AUTOMATED NEUROVASCULAR TRACING AND ANALYSIS OF THE KNIFE-EDGE SCANNING MICROSCOPE RAT NISSL DATA SET USING A COMPUTING CLUSTER

A Thesis

by

SUNGJUN LIM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Yoonsuck Choe
Committee Members,	Anxiao Jiang
	Li Peng
Head of Department,	Dilma Da Silva

December 2015

Major Subject: Computer Science

ABSTRACT

3D reconstruction of the neurovascular networks in the brain is a first step toward the analysis of their function. However, existing three dimensional imaging techniques have not been able to image tissues on a large scale at a high resolution in all three dimensions. For creating high-resolution neurovascular models, the Knife-Edge Scanning Microscope (KESM) at Texas A&M University has been developed and used to image whole rat brain vascular networks at submicrometer resolution.

In this thesis, I describe algorithms that are fully automatic and compatible with the large KESM rat Nissl data set. The method consists of image enhancement, binarization, 3D neurovascular networks tracing, and quantizing anatomical statistics. These methods are easily parallelizable and are compatible with high-throughput microscopy data. A computing cluster has been used to increase the throughput of the methods. Using the method developed, I analyzed a large volume of rat brain vasculature data. The results are expected to shed light on the structural organization of the vascular network that underlies the delivery of oxygen, nutrients, and signalling molecules throughout the brain.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I truly appreciate my advisor, Dr. Yoonsuck Choe for his constant support, guidance, and encouragement throughout the course of this research. I would also like to thank my committee members, Dr. Peng Li and Dr. Anxio Jiang for their guidance and valuable comments on my research.

I would like to thank Jaewook Yoo, Michael Nowak, Junseok Lee, Wookyung Ahn, Wenjie Yang, Ankur Singhal, and all the other lab mates of mine. Finally, my deepest gratitude goes to my family for their caring and great support in me all through these years.

This research was partially funded by the National Science Foundation, grants #1208174 and #0905041 (PI: Yoonsuck Choe).

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

FIGURE Page

Page

Page

Page

Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Motivation

Neurovascular networks play an important role in analyzing neurological conditions [34, 39, 23, 28, 31]. However, the structure of the microvascular networks is not known completely, because the vessels are very small in diameter, and the structure is very complex. Knife-Edge Scanning Microscope(KESM) has enabled staining and imaging a large volume of biological tissues at sub-micro meter resolution [17]. Compared to other brain vasculature atlases [11, 15, 24] KESM data has better z-axis resolutions. KESM has been used to image tissue stained with Nissl as well as other stains [18]. Nissl stains the cell bodies of neurons and other cells while leaving the dendrites and axons unstained. Also, blood vessels are not stained, thus neurovascular relations can be investigated.

The microvasculature is known as to deliver nutrients to the nearby cells and exchanges molecules with its surrounding tissues. In this sense, brain microvasculature plays an important role in chronic disease such as Alzheimer's Disease, Multiple Sclerosis, and Parkinson's Disease. The key interest in the microvasculature research is how much access the tissues have toward the blood-bourne molecules. However, details about the microvasculature are not known due to its small size and exceptional complexity. Therefore, the traced vascular structure and estimated anatomical statistics from this thesis will be useful for the research on mechanism between the microvasculature and the cells.

In this thesis, image enhancement and segmentation methods will be developed for the extraction of vessel cross sectional regions of the Nissl-stained data. The image enhancement algorithm minimizes both random noise and lighting artifacts present in the Nissl-stained data. For tracing the neurovascular network in the Nissl data, a simple and robust tracing algorithm will be developed. Since the tracing algorithm is simple, it can reduce computation time while tracing large volumes of neurovascular network.

1.2 Goal of the research

The goal of this thesis is to extract the vascular cross sectional regions from the sequence of the Rat Nissl-stained images and to analyze its vascular structure. The thesis focuses on finding anatomical statistics such as the length and diameter of the traced vascular structure.

1.3 Approach

For tracing the vascular structure automatically in the image sequence, the following steps are required: (1) image enhancement, (2) binarization, (3) centroid selection, and (4) neurovascular network tracing. In the Nissl-stained data, random noise and lighting artifacts exist [18].

To minimize the random noise and lighting artifacts, images need to be enhanced. In image enhancement, every image in the original image sequence is paired with their subsequent images. The two subsequent images in the pair are averaged together so that the averaged images make up a new image sequence. The result of the averaging process is useful to minimize the lighting artifacts on the background of the image. In other words, the background pixel intensities stays low while the foreground (the vessel cross sectional region) pixel intensities stay high. Finally, the result image does not have the lighting artifacts which could become noise. In the end, compared to the original image sequence, the number of the new image sequence is half that of the original image sequence. After the images are enhanced, every enhanced image is subtracted by their duplicate image convolved with a Gaussian filter. The image subtraction step removes unnecessary background except for the vessel cross sectional regions.

