



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/89127/>

Version: Accepted Version

Article:

Liang, Y, Wang, F, Treanor, D et al. (2016) A framework for 3D vessel analysis using whole slide images of liver tissue sections. *International Journal of Computational Biology and Drug Design*, 9 (1/2). pp. 102-119. ISSN: 1756-0756

<https://doi.org/10.1504/IJCBDD.2016.074983>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A Framework for 3D Vessel Analysis using Whole Slide Images of Liver Tissue Sections

Yanhui Liang

Department of Biomedical Informatics,
Emory University,
Atlanta, GA, 30322 USA
E-mail: yanhui.liang@emory.edu

Fusheng Wang

Department of Biomedical Informatics,
Emory University,
Atlanta, GA, 30322 USA
E-mail: fusheng.wang@emory.edu

Darren Treanor

Department of Pathology Leeds Teaching Hospitals NHS Trust
Leeds Institute of Cancer and Pathology
The University of Leeds,
Leeds, LS9 7TF United Kingdom
E-mail: darrentreanor@nhs.net

Derek Magee

School of Computing,
The University of Leeds,
Leeds, LS2 9JT United Kingdom
E-mail: D.R.Magee@leeds.ac.uk

Nick Roberts

Leeds Institute of Cancer and Pathology
The University of Leeds,
Leeds, LS9 7TF United Kingdom
E-mail: n.j.roberts@leeds.ac.uk

George Teodoro

Department of Computer Science,

University of Brasília,
Brasília, DF, 70910-900 Brazil
E-mail: teodoro@unb.br

Yangyang Zhu

Department of Mathematics and Computer Science,
Emory University,
Atlanta, GA, 30322 USA
E-mail: yangyang.zhu@emory.edu

Jun Kong*

Department of Biomedical Informatics,
Emory University,
Atlanta, GA, 30322 USA
E-mail: jun.kong@emory.edu

* Corresponding author

Abstract: Three-dimensional (3D) high resolution microscopic images have high potential for improving the understanding of both normal and disease processes where structural changes or spatial relationship of disease features are significant. In this paper, we develop a complete framework applicable to 3D pathology analytical imaging, with an application to whole slide images of sequential liver slices for 3D vessel structure analysis. The analysis workflow consists of image registration, segmentation, vessel cross-section association, interpolation, and volumetric rendering. To identify biologically-meaningful correspondence across adjacent slides, we formulate a similarity function for four association cases. The optimal solution is then obtained by constrained Integer Programming. We quantitatively and qualitatively compare our vessel reconstruction results with human annotations. Validation results indicate a satisfactory concordance as measured both by region-based and distance-based metrics. These results demonstrate a promising 3D vessel analysis framework for whole slide images of liver tissue sections.

Keywords: Whole Slide Imaging; Digital Pathology; Pathology Image Analysis; 3D Vessel Structure; Liver Pathology.

Reference to this paper should be made as follows: Liang, Y.H., Wang, F.S., Treanor, D., Magee, D., Teodoro, G., Roberts, N., Zhu, Y.Y., and Kong, J. (2014) 'A Framework for 3D Vessel Analysis using Whole Slide Images of Liver Tissue Sections', *Int. J. Computational Biology and Drug Design*, Vol. x, No. x, pp.xxx-xxx.

Biographical notes: Yanhui Liang is a PhD student on Biomedical Informatics track in the Department of Mathematics and Computer Science at Emory University.

Fusheng Wang is an assistant Professor at Department of Biomedical Informatics at Emory University. His research interests include Big Data Management and Analytics, Spatial and Temporal Data Management and Analytics, Medical

*A Framework for 3D Vessel Analysis using Whole Slide Images of Liver Tissue Sections*³

Imaging Informatics, Heterogeneous Data Management and Integration, Clinical Natural Language Processing, and Data Semantics and Standardization.

Darren Treanor is a consultant liver and gastrointestinal pathologist at Leeds Teaching Hospitals NHS Trust, UK, and an honorary senior lecturer at the Leeds Institute of Molecular Medicine, University of Leeds.

Derek Magee is a lecturer within the School of Computing at the University of Leeds. His main research interests are in the medical domain, relating to medical image/volume (MRI, CT, PET and ultrasound) analysis and surgical simulation, and in visual surveillance, analysis of remote sensing data, and visual learning.

Nick Roberts is a research student in Leeds Institute of Cancer and Pathology at the University of Leeds. He works in the Digital Pathology Team specializing in the 3D reconstruction and 3D modeling of a range of different tissues.

George Teodoro holds a PhD Degree in Computer Science from the Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, and a MS Degree from the same university. His primary areas of expertise include high performance computing runtime systems to support efficient execution of biomedical and data-mining applications on distributed heterogeneous environments.

Yangyang Zhu is a Master student on Computational Science track in the Department of Mathematics and Computer Science at Emory University.

Jun Kong is an assistant Professor at Department of Biomedical Informatics at Emory University. His research interests include biomedical image analysis, computer-aided diagnosis, machine learning, whole slide image processing, computer vision, bioimaging informatics, and signal processing for large-scale biomedical translational research.

This paper is a revised and expanded version of a paper entitled 'A Framework for 3D Vessel Analysis with Whole Slide Images of Serial Liver Sections' presented at 'International Conference on Intelligent Biology and Medicine (ICIBM 2014), San Antonio, TX USA, December 4-6, 2014'

1 Introduction

With the rapid and significant advances in large-throughput scanning technologies, digital pathology whole slide imaging has recently become an emerging tool not only promising for disease diagnosis and treatment evaluation, but also for high-throughput quantitative information extraction for a wide scope of diseases, ranging from brain (Han et al., 2011; Kong et al., 2011), breast (Petushi et al., 2006), lung (Xing and Yang 2013), colorectal, neuroblastoma (Kong et al., 2009), lymphoma (Cooper et al., 2009), to prostate cancer (Jafari-Khouzani and Soltanian-Zadeh, 2003; Tabesh et al., 2007). Importantly, this new field presents salient merits that help researchers and clinicians better understand the underlying biological mechanisms of pathological evolutions and progressions of distinct diseases (Jara-Lazaro et al., 2010; Kong et al., 2012). Due to the large pathology microcopy imaging scale and overwhelming number of histopathology structures of interest to research, human-based reviewing approaches are not practically feasible for large data sets. In addition, the traditional labor-intensive reviewing process remains a qualitative discipline prone to significant intra-/inter-reader variations and poor reproducibility for diagnosis (Coons et

al., 2007). This results in a significant potential utility of whole slide image analysis and the quantitative characterization of pathological structures. A large suite of imaging analytical methods for quantitative analysis of pathology images has, therefore, been developed to meet these needs (Qi et al., 2012; Foran et al., 2011; Yang et al., 2008; Teodoro et al., 2013). In addition, commercial and open-source products are also available (Aperio, Definiens, ImageJ, among others). However, these tools are bounded by 2D biological structure analysis and limited for processing objects in 3D space.

