The Giga-pixel Challenge: Full Resolution Image Analysis – Without Losing the Big Picture

An open-source approach for multi-scale analysis and visualization of slide-scanner data

Petter Ranefall[1,2], Alexandra Pacureanu[1,2], Christophe Avenel[1], Anne E. Carpenter[3], Carolina Wählby[1,2,3]
[1]Centre for Image Analysis, Uppsala University
[2]Science for Life Laboratory, SciLifeLab, Uppsala University
[3]Broad Institute of Harvard and MIT
Email: petter.ranefall@it.uu.se, {alexandra.pacureanu, christophe.avenel, carolina}@cb.uu.se, anne@broadinstitute.org

***Abstract* — The amount of image data in a slide scanner image is usually in the giga-pixel range, inducing challenges for image analysis and visualization. At the same time, in order to maximize the information gained in these experiments and integrate expert knowledge with quantitative measurements, it is crucial to maintain the connection between high-resolution per-cell and per-region metrics with lower resolution relational metrics. We present a free and open-source framework for full resolution image analysis of large images, e.g. slide scanner data, with the possibility of visual examination and interaction at multiple resolutions. The interface enables seamless zooming and panning, with the option to toggle multiple layers, such as segmentation masks and classification results, on or off. We make use of the strength and flexibility of existing state-of-the-art open-source software for visualization, creating resolution pyramids, image registration and image analysis.**

Keywords—multi-scale image analysis, slide scanner data, open source

I. INTRODUCTION AND BACKGROUND

Digital imaging in pathology is currently experiencing exponential growth and expansion catalyzed by more advanced imaging hardware and computational processing techniques. Thin slices of tissue mounted on glass slides can be digitized at near the optical resolution limits of light in only a few minutes' time [5], [6]. Novel staining techniques [4], [15] and molecular detection methods [3], [7], [9], [16] make it possible to record the localization of many types of biomedical processes by fluorescence and bright-field microscopy in parallel.

Modern signal processing algorithms make it possible extract rich localized information visualizing spatial heterogeneity of gene and protein expression, cell morphology, localized relational patterns and tissue structures. Digital imaging using slide scanners thus has the potential to integrate traditional histopathological staining approaches and expert visual examination with advanced digital image metrics and advanced staining protocols, providing unique spatially resolved information of tissue samples at multiple resolutions.

The amount of image data in a slide scanner image is usually in the giga-pixel range, inducing challenges for image analysis and visualization. At the same time, in order to maximize the information gained in these experiments and integrate expert knowledge with quantitative measurements, it is crucial to maintain the connection between high-resolution per-cell and per-region metrics with lower resolution relational metrics. This is a severe limitation for existing analysis software (commercial as well as free and open-source) as few can handle analysis of full giga-pixel data. Several commercial systems providing multi-resolution sample viewing with image processing are starting to appear (some examples include Definiens, Indicalab, Visisopharm, Aperio). However, every tissue type, combination of biomarker, and disease process will pose unique analytical questions, meaning that customized solutions will be needed. Commercial solutions are also difficult to share between research labs, limiting reproducibility and comparison of research results. We believe there is a need for free and open-source tools for handling this giga-pixel challenge.

Rather than implementing analysis algorithms and metrics from scratch, we make use of the strength and flexibility of existing state-of-the-art open-source software for the image analysis, benefitting from the interoperability between primarily CellProfiler [A] for single-cell metrics Fiji [C] and Image J [E] for image filtering, image segmentation and feature extraction, and Ilastik [D] for texture-based classification based on machine learning. However, performing high-resolution image analysis on large slide-scanner images using these software packages is limited on today's desktop computers due to memory limitations. At the same time, we would like to visualize the results of the analysis in the context of the full tissue slide so that for example phenotypic variations at the subcellular level can be related to lower-resolution structures such as vessels, ducts, or tumors in the tissue. Here we present a free and open-source framework for visual examination at multiple resolutions with the option to toggle results on or off, such as segmentation masks, classification results, and tissue morphology measurements, using a map view with seamless zooming and panning capabilities, allowing for fast navigation between a full-tissue view and high-resolution sub-cellular observations. The interface also enables visual/manual selection of regions of interest, target discovery, and understanding of novel spatial relationships within the tissue environment.

