2015 IEEE 28th International Symposium on Computer-Based Medical Systems

INTERACTIVE SEGMENTATION RELABELING FOR CLASSIFICATION OF WHOLE-SLIDE HISTOPATHOLOGY IMAGERY

Anoop Haridas, Filiz Bunyak, Kannappan Palaniappan

Computational Imaging and VisAnalysis Lab
University of Missouri
Columbia, MO USA 65211

ABSTRACT

Collecting ground-truth or gold standard annotations from expert pathologists for developing histopathology analytic algorithms and computer-aided diagnosis for cancer grading is an expensive and time consuming process. Efficient visualization and annotation tools are needed to enable ground-truthing large whole-slide imagery. KOLAM is our scalable, cross-platform framework for interactive visualization of 2D, 2D+t and 3D imagery of high spatial, temporal and spectral resolution. In the current work KOLAM has been extended to support rapid interactive labeling and correction of automatic image classifier-based region labels of the tissue microenvironment by pathologists. Besides annotating regions-of-interest (ROIs), KOLAM enables extraction of the corresponding large polygonal image sub-regions for input into automatic segmentation algorithms, single-click region label reassignment and maintaining hierarchical image sub-regions. Experience indicates that clinicians prefer simple-to-use interfaces that support rapid labeling of large image regions with minimal effort. The incorporation of easy-to-use tissue annotation features in KOLAM makes it an attractive candidate for integration within a multi-stage histopathology image analysis pipeline supporting assisted segmentation and labeling to improve whole-slide imagery (WSI) analytics.

Index Terms— histopathology, tissue microenvironment, interactive segmentation, whole-slide imagery, supervised classification, visualization

1. INTRODUCTION

Clinical decision support systems (CDSSs) play a crucial role in the clinical process, from diagnosis and investigation to treatment and long-term care [1]. CDSSs are currently facing a deluge of image data from multiple sources including radiology and pathology. Advances in imaging techniques, high-throughput technologies, pathology informatics, bioinformatics, need for exploratory analysis of WSI and personalized medicine data sets continue to generate big data volumes in need of immediate analysis. The web-based Cancer Digital Slide Archive [2] is one such repository, that integrates imaging, genomic and clinical datasets.

The approach of manually labeling patient biopsy slides for automatic algorithm development is not scalable with the burgeoning volume of biopsy data, due to subjectivity of interpretation by pathologists, fatigue and likelihood of errors. The expert variability in the visual inspection process is quite high with up to a 20 percent discrepancy between different reviewers [3]. By providing efficient methods for objective quantification of image features, CDSSs have emerged as useful decision-making tools involving biomedical imaging.

In this paper, we focus on an improved human-computer interface for rapid annotation of the tissue microenvironment in histopathology WSI using a combination of automatic image classification with manual expert region label correction. Our system is focused on further investigating the correlation between microenvironment image features derived from stromal and epithelial regions and patient survivability beyond five years post-diagnosis [4]. We apply automated image analysis methods to identify and distinguish between stromal and epithelial regions using hematoxylin and eosin (H&E) stained histopathology digital slides [5][6][7][8][9]. KOLAM[10][11][12][13], our cross-platform and scalable digital light table or virtual microscope framework has been extended to support visualization and fast, reproducible semi-supervised annotation of histopathology slides. KOLAM also provides interfaces for (semi-)automated labeling and manual correction or relabeling of stromal, epithelial and other regions of interest in the tissue microenvironment.

2. RELATED WORK

Several new methodologies focusing on WSI data analysis have been developed over the past few years [14][15][16] [17][18][19]. Another aspect of handling WSIs is the increasing number of custom data formats. Increasing clinical and commercial interest in digital (histo)pathology alongwith the current non-standardized proprietary large image formats has led to the Digital Imaging and Communications in Medicine (DICOM) Standards Committee proposing Supplement 145 for WSI, compatible with current enterprise-wide

2372-9198/15 $3.00 © 2015 IEEE
DOI 10.1109/CBMS.2015.89

Picture Archiving and Communication System (PACS) software widely used by hospital information systems for radiology imagery [20]. The DICOM standard for histopathology will enable the electronic display, sharing, storage and management of large multi-resolution multi-planar WSI in a tiled pyramid image format. Each tile can be compressed using JPEG, JPEG2000 or other compression and coding methods. The OME Remote Objects (OMERO) [21] software platform is a collaborative effort towards building open-source data access and interoperability tools for WSI data. WSI handling also requires capable visualization systems. Popular examples are Imaris [22], Fiji [23], ePathology (Aperio) ImageScope [24], and Ventana Virtuoso and PathXchange [25].