As micro-anatomic objects in 2D pathology images are only approximate representations of their true 3D structures, information is substantially lost when 3D structures are projected to the 2D image focal plane. In liver disease diagnosis, for instance, structural changes in liver vessels and their spatial relationships in 3D are important (Roberts et al., 2012). 2D profiles of blood vessels in microscopy images of liver tissue are highly dependent on the location and angle of the cutting plane during tissue slide preparation and are substantially different from their 3D shapes in most cases. By contrast, a sequence of microscopy images of liver tissue slices allows for analysis of vessels in 3D space. Therefore, it presents better potential to improve vessel characterization accuracy and to support disease diagnosis and understanding.

Numerous studies on modeling and analysis of vessel structures have been proposed, ranging from vessel segmentation (Lesage et al., 2009), vessel structure tracking (Friman et al., 2009), to the 3D visualization of vascular structures (Kubisch et al., 2012). However, these investigations are limited to either radiology images (such as Computed Tomography, and Magnetic Resonance Imaging), or 3D contrast-enhanced imaging modalities (including Computed Tomography Angiography and Magnetic Resonance Angiography). Compared with radiology image modalities, digital pathology images have much higher resolutions and contain a wealth of information at the cellular level. Therefore, applying 3D vessel analysis methods for radiology images directly to whole slide images is not feasible. Other vessel studies involve the usage of Optical Coherent Tomography (OCT) and confocal microscopy imaging data (Hu et al., 2010; Arribas et al., 2007). Although OCT data modality is non-destructive and accurate for micro-vessels, it loses contextual information of digital pathology. Alternatively, confocal microscopy can also be used to accurately characterize vessels, but limited to small tissue samples.

In this paper, we present a complete 3D vessel analysis framework that can help researchers and clinicians better visualize 3D vessel structures and quantitatively characterize vessel features with 3D microscopy images of liver tissues. With access to a sequence of liver 2D microscopy images, we register all these images to allow for the following analysis in the same spatial coordinate system. Vessels, along with other key micro-anatomic objects of interest in pathology diagnosis, are segmented on each 2D microscopy image. The large number of vessels within livers poses major challenges to recover vessel shapes in 3D by human reviewers. With our computerized analysis, we use Fourier shape descriptors and spatial similarity of vascular structures as visual cues and formulate similarity functions for one-to-one, one-to-two, and one-to-none correspondence cases across adjacent tissue slides. The best multiple vessel associations are found by maximizing the resulting aggregated similarity function with constrained Integer Programming. The resulting mapped vessel cross sections are interpolated to create more intermediate virtual slides for better vessel characterization. Finally, each associated vessel structure is reconstructed and volumetrically rendered in 3D space. We apply our method to a liver microscopy image dataset with sequential tissue slides and reconstruct primary vessels for 3D visualization. By comparing machine-identified vessels with human

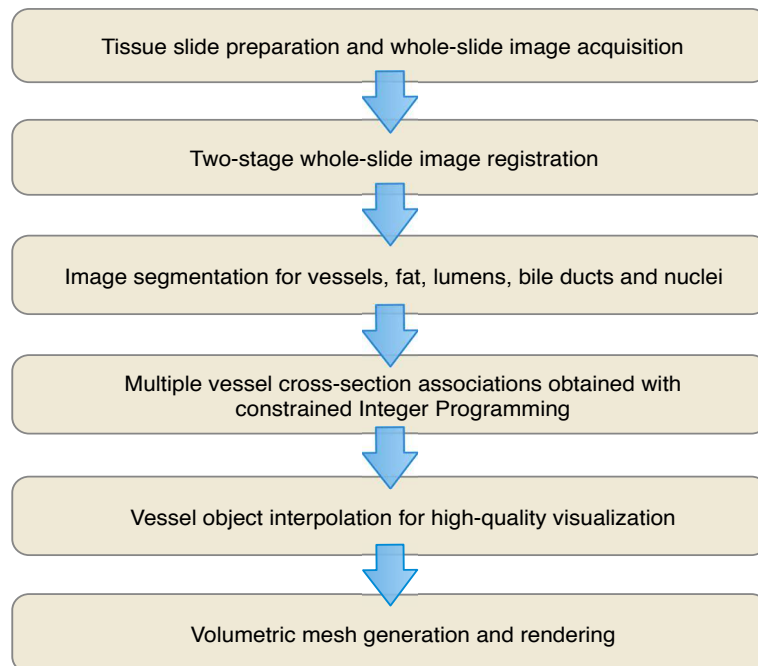
annotations, we observe a satisfactory concordance suggested by both region-based and distance-based measures. Meanwhile, visual inspections on vessel structures by our method and human annotations also present negligible difference. These results suggest the efficacy of our 3D vessel analysis method for whole slide images of liver tissue sections. Notably, our 3D vessel analysis framework is generic and can be readily applied and extended to the analytics of other 3D biological entities with 3D image datasets of diverse diseases.

The remainder of the paper is organized as follows. Data and Methods used for this study are described in Section 2. We then present and validate the 3D vessel analysis results in Section 3. In Section 4, we discuss how to apply this software to future integrative analysis linked with clinical and molecular endpoints and close the paper with conclusions.

2 Materials and methods

Our workflow for 3D vessel structure analysis consists of multiple components: image registration, vessel segmentation, vessel cross-section association, intermediate slide interpolation, and volumetric rendering. We present the overall workflow in Figure 1. Detailed discussions on these modules and the dataset involved in this study are given as follows.

Figure 1 The overall workflow of 3D vessel analysis with whole slide imaging data of liver tissue sections.



2.1 *Slide preparation and imaging data acquisition*

Whole slide images provide rich information about morphological and functional characteristics of biological systems. As a result, such image modality serves as a promising data source to provide insights on the underlying mechanisms of disease onset and progression, given the availability of scalable image analysis tools. In this study, we have access to a sequence of 12 whole slide images of liver sections separated by $50\mu\text{m}$ from the University of Leeds. Resected human liver samples are paraffin embedded and formalin fixed. Then the embedded samples are sectioned into slices with thin thickness. The resulting slices are dyed with dual chromogen immunohistochemistry (IHC) (Ismail et al., 2010). DAB IHC is directed against CD34 antigen as a blood vessel marker. Haemotoxylin nuclear counterstain is also used. Finally, these slides are scanned with ScanScope AT scanner (Aperio, <http://www.leicabiosystems.com/>) at 40x magnification and converted into digital images compressed with JPEG2000 and Matrox Imaging Library (MIL 8.0, <http://www.matrox.com/>).