II. OVERVIEW

Initially we split the image into smaller tiles that are more suitable for high-throughput analysis using existing software. We register images from multiple imaging mo-

dalities/staining cycles and keeping track of the coordinates of the tiles. Thereafter we analyze the data and present analysis results in the form of color-coded maps and/or coordinates for specific signals. Using the same approach as is currently widely available for viewing maps of the world overlaid with different kinds of labels and imaging modalities we can display the results of the analysis on the original, full-size image at multiple resolutions.

III. ANALYSIS FRAMEWORK

The input slide scanner images, which in our application are about 30000×40000 pixels large, are divided into tiles forming a resolution pyramid (see Fig. 1) using VIPS [H], a free image processing system that is designed to handle large images (images larger than the amount of RAM you have available).

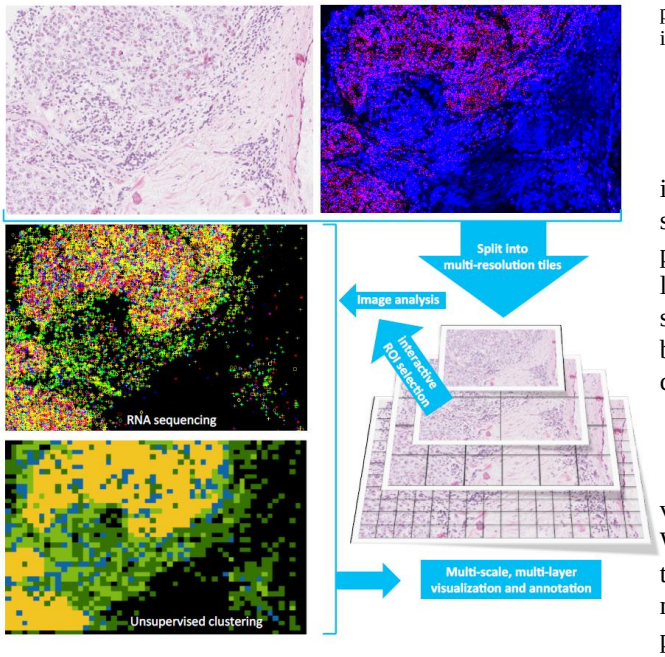

Fig. 1. Illustration of the work flow. The input images are split into tiles, that are used for both image analysis and visualization/interaction. The output from the image analysis can be viewed as overlay and annotations in the viewer.

Optionally fully- or semi-automated registration of images from different staining cycles and/or imaging modalities can be applied, using Elastix [B] or a manual segmentation tool we created based on OpenSeadragon [G] (see Fig. 2). In the image analysis pipeline a local fine registration is also used [14].

The image analysis is performed on the tiles separately and resulting regions of interest are put together to a resolution pyramid using a Python [I] script. When selecting the tile size there is a trade-off between speed when zooming and panning, and suitability for image analysis. The smaller the tiles, the faster zooming and panning, but also more edge artifacts in the image analysis. In our case tile sizes in the range 1000 to 2000 pixels seem to be suitable. We are using tiles without overlap, but both the tiling and visualization tools are able to handle overlap. In case of overlap we only need to make sure that from the image analysis pipeline handles the overlapping tiles in a consistent manner.

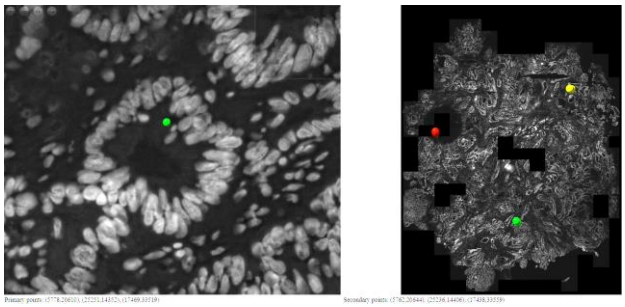

Fig. 2. Manual selection of control points for semi-automated image registration in the OpenSeadragon interface. Left is the reference image (zoomed) and right is the floating image. The user marks at least three pairs of corresponding points in the two images. Using those points as input a transformation is computed and applied.