3. KOLAM FOR HISTOPATHOLOGY VISANALYSIS

KOLAM is a large data visualization environment with robust support for out-of-core multiresolution tiled imagery across application domains [10][11][12][13]. In addition to the functionality of a virtual microscope [19] or light table, KOLAM has been extended to support WSI data and is currently being used for generating ground-truth or supervised region labels for input to image classification training algorithms. The suitability of KOLAM for WSI (big data) handling stems from the following aspects of WSI. Firstly, WSI come with a wide variety of image and metadata formats, with a variety of format specifications and proprietary aspects; secondly, issues pertaining to the large images, of the order of tens of gigabytes; and finally, the need for an intuitive yet efficient human computer interface for interactive visualization, navigation, manipulation and labeling. The KOLAM interface environment aims to provide the pathologist with easy-to-use visualization and annotation capabilities in a way that captures the expert's knowledge that is scalable [21]. The KOLAM GUI facilitates and enhances opportunities for scientists and experts to make new inferences using the rich feature set of visanalysis tools, that would otherwise be difficult.

3.1. Big Data Out-of-Core Visualization

KOLAM handles big data imagery by exploiting a multi-scale tiled memory efficient data structure [13] and out-of-core cache management strategies [10] as shown in Figure 1. At the heart of KOLAM is a multi-threaded image-rendering engine, capable of rapidly displaying multiple images and information layers as well as a temporal collection of multi-image (sequence) layers for display. The partitioning of each image into tiles allows fast, scalable random access to spatially coherent regions of the imagery across resolutions. This permits on-demand access to arbitrary regions and scales of the image needed for display, as well as application-level out-of-core management of memory [10]. Multiresolution (pyramidal) tiling allows for the interactive zooming of the image across scale [13]. Use of thread synchronization and

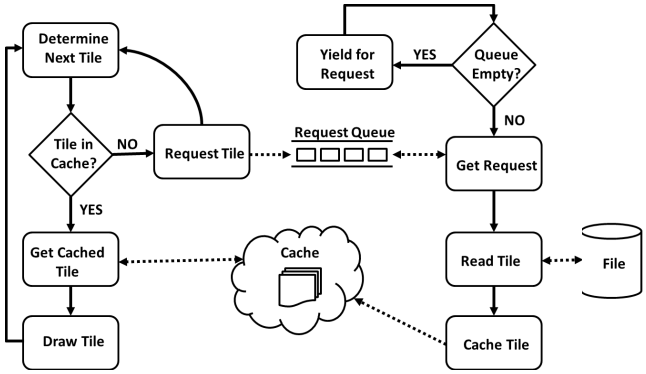

Fig. 1. KOLAM's multi-threaded image rendering engine, showing the roles and interactions of the tile cache, tile request queue, display and reader threads for managing large multiscale tiled imagery with multiple layers of information including segmentation and classifier label annotations.

CPU load balancing schemes allow KOLAM to manage out-of-core I/O and CPU demand patterns for visualizing large, tiled imagery. Furthermore, KOLAM's spatio-temporal dual cache system [10] permits the interactive visualization and playback of multiresolution image sequences. By using KOLAM features such as overlays, altering projection parameters, or interactively modifying the image resolution level the user can generate multiple views of the data by composing and organizing it in different ways. Additionally, custom view settings for data generated by one user may be stored and thus shared with other users of the WSI.

3.2. Support for Multiple WSIs and Metadata Types

A number of image formats capable of supporting the huge dimensions and metadata requirements of WSIs have emerged recently. The Aperio '.svs' format is being supported in microscopy software from multiple vendors. The BigTIFF format can also handle the large sizes of WSI bio-images. KOLAM supports these as well as the formats provided through the LOCI Bio-Formats Java library [26] by interfacing with the appropriate readers. KOLAM provides a uniform interface for access to metadata and can query or update metadata information for all supported image file formats.