The resulting images have high resolutions and large file sizes, typically with $75\text{k} \times 65\text{k}$ pixels and 300 Megabytes per image. Each whole slide image file consists of four image representations down sampled from the base image by 4:1, 16:1, 32:1, and 88:1 (thumbnail), respectively. Each resolution level of image representation has three 8-bit image channels. At the base level, the physical resolution is about $2.508\text{e-}1 \mu\text{m}$ per pixel. From computer-based analysis perspective, we would like to process images at a high resolution for sufficient tissue structure detail. By contrast, efficient analysis welcomes images of a low resolution. After our extensive tests, we reach the balance between image analysis quality and execution speed via applying our method to images down sampled from the base image level by 64:1.

Scanned whole slide images of liver slides stained by our staining process present blue, brown, red colors for nuclei, vessels and bile ducts, respectively. Locations for fat in the resulting liver microscopy images are visualized by bright white color. Lumens inside vessels show white in color as well, but surrounded with vascular walls in brown.

2.2 *Image registration and segmentation*

All slides in the same series are registered with a two-stage process. Initially, images are rigidly registered using phase correlation (Decastro et al., 1987). For non-rigid registration, the whole image is partitioned into a set of evenly spaced patches that are aligned separately with phase correlation. Then, B-Spline transformation is applied to the rigidly transformed image patches to estimate the global non-rigid pixel spatial change by a set of 64 control points (Roberts et al., 2012). Given the down sampled image resolution, more control points would only improve registration accuracy marginally. With this sequence of registered images, we have managed to automatically segment vessels, along with nuclei, fat, lumens, and bile ducts by morphology reconstruction segmentation method (Vincent, 1993) on each image channel found by color deconvolution technique (Ruifrok, 2001). The final vessel boundaries are smoothed.

2.3 *3D vessel association*

After blood vessels are identified by our analysis, we find some are recognized as numerous disconnected small parts. To group these disconnected pieces back to their intact vessel cross-sections, we perform intra-frame vessel object grouping, followed by vessel object filtering. Grouping is achieved by an image dilation operation (Vincent, 1993) that grows

small vessel parts belonging to the same vessel object and that forces them to be cohesive to a single vessel object. Among all grouped vessel objects, we proceed with selecting the top- K candidates of primary vessels for further analysis based on their sizes, as inclusion of unduly small vessels for analysis not only significantly increases the computational complexity for object association, but also reduces the vessel association accuracy.

The selected top- K vessel objects in each frame are characterized based on their shapes and spatial relationships. We define the problem of frame-by-frame object association in a constrained Integer Programming framework (Hillier and Lieberman, 2001). For our data, three distinct vessel association cases are considered: one-to-one (growing), one-to-two (bifurcation) and one-to-none (disappearing). For each case, the similarity function \mathcal{S} is defined as follows:

- one-to-one: the main vessel still continues to extend to the next frame;

$$\mathcal{S}(v_i^t, v_j^{t+1}) = \lambda_1 \mathbf{g}(v_i^t, v_j^{t+1}) + \lambda_2 \mathbf{d}(v_i^t, v_j^{t+1}) \quad (1)$$

- one-to-two: bifurcation occurs and the vessel grows into two small branches;

$$\mathcal{S}(v_i^t, v_{j_1}^{t+1}, v_{j_2}^{t+1}) = \lambda_1 \mathbf{g}(v_i^t, v_{j_1}^{t+1} \cup v_{j_2}^{t+1}) + \lambda_2 \mathbf{d}(v_i^t, v_{j_1}^{t+1} \cup v_{j_2}^{t+1}) \quad (2)$$

- one-to-none: the vessel stops growing;

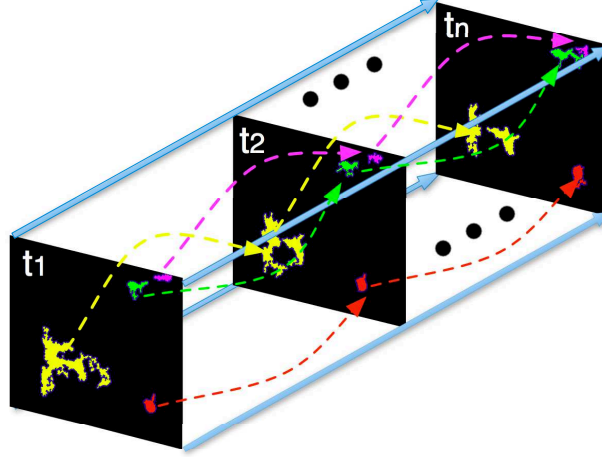
$$\mathcal{S}(v_i^t, v_{\emptyset}^{t+1}) = \mathbf{d}(v_i^t, \Omega_I) \quad (3)$$

where function $\mathbf{g}(\cdot)$ compares the similarity of vessel appearance by comparing the Fourier shape descriptors (Zahn et al., 1972) derived from vessel boundaries; $\mathbf{d}(\cdot)$ computes the spatial similarity of two vessel objects with their centroids; v_i^t is the i -th vessel object in frame t ; Ω_I is the boundary of image I ; λ_1 and λ_2 are two constant weights to control the association smoothness subject to $\lambda_1 + \lambda_2 = 1$.

Identification of vessel association is a process of establishing mappings between vessel objects in adjacent image frames based on given criteria and constraints. Assuming we have n_t and n_{t+1} vessel objects in frame t and $t + 1$ and h possible associations between the two frames, we deem such association identification process identical to a multi-object tracking problem (Jiang et al., 2007). Therefore, the optimal associations among vessel objects can be achieved by solving the constrained Integer Programming (Hillier and Lieberman, 2001), based on the pre-defined similarity function:

$$\begin{aligned} & \text{Arg max}_{\mathbf{x}} \quad \mathbf{s}^T \mathbf{x} \\ & \text{subject to} \quad (\mathbf{H}^T \mathbf{x})_i \leq 1 \quad \text{for } i = 1, \dots, n_t + n_{t+1} \\ & \quad \quad \quad x_j \in \{0, 1\} \quad \text{for } j = 1, \dots, h \end{aligned} \quad (4)$$

where \mathbf{s} is an $h \times 1$ vector with each entry representing the similarity of one vessel object association; \mathbf{H} is an $h \times (n_t + n_{t+1})$ binary matrix with H_{ij} set to 1 if and only if the i -th object association involves the j -th element in a set concatenated by n_t vessel objects from frame t and n_{t+1} vessel objects from frame $t + 1$; $(\mathbf{H}^T \mathbf{x})_i$ denotes the i -th element of $(\mathbf{H}^T \mathbf{x})$ and the constraint $(\mathbf{H}^T \mathbf{x})_i \leq 1$ guarantees that each vessel object in a given frame (i.e. t or

Figure 2 Vessel associations of four chains of color-coded vessels in frame t_1, t_2, \dots, t_n .

