NAS: Neuron Analyzer Suite for Automatic Analysis of Neuronal Activities from Calcium Imaging Data

Mohamed Hamed
Center for bioinformatics (CBI), Saarland University, Saarbrucken, Germany
Institute for Biostatistics and Informatics and Ageing Research, Rostock Uni Medical Center, Rostock, Germany
Email: mhamed@bioinformatik.uni-saarland.de

Jonathan Odul
National Institute of Informatics (NII), Tokyo, Japan

Andrew Steven Miller, Koichi Kawakami, and Kitamoto Asanobu
National Institute of Genetics (NIG), Mishima, Japan
Email: kitamoto@nii.ac.jp

***Abstract*—Neuronal activities can be visualized through functional Multi-neuron Calcium Imaging (fMCI) using fluorescence microscopes. However, the necessary components of data processing pipelines have been separately implemented and not well integrated previously. To this end, this paper introduces a Neuron Analyzer Suite (NAS) that integrates several components, from image preprocessing to inferred functional connectivity, in order to establish a processing pipeline for calcium imaging data. NAS is composed of three main modules. First, a motion removal module employing image registration is applied to raw images to remove artifacts caused by the motion of model organisms. Second, a neuron segmentation module is utilized to segment neurons and to extract their time series signals. Third, a neuron connectivity module is applied to time series signals to bi/cluster neurons based on their activity patterns and to reverse engineer their putative functional connectivity using a Bayesian learning approach. The NAS suite was tested on zebrafish calcium images to infer functional connectivity that may change over time.**

***Index Terms*—calcium imaging data, neuron activity, neuron connectivity, and bayesian network**

I. Introduction

Calcium imaging data have been widely used for the real-time observation of neuronal activity, and the analysis of calcium transients is currently an active area of research. The challenge of analyzing calcium activity images is two-fold. First, the preprocessing and segmentation of image data requires the identification and extraction of neurons from images with motion artifacts and noise. Second, the inference of neuron connectivity requires the learning of a connectivity model from incomplete time series data. To solve these two challenges, we propose an integrated software tool called NAS (Neuron Analyzer Suite) to help biologists understand relationships between neurons observed as temporal signals in calcium images.

A number of software tools have been proposed to analyze calcium imaging data. CellSort [1] is widely used for image segmentation and time series signal extraction, but it must be combined with other tools for image preprocessing and connectivity inference. In order to infer network connectivity, the lack of true connectivity data for a population of neurons, such as data validated by anatomical measures, is a major obstacle in calcium imaging studies. To cope with this problem, simulated calcium imaging data is widely used to validate learning algorithms, as in Bayesian approaches [2], but our target is the real data without simulation data. To work on the real data, it is important to note that inference of functional connections has fundamental limitations due to incomplete information, such as the problem of unobserved common inputs. What we infer is, at best, an approximate description of the network, and the functional connectivity is a reconstruction of the pair-wise connections that best reproduces the data [3]. Related software tools are also summarized in a survey [4].

This paper deals with real calcium imaging data observed using transgenic zebrafish expressing the GECI (Genetically Encoded Calcium Indicator), GCaMP7a. The zebrafish is a suitable model animal for fluorescence imaging studies to visualize neuronal activity because its body is transparent throughout the embryonic and larval stages. Details of the images used in the experiment are described in [5], and our motivation is to develop software tools to infer neuron connectivity from the conditions found in real data.

Manuscript received June 3, 2015; revised August 10, 2015.

doi: 10.18178/ijpmbs.4.3.167-170

II. Description

NAS is a freely available desktop application that can provide initial insights on how signals are transmitted between neurons. It processes two-dimensional, single-cell resolution calcium imaging frames to infer the functional connectivity of neurons in response to cognitive processes, perceptions, or behaviors. Furthermore, NAS integrates several analytical algorithms starting from image processing to data portioning and probabilistic graphical models. Fig. 1 illustrates the analytical stages of NAS. Section 2.1 describes the motion removal module, which is applied to raw images to remove artifacts caused by the motion of model organisms. Section 2.2 describes the neuron segmentation module, which is applied to registered image sequences to segment neurons and to extract their time series signals along entire frames. Finally, in Section 2.3, we describe the neuron connectivity module, which is applied to time series signals to bi/cluster the neurons based on their signal patterns and to reverse engineer their functional connectivity using a Bayesian learning approach. The major contribution of the paper is in the neuron connectivity module, which includes the clustering of neurons that share similar activity patterns and the inference of connectivity among neurons in a cluster.