After the background is subtracted from the image, Max-entropy algorithm is applied to the subtracted image and results in a binarized image where a vessel cross sectional region has 255 pixel intensity and the background 0 pixel intensity. Max-entropy algorithm enables dynamic threshold selection for the subtracted images [14, 10]. In the end, the selected threshold values result in binarized images of the vessel cross sectional regions,

and the binarized images will be used for tracing the vascular structure.

Although the binarization process is successful, the result images could contain damaged vessel cross sectional regions. Due to their damaged vessel cross sectional regions, they result in gaps that prevent tracing the vascular structure. However, morphological operations (2D closing and filling holes) resolve the gap problem.

Although the result image is completely binarized, the result image still contains noise. To get rid of the noise without any damage on vessel cross sectional regions, a guidance map is introduced. The guidance map is generated using the following formula: $\min(I_x I_y)$ where I_x and I_y are two vertically adjacent images. In other words, the minimum pixel intensity values of two vertically adjacent images make up the guidance map. For example, if vessel cross sectional regions are vertically adjacent, their adjacent region shows up in the guidance map. However, noise does not show up in the guidance map since it is unlikely to be adjacent through the z-axis. In this sense, every connected component that is extracted based on the guidance map is most likely vessel cross sectional region in the result image. Finally, noise is removed in the result image.

After preprocessing binarized images, their vascular structure is traced. For tracing the vascular structure, a main problem is to find the connectivity among the vessel cross sectional regions. The following assumption is applied when the connectivity among the vessel cross sectional regions is defined: two vessel cross sectional regions are connected if they are adjacent through the z-axis. The guidance map which was previously generated to remove the noise is reused to define the connectivity among the vessel cross sectional regions. Similar to the noise removal step, connected components from the guidance map are used to label two connected vessel cross sectional regions on two vertically adjacent images. After the connected vessel cross sectional regions are labeled, their centroids are extracted. The way to calculate the centroid of vessel cross sectional regions is as follows: $\frac{\sum_{i=1}^{n} x_i}{n}$ where $x_1 ... \ x_n$ are pixel locations of the extracted vessel cross sectional regions [3].

Finally, the anatomical statistics are analyzed. First, the vasculature length is estimated using the following formula: $\sum_{i=1}^{n-1} \|C_{i+1} - C_i\|$ where C_i is the centroid of the connected

vessel cross sectional region, and n is the last vessel cross sectional region of the vasculature. Additionally, the vasculature diameter calculation is done by a 6-neighborhood exhaustive search around the centroids on the vessel cross sectional regions.

To increase the accuracy of the diameter estimation, both mean and standard deviation of 6 candidate radii are calculated. The mean of the radii is added to the standard deviation that is divided by a constant value. For validation of the estimated diameter, the adjusted mean values of the radii are utilized to estimate the volume of the vascular structure. Based on the equation $S_i = \pi\ r^2$ where r is the estimated radius, the volume $\sum_{i=1}^{m}\sum_{j=1}^{n} S_j$ is estimated, where n is the number of vessel cross sectional regions in an image, and m is the depth of the volume. The estimated volume is validated by the ratio of the original volume to the estimated volume: $\frac{E}{S}$ where S is the estimated volume, and E the original volume.

The number of vessel cross sectional regions where their bifurcation begins is also estimated. The bifurcation point is a vessel cross sectional regions where a blood could flow in more than one direction. In other words, if the vessel cross sectional regions in a image are adjancent with more than single vessel cross sectional region in a subsequent image, the vessel cross sectional region is considered as the branch point.

1.4 Significance

This thesis introduces efficient, automatic methods for image enhancement, binarization, vascular tracing, and anatomical statistic analysis on the Nissl-stained data. As a result, the methods accurately extract vessel cross sectional regions of the Nissl-stained data and trace the vascular structure without human assistance. The tracing method is easily parallelizable and fully exploits a computing cluster. In the end, its throughput and performance significantly increases, so a large volume of rat brain vasculature data is processed, offering the interesting analysis on its anatomical statistic.

1.5 Outline of the thesis

The thesis consists of several sections as follows. In section 2, background of KESM, its datasets, and the Nissl-stained data; and Related works similar to this thesis are also introduced. In section 3, the methods used for extracting vessel cross sectional regions, tracing vascular structures, and analyzing the anatomical statistics are presented. In section 4, experimental results and analysis are discussed. In section 5, contribution, open issues, and future work of this thesis are discussed.

2. BACKGROUND AND RELATED WORK

2.1 Knife-Edge Scanning Microscope

Knife-Edge Scanning Microscope (KESM) was developed in the Brain Network Laboratory (BNL) at Texas A&M University (TAMU). KESM is able to acquire a high-resolution whole-brain scale data [17]. KESM has acquired three different mouse brain data stained in Golgi, India ink, and Nissl [6]. Nissl stains all cell bodies of neurons leaving vasculatures unstained in the brain. In this sense, the Nissl-stained data is useful to find the relationship between cellular information and the vascular structure [18].