$t + 1$) can be selected at most once in the result; the optimal solution \mathbf{x} is an $h \times 1$ binary vector where $x_i = 1$ indicates the selection of the i -th association in the optimal solution. The algorithmic description of the above constrained Integer Programming with simplex solver is presented in Algorithm 1.

Solving the constrained Integer Programming problem above provides the optimal associations for vessel objects in adjacent frame pairs. Therefore, by tracking vessel cross-sections through all adjacent frames with the identified associations, we can recover the profile of a vessel structure along the z -axis with shape descriptors and spatial similarity. In Figure 2, vessel in yellow bifurcates, corresponding to the one-to-two association case. Other vessels (green, red, and magenta) represent one-to-one tracking cases.

2.4 Interpolation and visualization

Since there is a limited number of whole slide images for this liver imaging dataset ($n = 12$) and the z -axis resolution is not sufficient to support high-quality 3D rendering, we perform interpolation between each pair of associated vessel objects in adjacent frames. Initially, we normalize each boundary by its arc length. Then we uniformly sample points along each normalized boundary arc. Finally, we apply linear interpolation (Hazewinkel, 2001) to the sampled points.

Let's denote m sampled points from frames t and $t + 1$ as $(x_1^t, y_1^t), (x_2^t, y_2^t), \dots, (x_m^t, y_m^t)$ and $(x_1^{t+1}, y_1^{t+1}), (x_2^{t+1}, y_2^{t+1}), \dots, (x_m^{t+1}, y_m^{t+1})$, respectively. We linearly interpolate the x -coordinate and y -coordinate independently. Thus the i -th point $(x_i^{t+\delta}, y_i^{t+\delta})$ in frame $t + \delta$ ($0 < \delta < 1$) can be interpolated as:

$$\begin{cases} x_i^{t+\delta} = \frac{x_i^{t+1} - x_i^t}{t+1-t} (t + \delta - t) + x_i^t = (x_i^{t+1} - x_i^t) \delta + x_i^t \\ y_i^{t+\delta} = \frac{y_i^{t+1} - y_i^t}{t+1-t} (t + \delta - t) + y_i^t = (y_i^{t+1} - y_i^t) \delta + y_i^t \end{cases} \quad (5)$$

In this way, we get all point coordinates (x_i, y_i) by performing linear interpolation for each interpolated frame between frame t and $t + 1$. Combining the original dataset

Algorithm 1 Constrained Integer Programming with simplex solver for Vessel Object Association (VOA): $(f_I^*, \mathbf{x}^*) = VOA(\mathbf{x}, j, f_I)$ (Hillier and Lieberman, 2001)

Require: $(\mathbf{H}^T \mathbf{x})_i \leq 1$ for $i = 1, \dots, n_t + n_{t+1}$ and $x_j \in \{0, 1\}$ for $j = 1, \dots, h$

Ensure: $\text{Arg max}_{\mathbf{x}} s^T \mathbf{x}$

1: Suppose f_I and f_r are current objective values obtained by solving the Integer Programming(IP) problem and *LP*-relaxation problem, respectively; f_I^* is the optimal objective value and \mathbf{x}^* is the optimal integer solution

2: **Initialization:** $\mathbf{x} \leftarrow [0, \dots, 0]_{h \times 1}$, $j \leftarrow 1$, and $f_I \leftarrow \infty$

3: **Output:** $(f_I^*, \mathbf{x}^*) = VOA(\mathbf{x}, j, f_I)$

4: **function** $VOA(\mathbf{x}, j, f_I)$ \triangleright The definition of function $VOA(\mathbf{x}, j, f_I)$

5: **if** $j > h$ **then**

6: **return** (f_I, \mathbf{x})

7: **end if**

8: Perform *LP*-relaxation by releasing $x_j \in \{0, 1\}$ to a weaker constraint $0 \leq \tilde{x}_j \leq 1$

9: Solve the *LP*-relaxation problem: $(f_r, \tilde{x}_j) = LP_Relax(x_1, \dots, x_{j-1}, \tilde{x}_j, x_{j+1}, \dots, x_h)$

10: **if** $f_r > f_I$ and \tilde{x}_j is NOT an integer **then**

11: $x_j \leftarrow 0$; $(f_I^0, \mathbf{x}^0) = VOA(\mathbf{x}, j+1, f_I)$

12: $x_j \leftarrow 1$; $(f_I^1, \mathbf{x}^1) = VOA(\mathbf{x}, j+1, \max(f_I^0, f_I))$

13: **if** $f_I^0 \geq f_I^1$ **then**

14: **return** (f_I^0, \mathbf{x}^0)

15: **else**

16: **return** (f_I^1, \mathbf{x}^1)

17: **end if**

18: **else if** $f_r > f_I$ and \tilde{x}_j is an Integer but $\tilde{x}_j \neq x_j$ **then**

19: $f_I \leftarrow f_r$; $x_j \leftarrow \tilde{x}_j$

20: $(f_I, \mathbf{x}) = VOA(\mathbf{x}, j+1, f_I)$

21: **else if** $f_r < f_I$ **then**

22: $(f_I, \mathbf{x}) = VOA(\mathbf{x}, j+1, f_I)$

23: **end if**

24: **end function**

with the interpolated frames, we then render vessels in a 3D space. With interpolated 3D image volume, triangular isosurfaces with specified density are extracted by Constrained Delaunay tetrahedralization(CDT). The subvolumes bounded by the extracted isosurfaces are filled with tetrahedral elements and 3D mesh is generated with adaptive resolution. Specially, hollow structures or sub-domains corresponding to different tissue types are rendered according to their predefined labels in the volumetric data (iso2mesh, <http://iso2mesh.sourceforge.net>).

3 Results

We apply the proposed processing framework to a whole slide image dataset consisting of 12 sequential liver slides and extensively evaluate our approach implemented with C and Matlab. In this section, we present results from different modules in our framework and

demonstrate method performance measured by both quantitative and qualitative validation analysis.

3.1 Results of individual modules

The first processing module in the analysis pipeline is image registration that helps register to the first image all subsequent ones in the sequence. As this step provides the foundation to reconstruct 3D objects in a meaningful way with their 2D cross-sections from different coordinate systems, it has a profound impact on the following analysis. Our two-step registration method takes into account both global rigid and local non-rigid deformation with phase correlation and B-Spline interpolation by 64 control points. With our image resolution, 64 control points are sufficient to accurately recover non-rigid deformation at different regions in an image. Additional control points would only marginally improve registration accuracy. We present a typical registration result in Figure 3 where frame ($n = 10$) is registered to the reference image ($n = 0$). After registration, the registered image in Figure 3(right) is rotated in counterclockwise direction slightly, especially the upper and lower right corners. Nevertheless, not all parts of the registered image are rotated and shifted uniformly because of the non-rigid registration effect.