IV. IMAGE ANALYSIS

In our application we have used the CellProfiler [A] image analysis software, which is a free and open-source software designed for high-throughput image-based experiments. The CellProfiler pipeline can be run in headless mode using a Python command, which makes it possible to start the image analysis process from a web browser interface. Batch processing using a computer cluster is also possible.

V. DATA ANALYSIS AND MAPPING

The end goal is to enable statistical analysis as well as visual presentation of quantitative image analysis results. We therefore translate feature measurements (classification outputs, relational information etc.) to color-coded maps representing regions with similar features. Our applications include *in situ* sequencing of RNA [9], and we therefore also use methods for translating measurements of localized molecular markers to a format suitable for multi-resolution viewing.

VI. VISUALIZATION AND DATA EXPLORATION

The visualization is again based on OpenSeadragon [G] that allows us to efficiently zoom and pan data, and toggle on and off different imaging modalities and analysis results in multiple combinations of overlays. The viewing tool also incorporates manual drawing of regions of interest for further data exploration (see Fig. 3).

Manual selection of regions of interest opens up a new way of exploring the heterogeneity of tissue data. Data exploration integrates the expert's knowledge with quantitative measurements and can reveal e.g. differences in gene expression within tumors and in the surrounding stromal compartment.

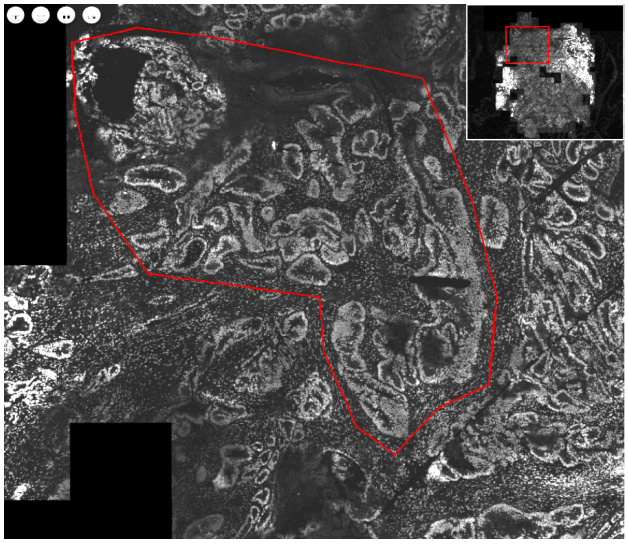

Fig. 3. Manual delineation of region of interest in the OpenSeadragon interface. The region of interest will be used for further analysis.

VII. Conclusion

We have presented a framework that enables exploration of different approaches for tissue analysis and makes the quantitative analysis readily compatible with visual analysis, bridging the gap between histopathology and advanced digital image analysis in a user-friendly, free and open-source environment incorporating existing tools. We believe this will leverage progress by ease of methods dissemination, shared development environments and reproducible science. The framework can of course also be applicable to many other types of large-scale 2D image data, and is not limited to slide-scanner data.

Acknowledgements

The authors would like to thank Mats Nilsson and his team for the collaboration regarding the in situ image analysis project, Ian Gilman, Antoine Vandecreme and Mark Salsbery for OpenSeadragon support, and John Cupitt for VIPS support.

List of used tools

[A] CellProfiler - a free and open-source software designed for high-throughput image analysis [2], [8]. http://cellprofiler.org/

[B] Elastix - a toolbox for rigid and nonrigid registration of images. http://elastix.isi.uu.nl/

[C] Fiji - Fiji is a distribution of ImageJ (and ImageJ2) together with Java, Java3D and a lot of plugins [11]. http://fiji.sc/Fiji

[D] Ilastik - a simple, user-friendly tool for image classification and segmentation [13]. http://www.ilastik.org/

[E] Image J - a public domain Java based image processing program [1], [12]. http://rsbweb.nih.gov/ij/index.html

[F] JavaScript - an interpreted computer programming language and a part of web browsers.
https://developer.mozilla.org/en-US/docs/Web/JavaScript

[G] OpenSeadragon - An open-source, web-based viewer for zoomable images, implemented in pure JavaScript. http:/openseadragon.github.io/

[H] PhP - a server-side scripting language designed for web development but also used as a general-purpose programming language. http://php.net/

[I] Python - a widely used general-purpose, high-level programming language. http://python.org/

[J] VIPS - an open source image processing software package that is designed to handle large images, and works with multi-core processors [10].
http://www.vips.ecs.soton.ac.uk/index.php?title=VIPS

References

[1] Abramoff MD, Magalhaes PJ, Ram SJ. Image processing with ImageJ. Biophotonics International. 2004;11:36–42.