KESM consists of the following main components: (1) high-speed line-scan camera, (2) microscope objective, (3) diamond knife assembly and light collimator, (4) specimen tank (for water immersion imaging), (5) three-axis precision air-bearing stage, (6) white-light microscope illuminator, (7) water pump (in the back) for the removal of sectioned tissue, (8) PC server for stage control and image acquisition, (9) granite base, and (10) granite bridge [17, 22]. Figure 2.4 shows the descriptions of the KESM parts

The air-bearing stage is operated to offer accurate movement in x,y axies (with 20 nm precision) and z axies (with 25 nm precision) [21]; it provides robust registration of KESM images. Also, the KESM provides the custom knife-collimator, which is classified into two kinds: a white-light collimator and a diamond microtome. KESM cuts thin sections of the specimen using the diamond knife while the diamond knife provides illumination. This whole process generates a thin section of tissue over the diamond knife. Line-scan imaging is done near the tip of the diamond knife as the tissue gets cut, resulting in high resolution images of the cut tissue [6]. Figure 2.2 shows the zoomed-in version of the KESM.

The KESM simultaneously sections and images tissue blocks. A brain tissue is embedded in hard polymer resin block; It is fixed on the specimen tank which is submerged under water. Instead of the diamond knife movement, the stage moves the specimen tank against the knife while generating a thin section of tissue over the tip of the diamond knife where

the distortion may be minimized. At the same time, the white-light source illuminates the tissue piece on the tip of the diamond knife. Once the microscope objective and the diamond knife aligned, the camera images the tissue piece through microscope objective and sends the image signal to the computer server. Figure 2.3 illustrates how KESM imaging operates.

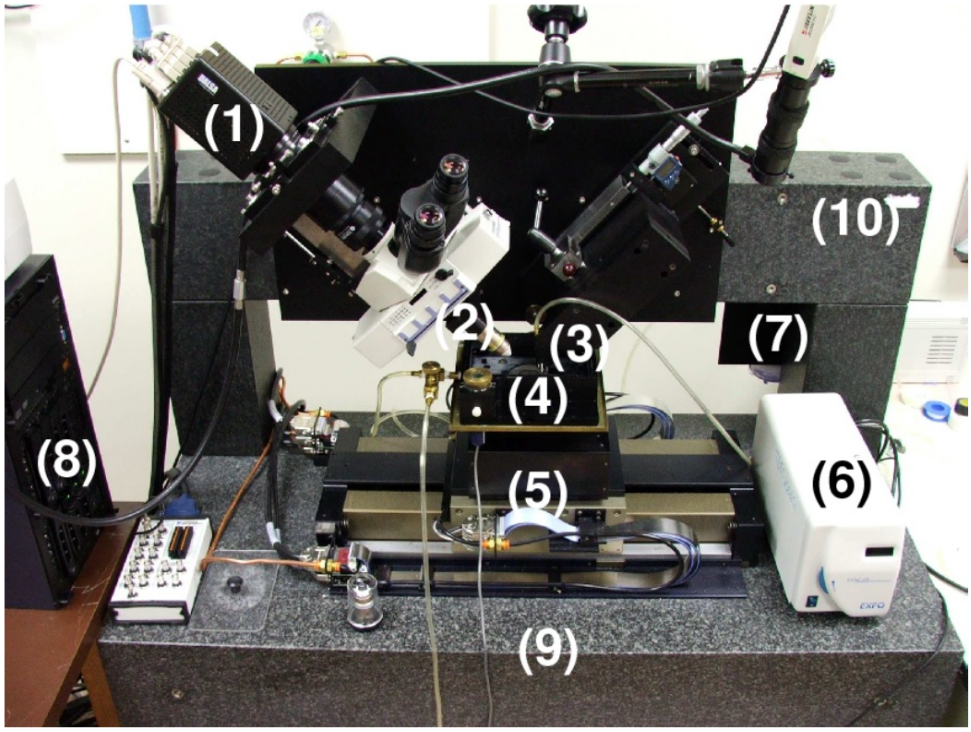


Figure 2.1: The KESM parts consist of as follows: (1) high-speed line-scan camera, (2) microscope objective, (3) diamond knife assembly and light collimator, (4) specimen tank (for water immersion imaging), (5) three-axis precision air-bearing stage, (6) white-light microscope illuminator, (7) water pump (in the back) for the removal of sectioned tissue, (8) PC server for stage control and image acquisition, (9) granite base, and (10) granite bridge. Reprinted from [6].

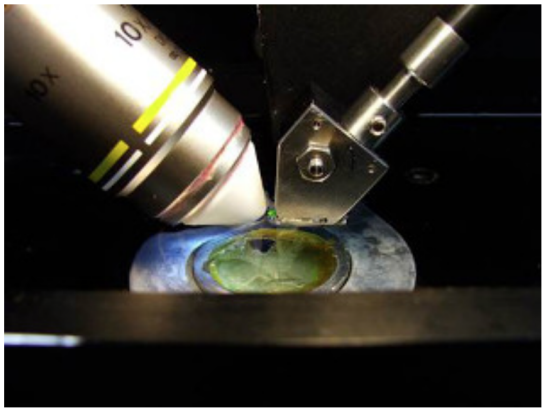


Figure 2.2: A zoomed-in version of the KESM. Reprinted from [22].

