Peta-scale lightsheet microscopy reconstruction and maintenance of existing software

Submitted by: Stephan Preibisch, Stephan Saalfeld

Background and current state-of-the-art

Lightsheet fluorescence microscopy (LFSM) is an emerging technology that creates a thin sheet of light to illuminate the sample. Emitted light is collected in an orthogonal direction to the lightsheet, which enables optical sectioning with high resolution, low photobleaching and high speed. While LSFM has initially been used to image developing organisms with high spatial and temporal resolution, it has recently gained a lot of attention for imaging large fixed and expanded samples (e.g. EASI-FISH project [1]).

To be able to acquire large volumes (e.g. entire mouse brains), many overlapping three-dimensional image tiles are acquired using microscopic stages, sample rotation, and/or different illumination directions. Additionally, acquired stacks can be sheared due to a non-orthogonal orientation (e.g. 45 degrees) of the stage relative to the detection objective.

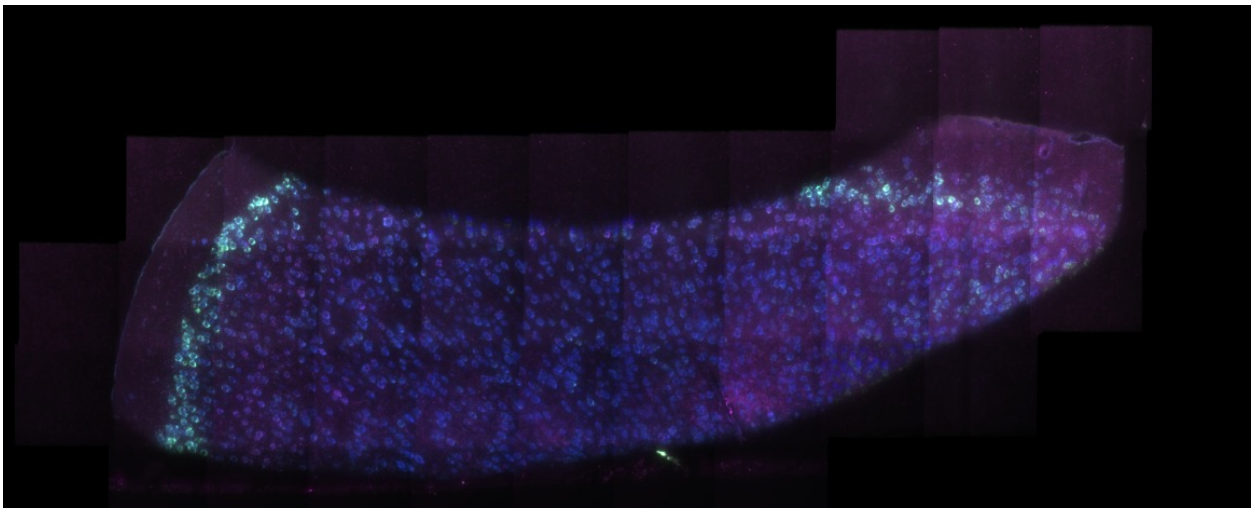

***Fig. 1**: Cut-plane through a reconstruction of a multi-terabyte lightsheet microscopy volume of a mouse cortex staining nuclei and several mRNA species with smFISH. The size of the isotropic dataset is 13864x22366x5891 pixel (2773 x 4473 x 1178 um) at 16 bit depth covering 6 channels (image data curtesy of Tim Wang & Karl Svoboda).*

To analyze such data, registration as well as fusion or deconvolution of the tiled images is required. The authors of this proposal have developed a powerful software ecosystem based on ImgLib2 [2] and BigDataViewer [3] over the years that is able to reconstruct such datasets accurately and efficiently. This includes the ImageJ Stitching plugin [4], BigSticher [5], and Stitching-Spark [6], which are heavily used by labs outside and within Janelia (e.g. EASI-FISH project, Keller lab, AIC, microscopy core, Singer lab, Svoboda lab, Spruston lab), highlighted by the fact that they have been cited more than 2500 times so far.