[2] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006;7(10):R100. Epub 2006 Oct 31. PubMed PMID: 17076895; PubMed Central PMCID: PMC1794559.

[3] Clausson CM, Allalou A, Weibrecht I, Mahmoudi S, Farnebo M, Landegren U, Wählby C, Söderberg O. Increasing the dynamic range of in situ PLA. Nat Methods. 2011 Oct 28;8(11):892-3. doi: 10.1038/nmeth.1743. PubMed PMID: 22036742.

[4] Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, Can A, Corwin A, Dinn S, Filkins RJ, Hollman D, Kamath V, Kaanumalle S, Kenny K, Larsen M, Lazare M, Li Q, Lowes C, McCulloch CC, McDonough E, Montalto MC, Pang Z, Rittscher J, Santamaria-Pang A, Sarachan BD, Seel ML, Seppo A, Shaikh K, Sui Y, Zhang J, Ginty F. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A. 2013 Jul 16;110(29):11982-7. doi: 10.1073/pnas.1300136110. Epub 2013 Jul 1. PubMed PMID: 23818604; PubMed Central PMCID: PMC3718135.

[5] Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol. 2013 Jan 24;8:331-59. doi: 10.1146/annurev-pathol-011811-120902. Epub 2012 Nov 15. Review. PubMed PMID: 23157334.

[6] Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. IEEE Rev Biomed Eng. 2009;2:147-71. doi: 10.1109/RBME.2009.2034865. Epub 2009 Oct 30. Review. PubMed PMID: 20671804; PubMed Central PMCID: PMC2910932.

[7] Jarvius M, Paulsson J, Weibrecht I, Leuchowius KJ, Andersson AC, Wählby C, Gullberg M, Botling J, Sjöblom T, Markova B, Ostman A, Landegren U, Söderberg O. In situ detection of phosphorylated platelet-derived growth factor receptor beta using a generalized proximity ligation method. Mol Cell Proteomics. 2007 Sep;6(9):1500-9. Epub 2007 Jun 12. PubMed PMID: 17565975.

[8] Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, Ljosa V, Rueden C, Eliceiri KW, Carpenter AE. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics. 2011 Apr 15;27(8):1179-80. doi: 10.1093/bioinformatics/btr095. Epub 2011 Feb 23. PMID: 21349861.

[9] Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, Nilsson M. In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods. 2013 Sep;10(9):857-60. doi: 10.1038/nmeth.2563. Epub 2013 Jul 14. PMID: 23852452.

[10] Martinez, K and Cupitt, J (2005) VIPS — a highly tuned image processing software architecture. In, IEEE International Conference on Image Processing, Genova, IEEE, 574-577.

[11] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jun 28;9(7):676-82. doi: 10.1038/nmeth.2019. PubMed PMID: 22743772.

[12] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul;9(7):671-5. PubMed PMID: 22930834.

[13] Sommer C, Straehle CN, Koethe U, Hamprecht FA. ilastik: interactive learning and segmentation toolkit. Proc IEEE Int Symp Biomed Imaging ISBI 2011.

[14] Thévenaz P, Ruttimann UE, Unser M. A pyramid approach to subpixel registration based on intensity. IEEE Trans Image Process. 1998;7(1):27-41. doi: 10.1109/83.650848. PubMed PMID: 18267377.

[15] Wählby C, Erlandsson F, Bengtsson E, Zetterberg A. Sequential immunofluorescence staining and image analysis for detection of large numbers of antigens in individual cell nuclei. Cytometry. 2002 Jan 1;47(1):32-41. PubMed PMID: 11774347.

[16] Zieba A, Wählby C, Hjelm F, Jordan L, Berg J, Landegren U, Pardali K. Bright-field microscopy visualization of proteins and protein complexes by in situ proximity ligation with peroxidase detection. Clin Chem. 2010 Jan;56(1):99-110. doi: 10.1373/clinchem.2009.134452. Epub 2009 Nov 19. PubMed PMID: 19926775.