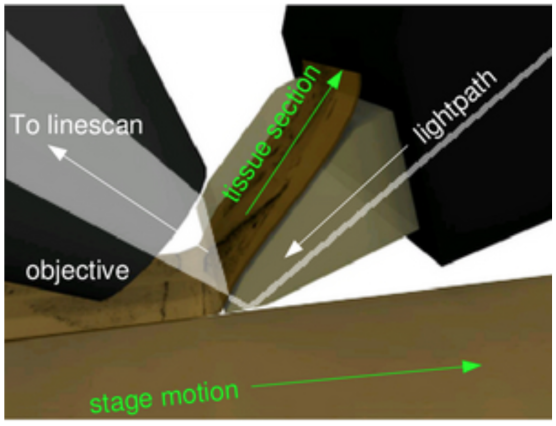


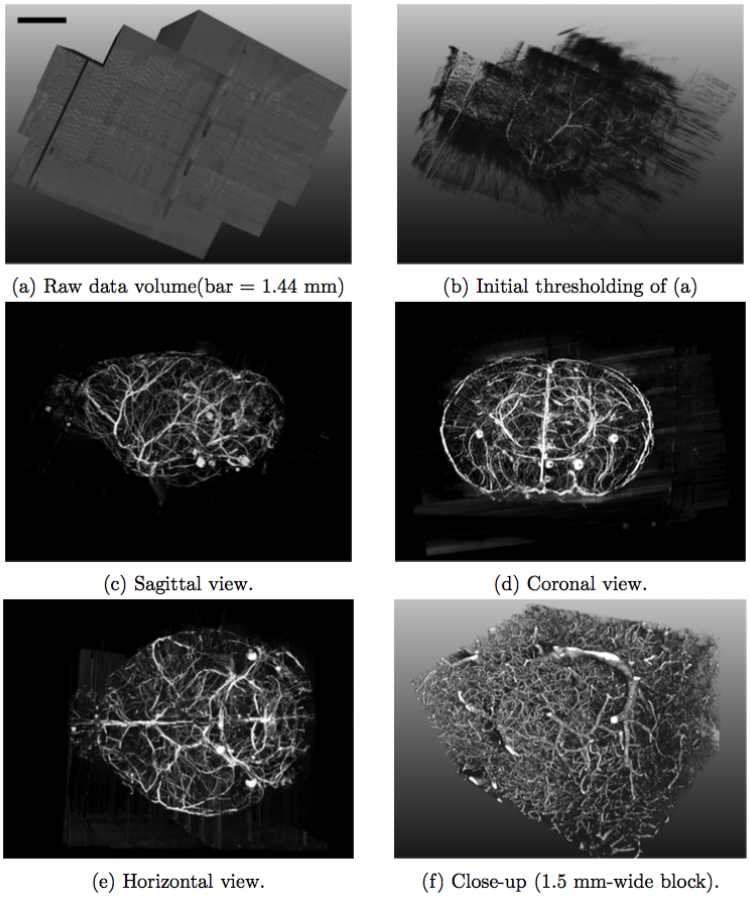
Figure 2.3: Imaging principle. Reprinted from [6].

2.2 KESM data

KESM has successfully scanned whole mouse brains stained in Golgi, Nissl, and India ink. In 2008, the earliest data were obtained from Golgi-stained (neuronal morphology) and India ink-stained (vasculature) mouse brain. In early 2010, a whole mouse brain stained with Nissl was scanned to observe its cell body distribution [6].

In this thesis, a dataset obtained from rat brain stained with Nissl. In this dataset, the cell bodies of neurons and other cells are stained while the dendrites and axons not stained. Since blood vessel are not stained, neurovascular relations can be observed by the proposed methods in this thesis.

The KESM datasets are basically two dimmensional images that can make up three dimmensional image stack. Figure 2.4 presents the KESM dataset in the three dimmensional stack form with its different views (sagital, coronal, and horizontal). Figure 2.6 presents the KESM rat Nissl dataset in the three dimmensional stack and in a mosaic form. In this thesis, to visualize the three dimmensional stack of the Nissl data, a visualization software ParaView is used.



(a) Raw data volume(bar = 1.44 mm)

(b) Initial thresholding of (a)

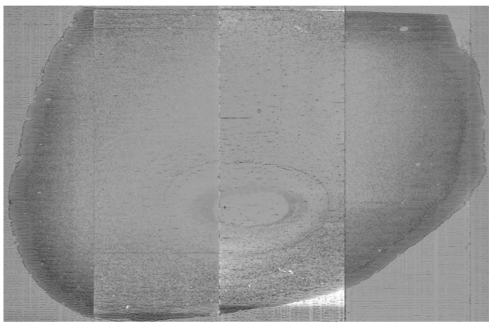
(c) Sagittal view.