In liver microscopy images, histopathologic objects of interest to this study at the segmentation step include vessels, nuclei, fat, lumens, and bile ducts. Each type of object is identified and represented with distinct colors. A representative segmentation result is demonstrated in Figure 4, where brown, red, dark blue, light purple, and light yellow colors are used to represent vessels, bile ducts, nuclei, lumens, and fat in liver tissues, respectively. Since we use vessel shape for association analysis in this study, we obtain the vessel boundaries from their masks and superimpose green vessel boundaries on the original image in Figure 4.

In the proposed method, a constrained Integer Programming framework is employed to identify the optimal vessel object association between adjacent frames. The optimal solution is achieved by iteratively solving a series of Linear Programming (*LP*) relaxation problems. The parameters for object association used in this study are $K = 25$, $\lambda_1 = 0.68$

Figure 3 Sequential liver microscopy image registration. (Left): a microscopy image of a liver slice ($n = 10$) before registration; (Right): same image after registration (slightly rotated counterclockwise).

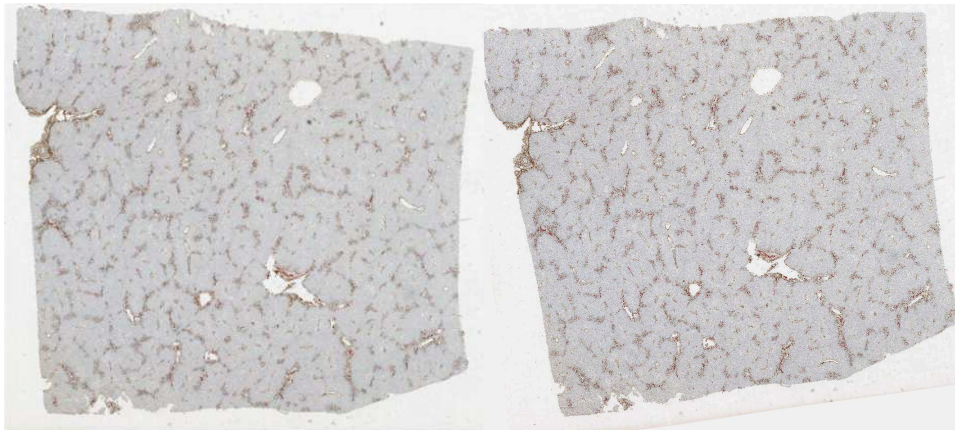
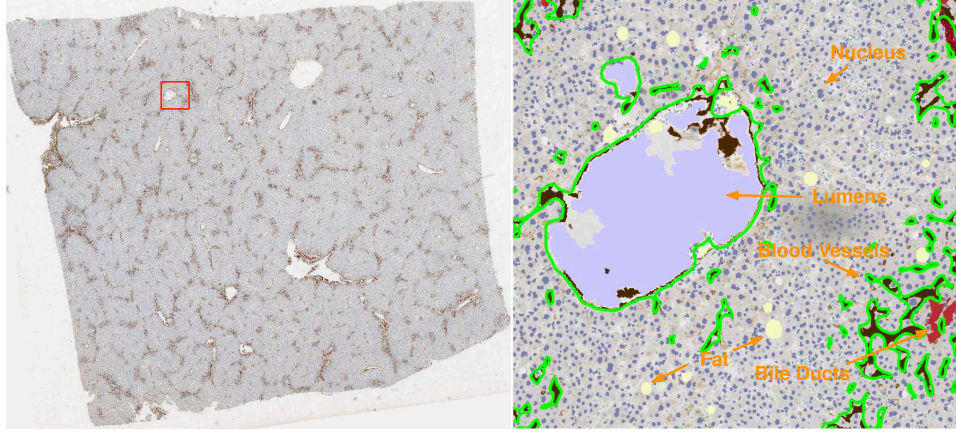


Figure 4 Liver microscopy image segmentation. (Left): a registered microscopy image of a liver slice ($n = 10$); (Right): a close-up view of segmentation result within the red box region overlaid on the left image. Note that brown, red, dark blue, light purple, and light yellow colors are used to represent vessels, bile ducts, nuclei, lumens, and fat in liver tissues; Additionally, vessel boundaries are overlaid in green.



in Equation (1), and $\lambda_1 = 0.35$ in Equation (2). We carefully choose these parameter values after intensive experiments. The value for λ_1 in Equation (1) is intentionally selected to be greater than that in Equation (2), in the sense that shape descriptors are more reliable and discriminative in the one-to-one than one-to-two case. When vessel bifurcates, shapes of the resulting vessel components are substantially deformed from their source. As a result, we assign more weight on spatial similarity for one-to-two case.

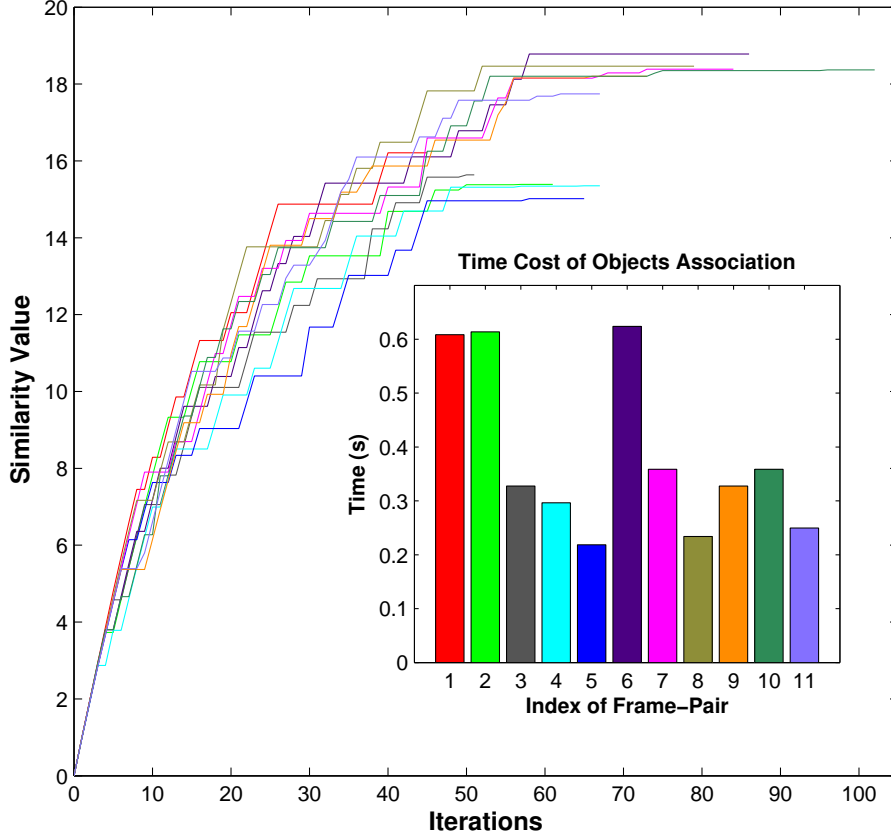
We plot objective similarity value for each adjacent frame pair as a function of iteration number in Figure 5, where the objective similarity function defined in Equation (4) increases as iteration number increases. Each curve in Figure 5 represents similarity function change of one frame pair, as indicated by a unique color. The optimal solution is achieved when the objective function reaches its maximum. In Figure 5, we also show time cost for finding the optimal vessel object association for each frame pair with the same color coding scheme of similarity function curves. Note that all time costs are less than 0.7 seconds, with the majority around 0.32 seconds. This suggests that it is efficient to use the constrained Integer Programming framework to establish vessel object association.