3.3. Interactive Region-of-Interest (ROI) Selection

The first step prior to ground-truth generation involves the clinicians interactively visualizing the WSI data, and annotating ROIs (ie. regions with abnormal tissue characteristics that may be cancerous) via KOLAM's overlay drawing planes. These annotations are saved in XML format and may be reloaded for later use or shared between users. These regions are sometimes very large themselves (*e.g.* a WSI image of 130K$\times$100K pixels, with ROIs of sizes 30K$\times$20K pixels).

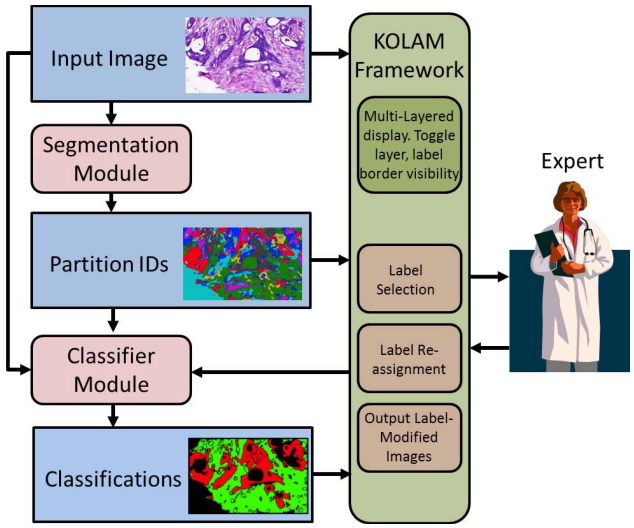

Fig. 2. KOLAM's role in the training and testing steps of the classifier algorithm. The input image and corresponding partition IDs are provided to the classifier module, which the outputs new tissue microenvironment labels. The expert relabels incorrectly labeled classes. These corrections are input to the classifier module for training.

In order to focus more closely on select areas within these ROIs, and to keep the burden of computational complexity as small as possible, the ROIs are further partitioned into image chips for individual processing. In the following steps, KOLAM is used first to assign new tissue labels for some of the segments generated by the applied segmentation algorithm. Subsequently, KOLAM is used to relabel some of the tissue classes assigned by the classifier algorithm. The details of the training and testing steps are presented below.

3.4. Ground-Truth Creation and Relabeling

The complexity of the ground-truth collection process is typically governed by (a) the quality of the imaging system, (b) standardization of data access [21], (c) clinical dignostic requirements, and (d) access to clinical patient data. The process of segmentation relabeling for classification involves the two stages of training and testing. Designation of ground-truth tissue microenvironment labels is the focus of the training step in the processing pipeline. In Figure 2, the training step includes every module except the classifier module and its input, which is solely part of the subsequent testing step. Prior to tissue label reassignment by the expert, the image chips need to be segmented and assigned preliminary labels. Our group has been working on various aspects of histopathology image segmentation ([5][6][7]) which we use to generate initial partitions. For the first training subimage, the expert sees the partition boundaries in a single class color that is overlaid on top of the underlying image. The pathologist can then change the color of any partition from epithelia to stroma or vice versa with just a few user interaction events. Tha manual assignment of ground-truth labels continues until the pathologist is satisfied with the labels assigned to partitions without having to tediously draw any boundary contours in the image. The partitions are essentially an unlabeled segmentation. These corrected tissue labels are then passed from KOLAM to the image region classifier module that can be applied to other test (sub)images with segmentations.

Prior to the testing step, the classifier module which received ground truth classes from KOLAM during the training step is trained on these classes in order to produce tissue labels for images outside of the original training image set. During the active learning testing phase (see Figure 2) the classifier generated tissue microenvironment labels on new (sub)images are used as input to the relabeling system. The expert visualizes these modified results, and may correct the annotations (generate ground truth) as necessary. Figure 3 illustrates an example of reassignment of a label or label annotation correction during the testing step. This shows the benefits of various KOLAM interface components supporting ground-truth region label annotations.