(d) Coronal view.

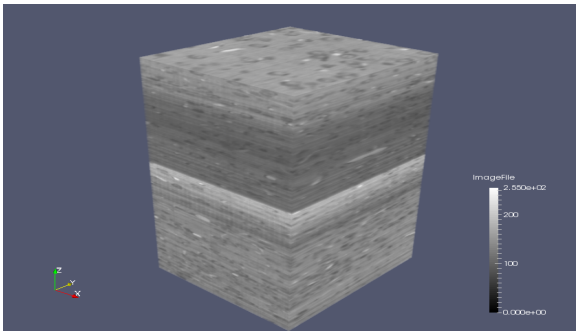
(e) Horizontal view.

(f) Close-up (1.5 mm-wide block).

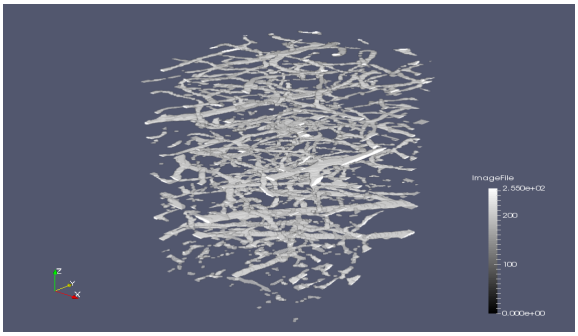
Figure 2.4: Visualization of vascular structures from the KESM mouse India ink data.(a) A raw data block in a sagittal view. (b) is resulted from (a) which is applied by the initial thresholding.(c)-(e) are fully thresholded in different angles (Sagittal, Coronal, Horizontal). (f) Zoomed in version of the complex details within a 1.5 mm wide block. Reprinted from [6].



(a) Image A



(b) Image B



(c) Image C

Figure 2.5: Visualization of the KESM rat Nissl data. (a) ImageA is a mosaic from imaged KESM rat Nissl data. Each column is independent data set. The columns are then merged into a single dataset. The illustration is reprinted from [19]. (b) Image B represents a image stack of the KESM rat Nissl data. (c) Image C represents a image stack of the binarized KESM rat Nissl data. The binarized image stack presents only vascular information.

2.3 Related work

With the Nissl-stained data, the cellular and the vascular information are available. However, the vascular information is not automatically available from the raw data. Different techniques have been used to extract vascular information. In this section, techniques for extracting microvasculature information are covered.

2.3.1 Automatic seed points extraction

Can et al. [5] used the vector technique to extract vessel information. The vector technique exploited 2D correlational kernel which is perpendicular to the direction of the vessel with low-pass averaging filters along the vessel. Vessel boundaries at various angles were detected by the peak response from the 2D correlational kernel. The template that consists of $[-1, -2, 0, 2, 1]^\intercal$ kernels found a next point and its orientation in a recursive manner. Al-kofahi et al. extended the technique in three dimensions for confocal images [2]. In order to select an initial seed point and an initial direction for the tracing, Mayerich et al. [40] made use of Doddapaneni's method [7] that found seedpoints by applying polymerization algorithm on a given data set and generating the endpoints of the clusters which would be served as seed points.

Xiao et al. [36] used ridge enhancement in order to emphasize vascular structure. Then, seed points were selected within emphasized vascular structure in terms of its local maxima intensity level when its hessian matrix stayed negative, and its gradient remained as zero. To remove remaining noise of low pixel intensity, any object consisting of low pixel intensity was removed by the thresholding technique. Y. Yin et al. [37] uses a probabilistic model (Gaussian model) to profile vessel's sectional intensity profile. The model can classify bifurcation, crossing configuration, and it improves on its results. In order to estimate local vessel edges, MAP(Maximum a posteriori) was utilized.

Shan et al. [35] developed a technique that selects seed points on breast ultrasound image segmentation. Preservation of edges and noise reduction were done by speckle reduction method. A histogram and breast lesion's spatial property were utilized to select a

threshold value iteratively. The selected threshold values filtered regions except for possible lesion areas. The scores of remaining regions were determined by considerations of the area, variance, and distance to the center. A region with the highest scores was considered as a final lesion. The center of lesions was considered as the seed point.

Zhang et al. [38] used automatically detected seed points in order to extract centerline in 3D Axon Images. Users selected initial seed points (centroids) for each axon object from a first slice of an image volume. Dynamic Programming (DP) searched other seed points in following slices of an image volume. However, estimated seed points by DP could be incorrect when multiple axon objects overlapped.

As described above, many techniques have been developed to detect seed points automatically; However, most of techniques above will not be compatible with KESM rat Nissl data since the techniques were developed for different kinds of data. Therefore, in this thesis, simple and fast segmentation technique will be introduced for automatic seed point detection. After vasculature cross section information is extracted from the images, centroids of vasculature cross section(seed points) are detected.