Interpolation used in our framework enables us to recover 3D vessel structures that exhibit a smooth transition process between two original frames. Without this step, the resulting reconstructed vessels present abrupt changes and step artifacts due to insufficient inter-frame resolution. We present a typical interpolation result in Figure 6, where the leftmost and the rightmost frames are original slices with two identified vessel objects in each. Six intermediate frames are evenly sampled from 50 interpolated frames.

Note that the interpolation error is bounded as follows (Hazewinkel, 2001):

$$\begin{aligned} error(k) &\approx f(k) - L(k) = \frac{f''(\xi_i)}{2} (k - k_1)(k - k_2) \\ &\leq \frac{l^2}{8} \left. \frac{f''(\xi^*)}{2} \right|_{\xi^* = \frac{k_1 + k_2}{2}} \end{aligned} \quad (6)$$

Figure 5 Similarity value as a function of iteration number, with an inset presenting time cost of vessel object association for each image pair in the sequence. Similarity function curves and time cost bars are coded with the same color scheme.



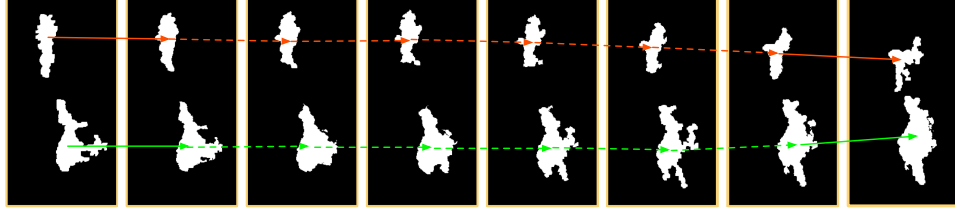
where $\xi_i \in [k_1, k_2]$, $i = 1, 2, \dots, P$, and $l = k_2 - k_1$. As a result, the interpolation error decreases as the interval l decreases and the number of sampled data points increases. This suggests that we can achieve improved accuracy by using more sampled points. In this study, the length of interval $[k_1, k_2]$ is normalized to $l = 1$ and the number of ξ_i is set to $P = 50$. With these parameter values, we observe continuously changed vessel cross sections as illustrated by two interpolated vessel chains in Figure 6.

As a last step in our pipeline, we render the matched and interpolated vessel objects in 3D space. After interpolation, each object is labeled with a unique integer. Coordinates of vertices on isosurfaces with same label values are recorded and the associations of these vertices for triangular isosurfaces are found. Then, tetrahedral elements are generated to fill up volumes contained by these isosurfaces and 3D mesh is generated with adaptive resolution. The 3D view of six primary vessel objects is presented in Figure 7(a).

3.2 Validation study

To validate the segmentation method, we quantitatively compare the machine-identified vessel segmentation result and human annotations on the same dataset consisting of 12 liver

Figure 6 Interpolation of intermediate frames. The leftmost and rightmost frames are original, while the six ones in the middle are interpolated.



tissue slides. With machine and manual results, we linearly interpolate segmentation results from adjacent frame pairs and validate our method on all interpolated frames. Specifically, region-based and distance-based measures are used for evaluation. Region-based measures used in our validation include Jaccard coefficient, precision, recall and F_1 score while Hausdorff distance is used as a distance-based metric.

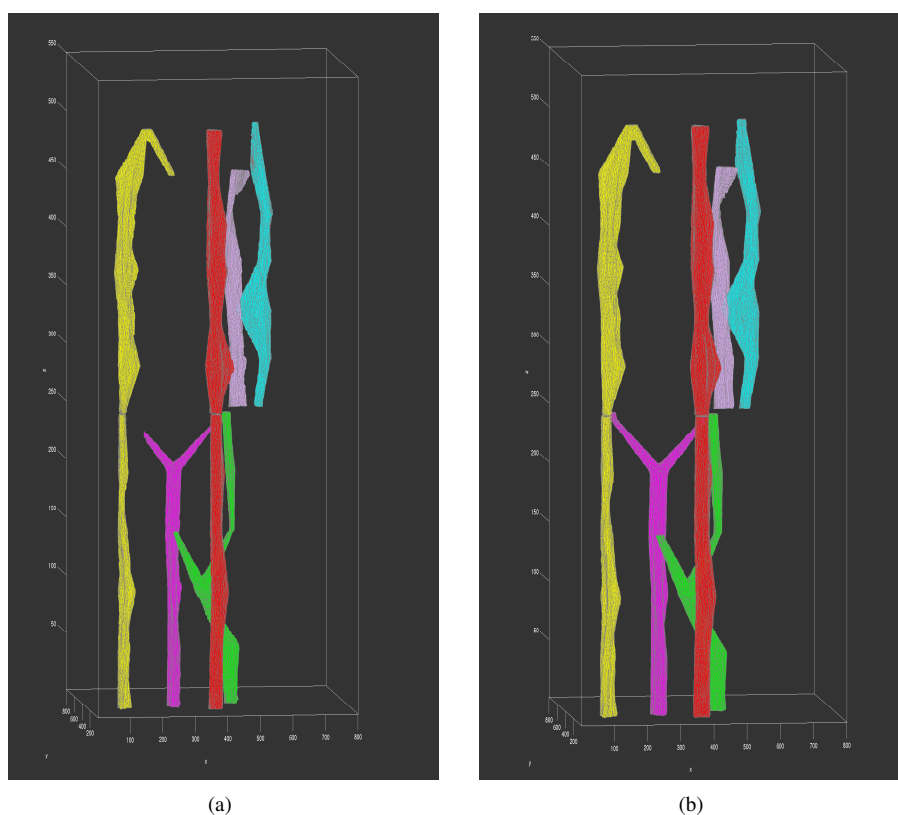
We take human annotations on each vessel as ground truth and denote them by A . We denote the machine segmentation results by B . For region-based measures, the Jaccard coefficient (J), precision (P), recall (R) and F_1 score (F_1) are defined as $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, $P = \frac{|A \cap B|}{|B|}$, $R = \frac{|A \cap B|}{|A|}$ and $F_1 = \frac{2PR}{P+R}$, respectively. Hausdorff distance between the ground truth and segmentation results is defined as $d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\}$

with \sup and \inf representing the supremum and the infimum, respectively. For each vessel, we present mean and relative standard error (RSE%) of each metric in Table 1. The best performance with each measure is in bold. The first column in Table 1 shows distinct vessel objects and their colors are demonstrated in Figure 7(a). Since vessel 1, 2 and 3 are more regular than vessel 4, 5 and 6 in shape, better segmentation agreement between the proposed method and human is achieved, as indicated in Table 1. Our validation results with all above measures suggest that our segmentation method has a satisfactory concordance with the manual segmentation results.