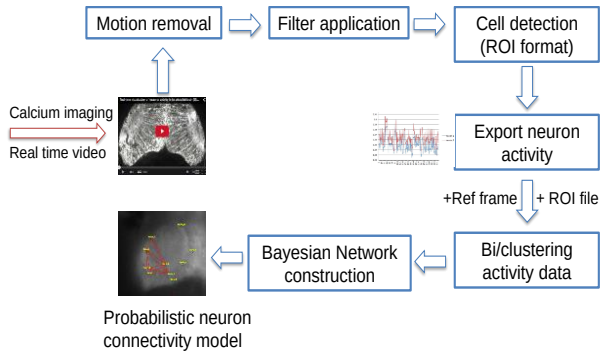

Figure 1. Analytical stages of the NAS tool.

A. Motion Removal Module

The first processing step of NAS is image registration in order to remove motion artifacts produced by movement of the model organism. Our method is based on a spatial normalization approach to deform frames so that a single feature in one frame corresponds to the same location in other frames. The SIFT (Scale-Invariant Feature Transformation) algorithm [6], [7] is used as local feature detector to identify distinctive key-points between different frames. By matching key-points across frames under the assumption of rigid motion on the two-dimensional focal plane, the tool can estimate motion vectors across frames. The tool has an option to select any frame as the reference frame, and all frames are warped to it, as shown in Fig. 2.

The results of the registration are then saved as a text file with a motion vector for each frame. These results can be checked using the included motion viewer tool, which displays an animation of the motion vectors over time, with a scroll bar for controlling the speed of the animation (Fig. 3).

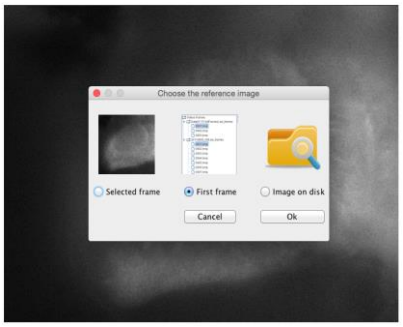

Figure 2. Selecting an image frame for image warping.

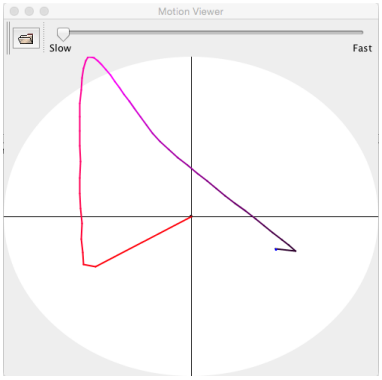

Figure 3. Motion viewer tool displaying the motion directions of the model organism while recording neuronal activity.

B. Neuron Segmentation Module

The purpose of this module is to segment the cell bodies of neurons in the registered image frames to extract neuronal signals (activities) along time. This process will be automated in future versions of NAS, but the current version provides an ROI editor for the manual segmentation of neurons in order to achieve reliable and flexible segmentation (Fig. 4). The output of the tool is an ROI file which contains the spatial locations and boundaries of neurons. The activity of each neuron is then measured by averaging pixel intensities inside the ROI. Repeating this process over the entire course of the frames produces time series signals representing neuronal activities. This result can be further exported in an XML format.

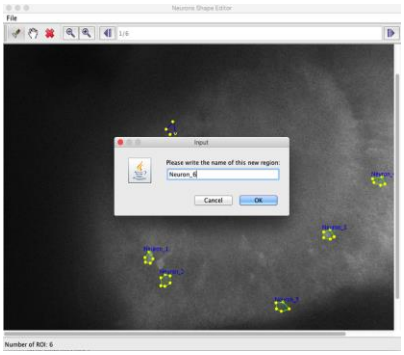

Figure 4. ROI editor to manually segment neuron shapes from the processed frames.

C. Neuron Connectivity Module

The last module concerns the analysis and visualization of neuronal activities recorded from the second module. Display panels (such as a heatmap or chart views) help with the interpretation of signal characteristics, but the main purpose of the module is to provide algorithms to infer the connectivity of neurons. The basic strategy of this module is a combination of neuron activity clustering and connectivity inference. Neurons are first clustered based on the similarity of activity signal patterns along certain time points (frames), and the connectivity of neurons is then inferred for each cluster. An important assumption here is the temporal dynamics of neuron groupings. Assuming that the grouping is the same over all time points, we can apply clustering algorithms on the activity signals. However, when the groupings of neurons can change dynamically over time, bi-clustering algorithms are useful for the simultaneous clustering of neurons and time points. Therefore, the tool offers two types of algorithms for a user to choose the best-fit clustering algorithm.