2.3.2 Vector tracing algorithms

Mayerich et al. [19, 20] made use of 3D vector tracing algorithm to extract 3D radius information and thread shape objects. Mayerich et al. initialized the vector tracking algorithm by finding initial seed points from thresholding technique. Templates are exploited according to seed point locations, and those templates are rotated and scaled in order to minimize a heuristic function which prevents excessive sampling. The templates are updated based on a selected orientation vector which is contributed to minimize the heuristic function. As the tracing proceeds, the current seed point is evaluated whether it has an intersection with a previous traced centerline. Figure 2.6 shows an overview of the vector tracing algorithm.

Palagyi et al. [27] introduced a thinning algorithm that extracted centerline. In the sequential thinning algorithm, one iteration step consisted of six successive sub-iterations based on the six main directions in 3D. Each sub-iteration contains two phases: (1) if the

border points of the actual type are both simple and not line-ended, those border points are considered as potential deletable points. (2) During the rechecking, if a marked point is simple and not a line-end point after previously visited marked points are deleted, the marked point is deleted.

Han et al. [13] introduced a fast and robust fiber tracing method for 2D and 3D data. For 2D tracing, Users selected seed points; these seed points were adjusted by two fiber boundary lines. Based on the seed points, a Moving Window (MW) was located, which determined the tracing direction. For handling of branching points, a Cubic Tangential Trace Spline (CTTS) was utilized as a nonlinear interpolation approach. Figure 2.8 illustrates how 2D MW is generated. For 3D tracing, local maximum intensity projection (MIP) was exploited for a method. MIP utilized 2D tracing for 3D tracing. In order to get an accurate projection, MIP cube size was set to be the medium length of the local X,Y,and Z axis. Figure 2.7 illustrates how MIP is exploited for 3D tracing.

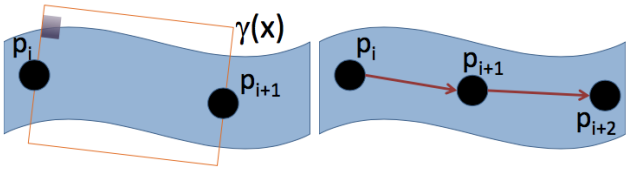


Figure 2.6: (left) Tracking the filament axis using a template function. (right) Successive steps along the filament axis are done after computing the template orientation. Reprinted from [19].

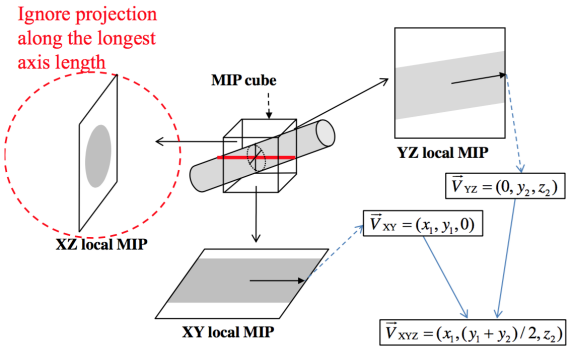


Figure 2.7: An illustration of exploiting MIP for 3D tracing. The longest axis is igonored because the local MIP with the logest axis does not result in useful information. The other two local MIPs will be used to determine the tracing direction. The obtained 2D tracing directions are used to determine 3D tracing direction, finally. Reprinted from [13].

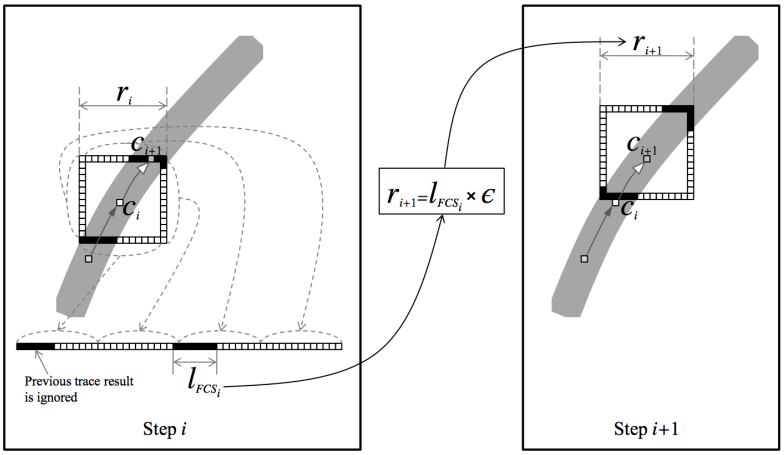


Figure 2.8: An illustration of the calculation of 2D MW. r_i is the side length of MW during the i^{th} step. c_i is the center of MW during the i^{th} step. l_{FCS_i} is the length of the fiber cross section during i^{th} step. ϵ scales the length of the fiber cross section to get the MW side length during the i^{th} step. c_{i+1} is the center of MW during the i^{th} step. Reprinted from [13].

