelerate the PSF evaluation, we prepare a lookup table at a resolution of 2 nm in the z axis and for each set of acquisition parameters. We then apply linear interpolation when required.

Noise sources and perturbations may include: shot noise for small photons count; EMCCD and read-out noise models; quantization; file storage; background auto-fluorescence; sample drift for long experiments; inaccurate values of numerical apertures, magnification and refractive indices. The random nature of the emission process of the fluorophore should be taken into account, as well.

2.3. Practical issues

Biomicroscopy data is often very large in terms of storage. This, in turn, limits the accessibility of such datasets over the Internet. The simulated dataset has to be then of a moderate size, say 100MB. It is also important that datasets will deliver metadata information that includes acquisition and other relevant parameters, which will make the simulation reproducible. This will allow for relatively simple adjustments when a new imaging modality is introduced.

3. BENCHMARKING EXAMPLES

3.1. Assessments

The simulated data should include an ensemble of scenarios that test the performance of the algorithms for several criteria.

The criteria should include quantitative and qualitative measures that are important for both the algorithm developer and the end-user, e.g. accuracy, exactitude, fidelity, computational time, visual inspection. The trade off between the various criteria is important to analyze, as well.

From an end-user point of view, the use of the algorithm should be as simple as possible and intuitive. Parameters have to be clearly defined. Installation should be effortless, as well. These features define the software usability [2], which is also interesting to evaluate.

3.2. Localization Microscopy

Super-resolution fluorescence microscopy is an emerging field that allows one to study the living cell at the nanometer scale. It overcomes the classical diffraction limit of Abbe, yielding images of biological structures that have lateral resolution down to 10 nm.

One of the most promising techniques in this field is single-molecule localization. It was originally introduced by three independent groups, giving rise to several acronyms: PALM (photo-activated localization), STORM (stochastic optical reconstruction microscopy) and F-PALM. Unlike classical fluorescence microscopy, however, the acquired data cannot be visualized directly and there is a need for a localization algorithm that determines the precise location of the blinking single-molecules. The algorithmic aspects of this method are therefore extremely important, and this is the focus of the localization microscopy challenge we have conducted at the ISBI 3013 conference.

To this aim, we generated ground-truth synthetic datasets that are inspired by the tubulin biological structure. We define a tubulin over the continuum and create a biological structure that is composed of several tubulins. Each tubulin is randomly positioned in 3D space and at different orientations. We then randomly marked them with fluorophores, as shown in Figure 1.

The acquired data is composed of consecutive frames that consist of a sparse set of excited fluorophores each. We randomly excite the fluorophores and assign a random lifetime model to each one of them. The fluorophores emit a random quantity of photons that are spatially spread by the 3D Gibson and Lanni PSF (we also the have the option of simulating a 3D defocussed Gaussian PSF). The noise sources we consider are: non-homogenous excitation intensity over the field of view, background scatter noise, auto-fluorescence, EMCCD multiplicative noise, read-out noise and dark pixels. For a specified frame rate, our simulator computes thousands of synthetic images that contain millions of active fluorophores. The computation is carried out at a high resolution of 5 nm/pixel and the images are then down-sampled to the camera resolution, which is typically 150 nm/pixel. The simulated data further includes acquisition perturbations such as the camera gain, saturation and quantization, as well as the file-format. The ground-truth data consists of the fluorophores positions and rendering of the data at high resolution.

To determine the rate of detection, standard statistical methods are used to compare two sets of positions: Jaccard index, F-Score, precision and recall. The accuracy is computed by evaluating the RMSE of the localization distance error. We also evaluate the algorithm run-time and the software usability. In addition, we provide visual comparison of the reconstructed data.

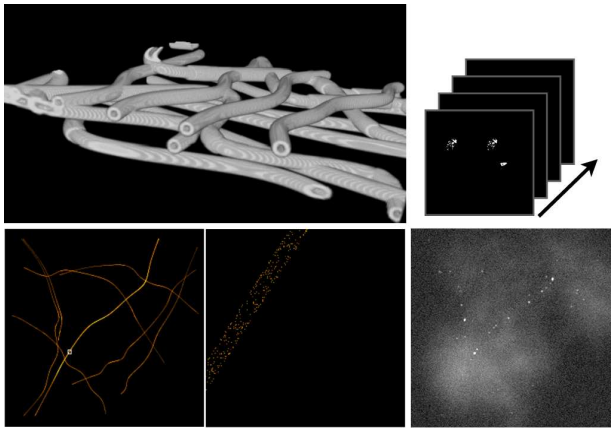

Figure 1. Simulated network of microtubules: these structures of 25 nm in diameter are continuously defined by B-spline functions. Part of the network of tubulins is shown here as a volume rendering for illustration purpose only (top-left); super-resolution PALM-type data that originates from the simulated network, only few fluorophores are activated in a frame (top-right); a single frame of the PALM-type sequence with fluorescent molecules and auto-fluorescent background (bottom-right); reconstructed images at 150 and 2 nm/pixel (bottom-left).

3.3. 3D Deconvolution Microscopy

The aim of the 3D Deconvolution Microscopy Challenge was to benchmark existing deconvolution algorithms, to stimulate the development of new methods as well as creating a network of deconvolution-software providers originating from distinct communities. Indeed, these providers have traditionally been segmented into three very different categories: academic developers, creators of free/open-source solutions and commercial software companies. We also spent a significant effort on making the conditions of the challenge reasonably realistic, in terms of data size, measurement model and processing pipeline.