4. CONCLUSION

Pathologists have a strong and growing need for tools to lessen the burden involved in analyzing and annotating large numbers of whole-slide histopathology imagery for training CMSS/CAD systems, and for sharing results among experts using a web-based interface. Algorithm designers require a means to interactively visualize the results of their segmentation algorithms on the WSI data, a means to generate ground-truth for training classifiers and improving and fine-tuning the performance of other algorithmic modules. KOLAM provides capabilities to meet both computational and clinical needs, as well as to bridge the gap between pathologists and algorithm developers. KOLAM provides a virtual microscope interface supporting a variety of WSI formats along with a novel method for rapidly labeling and correcting the tissue microenvironment stroma-epithelia region labels. KOLAM thus provides visanalysis software tools for medical and computational experts to develop the evolving field of *computational histopathology*.

5. REFERENCES

[1] The OpenClinical Project, "Decision Support Systems," Online at: http://www.openclinical.org/dss.html, 2001.

[2] D. A. Gutman, J. Cobb, D. Somanna, Y. Park, F. Wang, T. Kurc, J. H. Saltz, D. J. Brat, L. Cooper, and J. Kong, "Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data," *J. American Medical Informatics Association*, vol. 20, no. 6, pp. 1091–1098, 2013.

[3] L. A. Teot, R. Sposto, A. Khayat, S. Qualman, G. Reaman, and D. Parham, "The problems and promise of central pathology review: development of a standardized procedure for the children's oncology

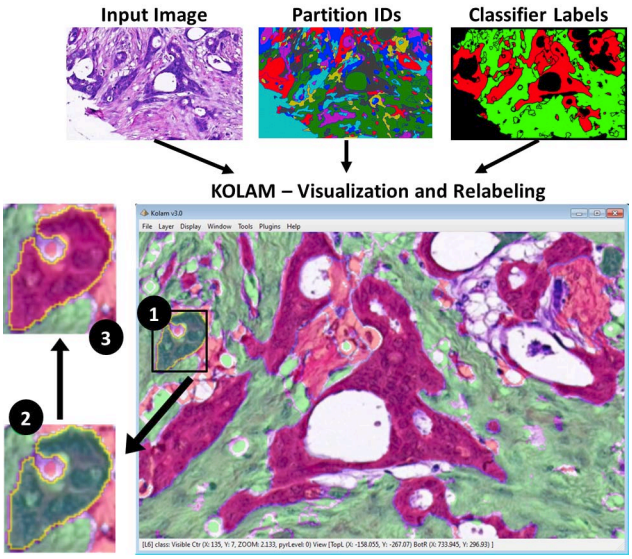

Fig. 3. An example of interactive relabeling of classifier-assigned tissue labels within the KOLAM visualization framework. The inputs to KOLAM for the testing step are shown. The labels for the stroma and epithelium, colored green and red respectively, are overlaid with partial transparency on the underlying imagery which is an image subregion from a breast cancer WSI. The tissue label in the inset box at the left, labeled '1', is being relabeled by the user. The label is zoomed up (Inset 2) on the far left, which clearly displays the boundary of the label. Finally, with one click, the expert changes the label, from stroma (Inset 2) to epithelia (Inset 3) class label.

group," *Ped. and Develop. Pathology*, vol. 10, no. 3, pp. 199–207, 2007.

[4] A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn, and D. Koller, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Science Translational Medicine*, vol. 3, no. 108, pp. 108–113, 2011.

[5] V. B. S. Prasath, F. Bunyak, P. Dale, S. R. Frazier, and K. Palaniappan, "Segmentation of breast cancer tissue microarrays for computer-aided diagnosis in pathology," in *IEEE Healthcare Innovation Conference*, 2012.

[6] F. Bunyak, A. Hafiane, and K. Palaniappan, "Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets," in *Software Tools and Algorithms for Biological Systems*, pp. 413–424. Springer, 2011.

[7] A. Hafiane, F. Bunyak, and K. Palaniappan, "Evaluation of level set-based histology image segmentation using geometric region criteria," in *IEEE Int. Symp. Biomedical Imaging*, 2009.