3. METHODS

In this section, the methodology of this thesis is discussed. The methodology consists of the following steps: image enhancement, binarization, centroid selection, automatic vascular structure tracing, Paraview visualization, and data analysis. The following sections describe each step of the methodology in detail.

3.1 Image enhancement

The raw images of the KESM data contain both random noise and lighting artifacts due to both lighting conditions and mechanical vibrations [10]. In other words, the lighting artifacts in the Nissl data cause uneven illumination problem which will result in noise in binarized result images. In order for image binarization to be successful with less noise, the raw images need to be enhanced. Usually, the area where lighting artifacts reside has high pixel intensity; that is, when the lighting artifacts reside on the image background, the background with the lighting artifacts may be classified as one of vessel cross sectional regions.

To minimize the lighting artifacts, an averaging operation is done as follows: Let X_i be the 2D raw image that is *ith* image out of the image sequence. Let Y_z be the enhanced image where z is $\frac{i+1}{2}$; it becomes the *zth* image out of a new averaged image sequence. In the end, the image X_i is averaged with the subsequnet image X_{i+1} in the following: $Y_z = \frac{X_i + X_{i+1}}{2}$. As a result of this, the area with the lighting artifacts (consisting of high pixel intensity) is averaged with the same area with darker background (consisting of low pixel intensity). In this sense, in the result image, the area with the lighting artifacts will consist of lower pixel intensity so that it less likely be classfied as the vessel cross sectional region during the binarization step. Figure 3.1 illustrates how the averaing operation on the image sequence is operated.

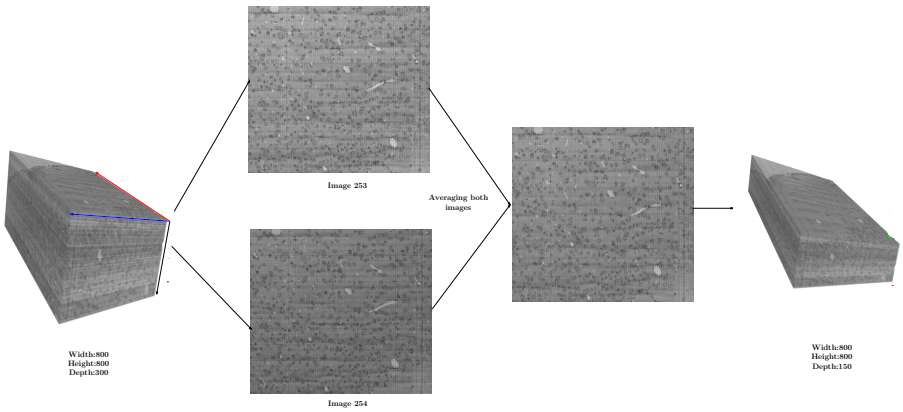


Figure 3.1: The figure shows the averaging operation between two adjacent images located at depths 253(image 253) and 254(image 254) of the image volume (Width:800 (blue colored axis), Height:800 (red colored axis), Depth:300 (black colored axis)). The new image volume has half the depth of the original volume.

3.2 Binarization

Once the image sequence is enhanced, the image sequence is ready to be binarized. To reduce noise in binarized result images, image subtraction technique gets rid of any other region than vessel cross sectional regions. After the image subtraction step, the image sequence is then binarized by Max-entropy method [14, 10] that results in a dynamic threshold value. During the binarization step, vessel cross sectional regions are made of pixel intensity 255, and background pixel intensity 0. However, due to low pixel intensity residing within vessel cross sectional regions, part of vessel cross sectional regions is not binarized correctly. As a result, damaged vessel cross sectional regions (with holes) are present in the binarized result images; Morphological 2D closing and hole filling techniques resolve the damage on vessel cross sectional regions. Finally, if every vessel cross sectional regions is not adjacent with any other vessel cross sectional region in its next subsequent image, it is considered as noise and discarded.

3.2.1 Background subtraction

When the raw image sequences is binarized after the image enhancement, many noise in result images are observed. In other words, background made of high pixel intensity is binarized as the vessel cross sectional regions. To prevent incorrect binarization, the image background needs to be removed as much as possible. As a simple and inexpensive approach, background subtraction technique is applied.

The background subtraction techinique is famous for detecting moving objects among many video frames; it is the result from the difference between an original image and its background image [29]. In other words, modeling the image background is the first step to be done. To model the image background, Gaussian smoothing technique [4] is applied to the duplicate image of the input image. During this step, the Gaussian smoothing techinique blurs the vessel cross sectional regions (high pixel intensity) by adapting the nearby background (low pixel intensity); Consequently, in the result image, background is only remained. Once the modeled background image is subtracted from the original image, vessel cross sectional regions remains more brighter (highter pixel intensity) while the background darker (lower pixel intensity). Figure 3.2 illustrates the result image that the background subtraction step generates.