The datasets for this challenge were computer-generated, with an effort to reproduce various resolution-critical features that are typically observed in cellular biology. Each class of features (e.g., filaments, cellular membrane) was allocated a specific channel.

In addition we simulated a complete image-acquisition pipeline, with rarely-combined realistic characteristics such as a Gibson and Lanni PSF model, Toeplitz boundary conditions, a Poisson+Gaussian noise mixture and quantization effects. Figure 2 shows maximum-intensity projections of the resulting datasets that were distributed to the participants, as well as deconvolution results obtained with the well-established Richardson-Lucy algorithm.

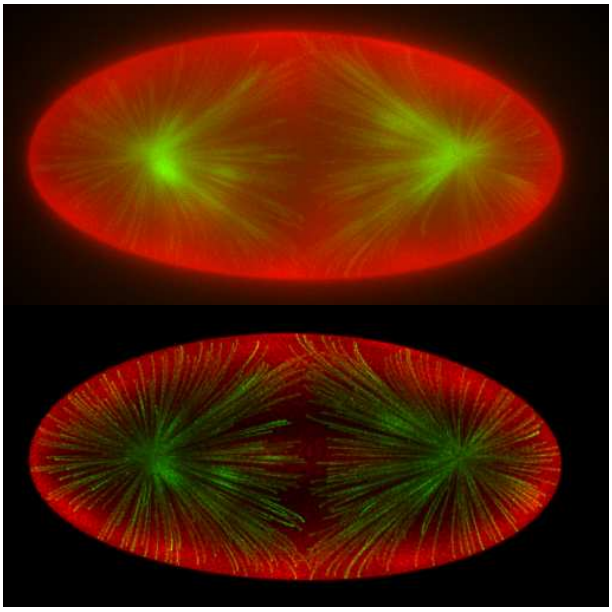

Figure 2: Maximum-intensity projections of two channels of the data used for the deconvolution challenge. Top: computer-generated phantom with simulated blurring and noise. Bottom: deconvolution result obtained with the Richardson-Lucy algorithm.

Finally we chose a number of standard and more advanced performance metrics to assess the results submitted by the participants. These metrics included for example the signal-to-noise ratio (SNR), the structure-similarity index (SSIM), a wavelet-domain sparsity measure, as well as total-variation and Hessian-based metrics.

4. DISCUSSION

Biomicroscopy benchmarking introduces several challenges such as the generation of realistic datasets in a computationally efficient manner and the choice of an acceptable set of performance metrics. The localization microscopy and the 3D deconvolution challenges, for example, addressed these aspects by describing biological structures in the continuum and by focusing on quantitative and qualitative metrics that are important for both the algorithm developers and the end-users. Special attention was given to the forward model which includes the PSF, the noise sources and the life-time of the fluorophores. Similar to other engineering fields, we expect biomicroscopy challenges to stimulate the development of better algorithms. Such a periodic effort will help in defining validation protocols, promote interdisciplinary dialogue and emphasize on the importance of quantitative tools for biomicroscopy.

However, innovation should not be discouraged by over-standardization that will prevent researchers from publishing new original solutions, which do not completely fit into the established model. The outcomes of such challenges will have great value for bioimaging communities such as the Open Bio Image Alliance.

5. REFERENCES

[1] Editorial, "The quest for quantitative microscopy," *Nature Methods, Special Issue Focus on Bioimage Informatics*, vol. 9, pp. 627, 2012.

[2] A.E. Carpenter, L. Kamentsky and K.W, Eliceiri, "A Call for Bioimaging Software Usability," *Nature Methods*, vol. 9, pp. 666-670, 2012.

[3] Grand Challenge in Medical Image Analysis, http://www.grand-challenge.org/, 2013.

[4] P.J. Phillips, P. J. Flynn, T. Scruggs, K.W. Bowyer and W. Worek. "Preliminary Face Recognition Grand Challenge Results," *7th International Conference on Automatic Face and Gesture Recognition*, pp. 15-24, 2006.

[5] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-view Stereo Reconstruction Algorithms," *2006 IEEE Computer Society Conference On Vision and Pattern Recognition*, vol. 1, pp. 519-528, 2006.

[8] D. Svoboda, M. Kozubek and S. Stejskal, "Generation of Digital Phantoms of Cell Nuclei and Simulation of Image Formation in 3D Image Cytometry," *Cytometry Part A*, vol. 75, pp. 494-509, 2009.

[9] A. Lehmussola, P. Ruusuvuori, J. Selinummi, H. Huttunen and O. Yli-Harja, "Computational Framework for Simulating Fluorescence Microscope Images with Cell Populations," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1010–1016, 2007.

[10] T. Zhao and R.F. Murphy, "Automated Learning of Generative Models for Subcellular Location: Building Blocks for Systems Biology," *Cytometry Part A*, vol. 71, pp. 978–990. 2007.

[11] S. Rajaram, B. Pavie, N. Hac, S.J. Altschuler and L.F. Wu, "SimuCell: a flexible framework for creating synthetic microscopy images," *Nature Methods*, vol. 9, pp. 634–635, 2012.

[12] E.D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval and B.S. Manjunath, "A Biosegmentation Benchmark for Evaluation of Bioimage Analysis Methods," *BMC Bioinformatics,* vol. 10, 2009.

[13] H. Kirshner, F. Aguet, D. Sage, M. Unser, "3-D PSF Fitting for Fluorescence Microscopy: Implementation and Localization Application," Journal of Microscopy, vol. 249, pp. 13-25, 2013.

This work is supported by the Eurobioimaging project, Work Package 11. http://www.eurobioimaging.eu/