[8] A. Hafiane, F. Bunyak, and K. Palaniappan, "Fuzzy clustering and active contours for histopathology image segmentation and nuclei detection," *Lecture Notes in Computer Science (ACIVS)*, vol. 5259, pp. 903–914, 2008.

[9] A. Hafiane, F. Bunyak, and K. Palaniappan, "Clustering initiated multiphase active contours and robust separation of nuclei groups for tissue segmentation," in *IEEE Int. Conf. Pattern Recognition*, 2008, p. Online.

[10] J. Fraser, A. Haridas, G. Seetharaman, R. Rao, and K. Palaniappan, "KOLAM: A cross-platform architecture for scalable visualization and tracking in wide-area motion imagery," in *Proc. SPIE Conf. Geospatial InfoFusion III*, 2013, vol. 8747, p. 87470N.

[11] A. Haridas, R. Pelapur, J. Fraser, F. Bunyak, and K. Palaniappan, "Visualization of automated and manual trajectories in wide-area motion imagery," in *15th Int. Conf. Information Visualization*, 2011, pp. 288–293.

[12] K. Palaniappan, F. Bunyak, P. Kumar, I. Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. Rao, and G. Seetharaman, "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in *13th Int. Conf. Information Fusion*, 2010.

[13] K. Palaniappan and J. Fraser, "Multiresolution tiling for interactive viewing of large datasets," in *17th Int. AMS Conf. Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology*, 2001, pp. 338–342.

[14] R. Gutiérrez, F. Gómez, L. Roa-Peña, and E. Romero, "A supervised visual model for finding regions of interest in basal cell carcinoma images," *Diag. Pathology*, vol. 6, no. 1, pp. 14, 2011.

[15] D. Romo, E. Romero, and F. González, "Learning regions of interest from low level maps in virtual microscopy," *Diag. Pathology*, vol. 6, no. Suppl 1, pp. 8, 2011.

[16] V. Roullier, O. Lézoray, V-T. Ta, and A. Elmoataz, "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization," *Comp. Med. Imaging and Graphics*, vol. 35, no. 7, pp. 603–615, 2011.

[17] W. Jeong, J. Schneider, S. G. Turney, B. E. Faulkner-Jones, D. Meyer, R. Westermann, R. C. Reid, J. Lichtman, and H. Pfister, "Interactive histology of large-scale biomedical image stacks," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1386–1395, 2010.

[18] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 5, pp. 1205–1218, 2012.

[19] U. Catalyurek, M. D. Beynon, C. Chang, T. Kurc, A. Sussman, and J. Saltz, "The virtual microscope," *IEEE Trans. Information Technology in Biomedicine*, vol. 7, no. 4, pp. 230–248, 2003.

[20] R. Singh, L. Chubb, L. Pantanowitz, A. Parwani, et al., "Standardization in digital pathology: Supplement 145 of the DICOM standards," *Journal of Pathology Informatics*, vol. 2, no. 1, pp. 23, 2011.

[21] C. Allan, J. Burel, J. Moore, C. Blackburn, M. Linkert, S. Loynton, D. MacDonald, W. J. Moore, C. Neves, A. Patterson, et al., "OMERO: flexible, model-driven data management for experimental biology," *Nature Methods*, vol. 9, no. 3, pp. 245–253, 2012.

[22] Bitplane Inc., "Imaris v 8.0.1," Online at: http://bitplane.com, 2014.

[23] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al., "Fiji: an open-source platform for biological-image analysis," *Nature Methods*, vol. 9, no. 7, pp. 676–682, 2012.

[24] Leica Biosystems, "ePathology Imagescope," Online at: http://www.leicabiosystems.com/pathology-imaging/aperio-epathology/integrate/imagescope/, 2015.

[25] Roche (Ventana), "VENTANA Digital Pathology Solutions," Online at: http://www.ventana.com, 2015.

[26] M. Linkert, C. T. Rueden, C. Allan, J. Burel, W. Moore, A. Patterson, B. Loranger, J. Moore, C. Neves, D. MacDonald, et al., "Metadata matters: access to image data in the real world," *J. Cell Biology*, vol. 189, no. 5, pp. 777–782, 2010.