In addition to the quantitative segmentation method validation, we also take a qualitative assessment by visually comparing the reconstructed vessel objects in 3D view. We present the 3D vessel rendering results from human annotations in Figure 7(b). Reconstructed vessels in Figure 7(a) and Figure 7(b) look similar in 3D structure overall, although ones derived by human annotations have more regular surfaces. By contrast, the machine generated vessels tend to preserve more structural details.

Table 1 Evaluation of the segmentation results (mean \pm Relative Standard Error%) for primary vessels. Red (r), yellow (y), magenta (m), green (g), cyan (c) and purple (p) are used to represent the corresponding 3D objects in Figure 7(a). The best result with each measure is bold.

Vessel (color)	J	P	R	F_1	d_H
1 (r)	0.91 \pm 20%	0.87 \pm 32%	0.87 \pm 32%	0.87 \pm 32%	4.13 \pm 5%
2 (y)	0.77 \pm 17%	0.90 \pm 35%	0.91 \pm 35%	0.91 \pm 35%	4.24 \pm 3%
3 (m)	0.82 \pm 28%	0.92 \pm 33%	0.89 \pm 33%	0.90 \pm 33%	5.39 \pm 2%
4 (g)	0.76 \pm 20%	0.71 \pm 41%	0.76 \pm 41%	0.73 \pm 41%	4.14 \pm 6%
5 (c)	0.88 \pm 11%	0.70 \pm 37%	0.69 \pm 39%	0.70 \pm 38%	5.48 \pm 3%
6 (p)	0.68 \pm 50%	0.71 \pm 42%	0.73 \pm 42%	0.72 \pm 42%	4.24 \pm 4%

Figure 7 3D visualization result of (a) machine-identified and (b) human-annotated vessels.

4 Discussion and conclusion

As commercial whole slide imaging scanners are now capable of producing high-magnification, high-resolution images from whole slide tissues in a prompt manner, it increasingly becomes feasible even for a medium-scale study to routinely generate sequential whole slide images of tissue sections. The resulting high-resolution microscopy image data offers a rich source of tissue phenotypic information on clinical relevant histological features and spatial relationships among histopathology hallmarks in specimen architectures. However, 3D microscopy image data is still an underutilized data modality in biomedical research because of the following reasons:

- It is not technically straightforward to develop effective and efficient methods to extract useful information from large-scale whole slide image dataset;
- There is a lack of highly configurable image analysis software to meet diverse research demands from individual studies.
- Paucity of generic methods appropriate for 3D pathology structure investigations contrasts with 2D analysis tools in relative abundance.

As 2D microscopy images can only present histopathological structures at discrete planes, we aim to provide efficient and generic 3D image analysis tools to facilitate biomedical investigations where 3D tissue examinations at microscopic resolution are important. Our vessel structure analysis framework is one such effort to enrich microscopy image analysis tools applicable to a large spectrum of 3D pathology phenotypic hallmarks in numerous diseases. To achieve a reasonable efficiency and accuracy, our current method is focused on primary vessel structure modeling and analysis. As a result, inaccurate segmentation results of small vessels do not affect following modules. Nevertheless, the efficacy of our method is affected by such factors as unduly large inter-slice resolution and tissue slicing orientation during tissue slide preparation. In our future work, we will improve individual analysis modules to reduce errors propagated in our working pipeline. We will also develop methods to accommodate small vessels. Additionally, we will validate our method with more liver tissue sample datasets and extract informative 3D imaging features to help researchers and clinicians better analyze 3D biological structures.

As a follow-up research, we will characterize vessels and other tube-shaped structures with the following generic features. 1) Length: the length of the 3D vessel can be determined by the medial axis with the hamonic skeletonization technique (Yang et al., 2005); 2) Thickness of walls: the vessel wall thickness can be used to measure blood flow volume and to indicate the degree of vascular integrity closely related to oncogenesis; 3) 3D vessel bifurcations: the number of bifurcations along the artery could be a clinically relevant parameter that determines the complexity of the vessel structures; 4) Angles at bifurcations: the angles at bifurcations of the vessels could affect volume of blood flow; 5) Local cross-sectional area: cross-sectional areas at different locations along a vessel can characterize vessel structures. With these fundamental features measured, we will further describe high-level 3D shape features. For vessels, we will calculate the inter-branch distance, characterize the branching pattern, and generate a histogram of branch lengths and thicknesses for the entire 3D tissue volume. With these 3D structure features and their cross-relationships quantitatively characterized, better clinically relevant patient stratification protocols and molecularly correlated phenotypic information could be expected. Our end goal is to combine our methods with such techniques as immunohistochemistry, Fluorescent In Situ Hybridization, or mRNA gene expression data and potentially provide complementary information to enhance investigations on the anatomical features and micro-architecture of histologic tissues, mechanisms of tumor progression, spatial distributions of signatred molecular biomarkers, expression of master regulators in cancer pathways in relation to 3D vessel structural features.

In summary, we present in this paper a new 3D vessel structure analysis method consisting of registering, segmenting, matching, interpolating, and visualizing blood vessels in liver tissues with large-scale microscopy images. We vision this cohort of 3D analysis tool development effort as a new avenue to facilitate pathology review and biomedical research, and a better way to engage researchers and clinicians in using rich molecular and clinical related information from histopathological imaging data.

Acknowledgement

This research is supported in part by grants from National Institute of Health K25CA181503, National Science Foundation ACI 1443054 and NSF IIS 1350885, and CNPq.