While the use of lightsheet microscopy in biology is growing, we are lacking resources to maintain and teach these essential, highly successful lightsheet reconstruction software projects.

Excitingly, there are currently several efforts that aim at imaging significantly larger samples (human and monkey tissues, expanded mouse brain, multiplexed samples) by modifying existing lightsheet microscopes. These datasets are expected to reach the peta-byte range, for which no general software solution exists to date[1]. Further development of our software to support such large datasets requires significant development effort. However, it would enable pioneering key efforts at the Allen Institute, Cold Spring Harbor Laboratories, and HHMI Janelia and would potentially make peta-scale lightsheet imaging widely available as a technology to the community.

Scope of the project

Goal 1: Maintenance of existing lightsheet software (4 SCA months + 2 FTE months)

Maintenance of existing tools for lightsheet reconstruction requires constant adjustment to changes in file formats, underlying APIs & frameworks, as well as bug fixes, implementations of small requests by users to adjust them to new use cases, and update of existing documentation. Additionally, the software needs to be taught at workshops, conferences and within Janelia to maximize its usability and impact.

Goal 2: Combining existing tools into scalable software (10 SCA months + 2 FTE month)

BigStitcher (and with that Image Stitching) and Stitching-Spark are both implemented using the Java/ImgLib2 frameworks. However, they are currently separate software projects, where each of them possess unique capabilities.

Stitching-Spark provides functionality for efficient flatfield correction, registration based on phase correlation using translation and affine models, and image fusion. Importantly, it uses N5 as a file format and is implemented using Apache Spark, which enable efficient parallelization on workstations, clusters (at Janelia using IPP) and the cloud.

BigStitcher additionally provides registration based on interest points, which is more robust and supports translation, affine, and non-rigid transformation models required for precise alignment of large images. Furthermore, BigStitcher provides a powerful GUI that enables users to effectively visualize and interact with the reconstruction process, has a generic data import module supporting most types of microscopy data directly, can provide quality estimation for the acquired data, and can perform GPU-based deconvolution. Its biggest limitation compared to Stitching-Spark is that it only supports multithreaded processing and is thus limited to datasets of about 1-2 terabytes.

---

[1] An exception is the ad-hoc solution developed specifically for the rhesus monkey brain paper [7]. It is, however, not general and cannot not be directly applied to similar systems. Furthermore, it shows relatively high registration errors (Fig. 1i,j), which questions the general stability and applicability of the software approach for this purpose.

Our goal is to create a software package that combines current capabilities of Stitching-Spark and BigStitcher and is thereby able to scale to petabyte-sized datasets. We already developed a prototype software for the Svoboda lab that we intend to use as starting point for this project.

Goal 3: New features (10 SCA months + 2 FTE month)

The development of new features will additionally be necessary to reconstruct large-scale lightsheet microscopy datasets. First, most microscopes that are currently used to acquire very large datasets are based on a diSPIM-like setup [8]. Therefore, un-shearing of the data is a crucial first step. We already developed a prototype software that can compute per-plane un-shear transformations with unprecedented accuracy, significantly easing downstream processing. Additionally, the data is un-sheared virtually using ImgLib2, thereby reducing the memory footprint drastically and avoiding multiple rounds of pixel interpolation, which otherwise reduces the image quality. Second, imaging of large samples will most likely require physical cutting of the sample. We will therefore adapt our code that was originally developed for slab-stitching of the FlyEM samples to lightsheet microscopy. Finally, we plan to develop a scalable deconvolution software combining fusion and deconvolution packages including GPU support from BigStitcher and Stitching-Spark.

Milestones

This project represents a multi-year effort. At the end of the first year, we plan to have a working prototype software that combines the features of Stitching-Spark and BigStitcher as well as un-shearing, which will allow large-scale microscopy efforts to make use of the software. At the end of the second year, we plan to finish implementation of the current goals.