The module offers k-means and the hierarchical clustering algorithm (HCL) as options for clustering. In addition, the module provides two bi-clustering algorithms, Order-preserving submatrices (OPSM) [8] and maximum similarity bi-clusters (MSBE) [9], with options for additive and constant bi-clustering. The implementation of this module is based on the BicAT [10], [11] bi-clustering analysis toolbox, which has been customized here to fit our problem and extended with further functions and algorithms (primarily the MSBE algorithms). After clustering, a software component for Bayesian network construction is applied to infer neuron connectivity with a probabilistic graphical model. The probabilistic model is inferred within a cluster based on the assumption that neurons with common cellular activities suggest underlying connectivity. The Bayesian network approach is as previously described in [12].

The NAS tool accepts both neuron activity data and ROI files that are generated from the previous modules. The tool first visualizes the time series data in a heatmap view. A user then applies one of the four clustering or bi-clustering algorithms to group neurons that share correlated activity signals. For each identified bi/cluster, the tool outlines the neurons and time points involved, (Fig. 5) and generates a network file representing a putative connectivity model between the neurons. The network files are saved in a folder called "ncadatasets" and can be further visualized by other tools such as Cytoscape [13] for topological visualization, and SVG (Scalable Vector Graphics) for an overlay of network structure on the reference frame of calcium imaging data.

III. Case Study

Our tool was applied to calcium imaging data from zebrafish larvae and generated hypothetical models for the connectivity between zebrafish neurons. We used the MSBE algorithm (the additive version) as a bi-clustering method with the default parameters and this yielded one neuron bi-cluster involving six neurons (Neuron_1, Neuron_2, Neuron_8, Neuron_9, Neuron_7, Neuron_6) that show similar cellular activities along specific frames. A network model elucidating a potential signal transmission between these neurons was inferred and visualized in Fig. 6.

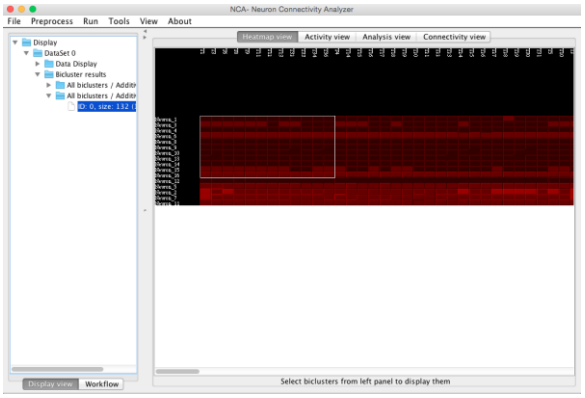

Figure 5. Heatmap view for the output neuron bi-clusters.

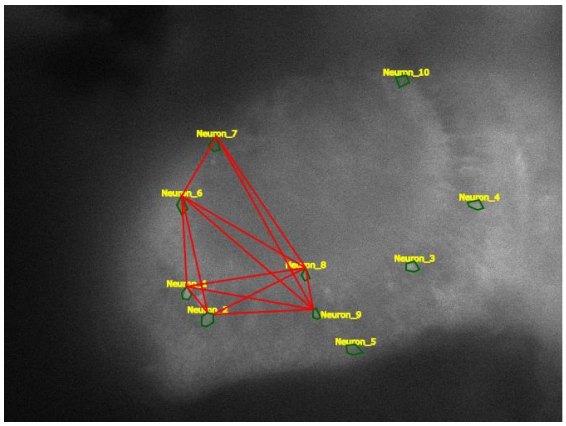

Figure 6. A putative connectivity model for zebrafish larvae in response to visual stimuli.

Interestingly, Neuron_8 and Neuron_9 are working as central hubs for the inferred connectivity model, which in turns reflects their putative critical role in transmitting the cellular signals in the zebrafish brain in response to visual stimuli. The zebrafish calcium imaging sample is bundled with the NAS package.

IV. Conclusion

We developed NAS as an automatic software suite that analyzes calcium imaging data of neuronal cells and infers the connectivity network of the neurons. NAS combines three independent modules of image registration, image segmentation, and network topology learning to extract neuron locations and their dynamics, and to therefore infer a hypothetical neuronal connectivity model with minimal human supervision. The application of the NAS suite to zebrafish calcium images demonstrated the usefulness of NAS in building a putative connectivity model for zebrafish neurons.

Although the NAS suite was tested on zebrafish calcium imaging data, the tool itself does not assume a specific model organism, and has the potential to be applied to other types of calcium imaging data, in order to

reach a better understanding of how the brain performs computations during perception and behaviors.