References

- Petushi, S., Garcia, F.U., Habe, M., Katsinis, C. and Tozeren, A. (2006) 'Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer', *BMC Med Imaging*, 6(14), doi: 10.1186/1471-2342-6-14 PMID: PMC1634843, page length: 11
- Xing, F.Y. and Yang, L. (2013) 'Robust cell segmentation for non-small cell lung cancer', *International Symposium on Biomedical Imaging*, Vol. 1, pp.386–389
- Jafari-Khouzani, K. and Soltanian-Zadeh, H. (2003) 'Multiwavelet grading of pathological images of prostate', *IEEE Trans Biomedical Eng.*, Vol. 50, pp.697–704
- Tabesh, A., Teverovskiy, M., Pang, H.Y., Kumar, V.P., Verbel, D., Kotsianti, A. and Saidi, O. (2007) 'Multifeature prostate cancer diagnosis and Gleason grading of histological images', *IEEE Trans Med Imaging*, 26(10), pp.1366–1378
- Han, J., Chang, H., Loss, L., Zhang, K., Baehner F.L., Gray, J.W., Spellman, P. and Parvin, B. (2011) 'Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme', *IEEE International Symposium on Biomedical Imaging*, pp.711–714.
- Cooper, L., Sertel, O., Kong, J., Lozanski, G. and Gurcan, M. (2009) 'Feature-Based Registration of Distinct Stained Histopathology Images: An Application for Computerized Follicular Lymphoma Prognosis', *Computer Methods and Programs in Biomedicine*, 96(3), pp.182–192
- Kong, J., Sertel, O., Shimada, H., Boyer, K.L., Saltz, J.H. and Gurcan, M. (2009) 'Computer-aided Evaluation of Neuroblastoma on Whole-slide Histology Images: Classifying Grade of Neuroblastic Differentiation', *Pattern Recognition*, 42(6), pp.1080-1092
- Kong, J., Cooper, L.D., Wang, F., Gutman, D.A., Gao, J., Chisolm, C., Sharma, A., Pan, T., Van Meir, E.G., Kurc, T., Moreno, C.S., Saltz, J.H. and Brat, D.J. (2011) 'Integrative, Multimodal Analysis of Glioblastoma Using TCGA Molecular Data, Pathology Images, and Clinical Outcomes', *IEEE Transactions on Biomedical Engineering*, 58(12), pp.3469–3474
- Jara-Lazaro, A.R., Thamboo, T.P., Teh, M. and Tan, P.H. (2010) 'Digital pathology: exploring its applications in diagnostic surgical pathology practice', *Pathology*, 42(6), pp.512–518
- Kong, J., Cooper, L.D., Wang, F.S., Gao, J., Teodoro, G., Scarpace, L., Mikkelsen, T., Moreno, C.S., Saltz J.H. and Brat, D.J. (2013) 'Generic, Computer-based Morphometric Human Disease Classification Using Large Pathology Images Uncovers Signature Molecular Correlates', *PLoS One*, 8(11):e81049. doi: 10.1371/journal.pone.0081049. eCollection 2013, page length: 17
- Coons, S.W., Johnson, P.C., Scheithauer, B.W., Yates, A.J. and Pearl, D.K. (1997) 'Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas', *Cancer*, 79(7), pp.1381–1393

- Qi, X., Xing, F.Y., Foran, D.J. and Yang, L. (2012) 'A fast, automatic segmentation algorithm for locating and delineating touching cell boundaries in Imaged histopathology', *Journal of Methods in Medical Informatics*, Vol. 51, No. 3, pp. 91–102
- Foran, D.J., Chen, W. and Yang, L. (2011) 'Automated image interpretation computer-assisted diagnosis', *Analytical Cellular Pathology*, Vol. 34, No. 6, pp.279-300
- Yang, L., Tuzel, O., Meer, P. and Foran, D.J. (2008) 'Automatic image analysis of histopathology specimens using concave vertex graph', *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, Vol. 11, pp. 833-841, 2008..
- Chang, H., Han, J., Spellman, P.T. and Parvin, B. (2012) 'Multireference level set for the characterization of nuclear morphology in glioblastoma multiforme', *IEEE Transactions on Biomedical Engineering*, 59(12), pp.3460–3467
- Teodoro, G., Pan, T., Kurc, T.M., Kong, J., Cooper, L.D., Podhorszki, N., Klasky, S. and Saltz, J.H. (2013) 'High-throughput Analysis of Large Microscopy Image Datasets on CPU-GPU Cluster Platforms', *IEEE International Symposium on Parallel and Distributed Processing*, pp.103–114
- Lesage, D., Angelini, E.D., Bloch, I. and Funka-Lea, G. (2009) 'A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes', *Medical Image Analysis*, 13(6), pp.819-845
- Friman, O., Hindennach, M., Kühnel, C. and Peitgen H. O. (2009) 'Multiple hypothesis template tracking of small 3D vessel structures', *Medical Image Analysis*, 14(2), pp.160-171
- Kubisch, C., Glaer, S., Neugebauer, M., and Preim, B. (2012) 'Vessel visualization with volume rendering', *Vis. in Med. & Life Sci.*, pp.109-134
- Hu, Z., Niemeijer, M., Abramoff, M. D., Lee, K. and Garvin, M. K. (2010) 'Automated segmentation of 3-D spectral OCT retinal blood vessels by neural canal opening false positive suppression', *MICCAI*, pp. 33–40
- Arribas, S. M., Daly, C. J., Gonzalez, M. C. and McGrath, J. C. (2007) 'Imaging the vascular wall using confocal microscopy', *J. Physiol.*, 584, pp. 5–9
- Jiang, H., Fels, S. and Little, J.J. (2007) 'A linear programming approach for multiple object tracking', *Proc. in IEEE CVPR*, pp.744-750
- Hillier, F. S. and Lieberman G. J. (2001) 'Introduction to Operations Research', *McGraw-Hill*, New York, NY, ISBN 0077298349
- Ismail, A., Gray, S., Jackson, P., Shires, M., Crellin, D. M., Magee, D., Quirke, P. and Treanor, D. (2010) '3D Histopathology of the Liver Using Dual Chromogen Histochemistry', *reAgents*, 3, pp. 20–22
- ecastro E. and Morandi C. (1987) 'Registration of Translated and Rotated Images Using Finite Fourier-Transforms', *IEEE T Pattern Anal*, 9, pp. 700–703

Roberts, N., Magee, D., Song, Y., Brabazon, K., Shires, M., Crellin, D., Orsi, N.M., Quirke, R., Quirke, P. and Treanor, D. (2012) 'Toward Routine Use of 3D Histopathology as a Research Tool', *Am J Pathol*, 180(5), pp.1835–1842

Vincent, L. (1993) 'Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms', *IEEE Transactions on Image Processing*, Vol. 2, No. 2, pp.176–201

Ruifrok, A.C. and Johnston D.A. (2001) 'Quantification of histochemical staining by color deconvolution', *Anal Quant Cytol Histol*, 23(4), pp.291–299

Zahn, C.T. and Roskies, R.Z. (1972) 'Fourier Descriptors for Plane Closed Curves', *IEEE Transactions on Computers*, Vol. 21, pp. 195–201

Hazewinkel, M. (2001) 'Linear interpolation', *Encyclopedia of Mathematics*, Springer, Netherlands, ISBN 9781556080104

Yang, Y., Zhu, L., Haker, S., Tannenbaum, A.R. and Giddens, D.P. (2005) 'Harmonic Skeleton Guided Evaluation of Stenoses in Human Coronary Arteries', *Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 1), pp.490–497