Documentation & outreach

We plan to conserve the GUI workflow of BigSticher as much as possible, which would allow us the extend existing documentation and tutorials. We plan to teach the new software at workshops, conferences, and internally. Towards the end of the project, we plan to release a bioRxiv (or similar) description of the project to maximize the impact on the community.

Required skillset

This project will require a dedicated developer (e.g. an SCA) that will be supervised and supported by the authors of this proposal. In addition, we plan to partially include Tobias Pietzsch, who is a co-author on frameworks central to this effort and who will help designing and implementing the code. Additionally, the Allen Institute as well as Cold Spring Habor Laboratories volunteered to support such an effort with local computer scientists helping to implement, adapt and test the software. For intuitive documentation and GUI layout, we plan to request help from a scientific writer and an UI expert if necessary.

Estimated impact

Our current software for lightsheet image reconstruction is already heavily used by the community (>2500 citations to date). Since the field of lightsheet microscopy is growing (more and more commercial and customs systems are becoming available), we expect that we can

further increase the usage of the software simply by maintaining it, teaching it, and adjusting it to new types of microscopes and use cases.

Peta-scale lightsheet projects are currently the new frontier and there are only a few labs/institutions around the world that are currently able to embark on such projects. Excitingly, most of them acknowledge the complexity of the software needed to reconstruct their data and are excited to collaborate, test, and adapt the software. Thus, we are confident that we'll be able to have most of the labs/institutions that work on such projects use and extend our software. We hope that this project will make the new field of petabyte-scale lightsheet microscopy accessible to the community.

References

[1] Wang, Y. et al., "Expansion-Assisted Iterative-FISH defines lateral hypothalamus spatio-molecular organization", bioRxiv 2021.03.08.434304 (2021).

[2] Pietzsch, T., Preibisch, S., Tomančák, P., & Saalfeld, S., "ImgLib2—generic image processing in Java", Bioinformatics, 28(22), 3009-3011 (2012).

[3] Pietzsch, T., Saalfeld, S., Preibisch, S. *et al.* BigDataViewer: visualization and processing for large image data sets. *Nat Methods* **12,** 481–483 (2015).

[4] S. Preibisch, S. Saalfeld, P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions", *Bioinformatics*, **25**(11):1463-1465 (2009).

[5] Hörl D, Rojas Rusak F, Preusser F, Tillberg P, Randel N, Chhetri RK, Cardona A, Keller PJ, Harz H, Leonhardt H, Treier M, Preibisch S., "BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples", *Nature Methods* **16**, 870–874 (2019).

[6] Gao R, Asano SM, Upadhyayula S, Pisarev I, Milkie DE, Liu TL, Singh V, Graves A, Huynh GH, Zhao Y, Bogovic J, Colonell J, Ott CM, Zugates C, Tappan S, Rodriguez A, Mosaliganti KR, Sheu SH, Pasolli HA, Pang S, Xu CS, Megason SG, Hess H, Lippincott-Schwartz J, Hantman A, Rubin GM, Kirchhausen T, Saalfeld S, Aso Y, Boyden ES, Betzig E., "Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution", *Science* **18**, eaau8302.

[7] Xu F. et al, "High-throughput mapping of a whole rhesus monkey brain at micrometer resolution", *Nat. Biotech*. https://doi.org/10.1038/s41587-021-00986-5 (2021).

[8] Kumar, A., Wu, Y., Christensen, R., Chandris, P., Gandler, W., McCreedy, E., Bokinsky, A., Colón-Ramos, D. A., Bao, Z., McAuliffe, M., Rondeau, G., & Shroff, H.,"Dual-view plane illumination microscopy for rapid and spatially isotropic imaging", *Nature protocols*, *9*(11), 2555–2573 (2014).