V. Outlook

NAS suite is planned to be extended with additional processing and analysis features in the near future. First, we plan to add automatic neuron segmentation so that a user can choose manual or automatic segmentation. Second, we plan to add visualization and validation functionality to help biologists check the evidence of inferred connectivity using temporal and spatial views. Another interesting direction in the future may be to support three-dimensional calcium imaging data in order to construct neuron connectivity models extending in the X, Y, and Z directions.

Acknowledgment

MH was supported by the international internship program from the National institute of informatics (NII), Tokyo, Japan, in collaboration with the GRADUS global program and graduate school of computer science, Saarland University, Saarbrucken, Germany.

References

[1] E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer, "Automated analysis of cellular signals from large-scale calcium imaging data," *Neuron*, vol. 63, no. 6, pp. 747-760, 2009.
[2] Y. Mishchencko, J. T. Vogelstein, and L. Paninski, "A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data," *The Annals of Applied Statistics*, pp. 1229-1261, 2001.
[3] I. H. Stevenson, *et al.*, "Inferring functional connections between neurons," *Current Opinion in Neurobiology*, vol. 18, no. 6, pp. 582-588, 2008.
[4] R. Mikut, *et al.*, "Automated processing of zebrafish imaging data: A survey," *Zebrafish*, vol. 10, no. 3, pp. 401-421, 2013.
[5] A. Muto, *et al.*, "Real-time visualization of neuronal activity during perception," *Current Biology*, vol. 23, no. 4, pp. 307-311, 2013.
[6] D. G. Lowe, "Object recognition from local scale-invariant features," *in the Proceedings of the Seventh IEEE International Conference*, 1999.
[7] D. G. Lowe, "Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image," Google Patents, 2004.
[8] A. Ben-Dor, *et al.*, "Discovering local structure in gene expression data: The order-preserving submatrix problem," *Journal of Computational Biology*, vol. 10, no. 3-4, pp. 373-384, 2003.
[9] X. Liu and L. Wang, "Computing the maximum similarity bi-clusters of gene expression data," *Bioinformatics*, vol. 23, no. 1, pp. 50-56, 2007.
[10] F. M. Al-Akwaa, M. H. Ali, and V. M. Kadah, "BicAT_Plus: An automatic comparative tool for bi/clustering of gene expression data obtained using microarrays," presented at Radio Science Conference, 2009.
[11] S. Barkow, *et al.*, "BicAT: A biclustering analysis toolbox," *Bioinformatics*, vol. 22, no. 10, pp. 1282-1283, 2006.
[12] M. Hamed, *et al.*, "Integrative network based approach identifies key genetic elements in breast invasive carcinoma," *BMC Genomics*, vol. 16, no. Suppl. 5, p. 2, 2015.
[13] P. Shannon, *et al.*, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498-2504, 2003.

Mohamed Hamed is a senior PhD student in the graduate school of computer science, Center for Bioinformatics (CBI) Saarbrucken, Germany. He received his master degree in information technology, from the school of computer science, Nottingham University, United Kingdom, in 2008. He worked as a software engineer for ESRI for 5 years. From April 2010 to March 2015, He was at the Graduate School of Computer Science as a PhD student. His current research involves developing and applying computational approaches to study the regulatory machinery of stem cell differentiation process as well as the pathogenicity of complex diseases.

Jonathan ODUL received his engineer's degree in Computer, L'École Polytechnique de l'universit é Grenoble-I, in 2012. He was involved in part of this research during his 6 months internship at the National Institute of Informatics the same year. Since his graduation, he is now working as a software engineer for a Japanese company in Tokyo.

Andrew Steven Miller is a PhD student at the Division of Molecular and Developmental Biology, National Institute of Genetics and Department of Genetics, the Graduate University for Advanced Studies (Sokendai), Mishima, Japan. He is currently studying the development of forebrain circuitry in the larval zebrafish brain though calcium imaging and closed loop behavioral assays.

Koichi Kawakami is a professor of Division of Molecular and Developmental Biology, National Institute of Genetics and Department of Genetics, the Graduate University for Advanced Studies (Sokendai), Mishima, Japan. He received his PhD from University of Tokyo (Molecular Biology). He developed a vertebrate transposon system, applied it to genetic methods in zebrafish including BAC transgenesis, gene trapping, enhancer trapping and the Gal4-UAS system, and also applied it to transgenesis in other model vertebrates and gene transfer in culture cells. He is now studying functional neuronal circuits in the zebrafish brain by calcium imaging and behavioral genetics.

Kitamoto Asanobu is an associate professor of National Institute of Informatics, and also an associate professor of SOKENDAI, Japan. He received his PhD in electronic engineering from the University of Tokyo in 1997, and joined National Institute of Informatics. His research focuses on the application of image processing algorithms and image database systems to real world problems, including bioimage informatics.