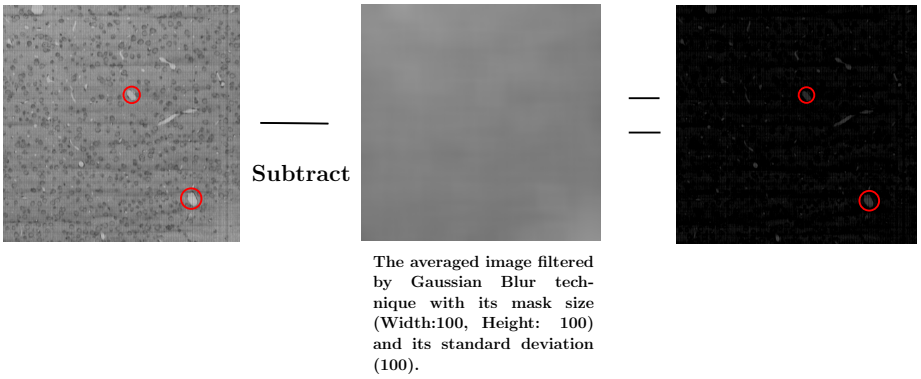


Figure 3.2: The averaged image is subtracted by its duplicate image filtered by Gaussian blur. White objects surrounded by red circles are vessel cross sectional regions.

3.2.2 The bilevel thresholding

Gray level thresholding is popular for image segmentation. In this thesis, bilevel thresholding technique is applied to partition the raw image into two regions: white (vessel cross sectional regions) and black (background) [26]. To find a global threshold of each image dynamically, Max-entropy thresholding method is applied to the image sequence [14]. For every images from the image sequence, Max-entropy method automatically searches a threshold value s that maximizes $\psi(s) = \ln P_s(1 - P_s) + \frac{H_s}{P_s} + \frac{H_n - H_s}{1 - P_s}$ where P_s is a distribution of gray level, and both H_s and H_n are posteriori entropies. While every image is scanned in the image seqeunce, different threshold values for the images are generated. By applying the global thresholding to the image sequence, the binarized image sequence is generated. In the end, the binarized result image consists of the vessel cross sectional regions (pixel intensity 255) and the background (pixel intensity 0).

3.3 2D closing and hole fillings

Although the bilevel thresholding step successfully results in the binarized result images, the binarized result images would consist of damaged vessel cross sectional regions. For example, in the raw image, part of the vessel cross sectional region may consist of low pixel intensity. In this sense, during the binarization step, the part of the vessel cross

sectional regions may be binarized as the background (pixel intensity 0); They may seem to contain holes within themselves. After obtaining the binarized image sequence, the binarized vessel cross sectional regions may consist of the vessel cross sectional regions.

The damaged vessel cross sectional regions result in gaps in 3D vasculature. The gaps prevent the vascular tracing correctly. In order to resolve this issue, morphological 2D closing and filling holes are used. The morphological 2D closing is known to fills holes, thin gulfs, or gaps in the contour. The morphological 2D closing consists of dilation and erosion step. When dilation is operated on binarized images, objects in the images will grow and get thicken. After the dilation operation, erosion is operated on the images, and thus the objects get shrinked or thinned. Both dilation and erosion that make up the morphological operation have the structuring element that controls the degree of effect both operations have to the binarized images. Since the excessive usage of either operations may lead to malformation of the vessel cross sectional regions, the size of structuring elements needs to be adjusted to result in appropriate dilation and erosion effect. In the end, by closing operation, the vessel cross sectional regions are enlarged while their ragged edges are pruned.

After the closing operation reconnects a disconnected boundary of the hole, the hole filling operation is fully operable. The hole filling operation fills the holes (background) inside region of interests (vessel cross sectional regions), consisting of dilation, complementation, and intersection. The complementation has its role to limit its operation to inside the vessel cross sectional regions [12]. By filling the holes, the gaps in 3D vasculature are resolved. Figure 3.3 illustrates how both the closing and the hole filling operations resolve damaged vessel cross sectional regions in 2D raw images. Figure 3.4 illustrates how both the closing and the hole filling operations resolve damaged 3D vasculature.

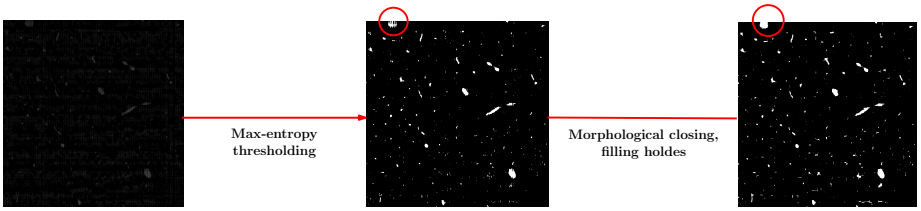


Figure 3.3: The figure shows the resulting images both after Max-entropy thresholding and Morphological closing,filling holes. The red circle emphasizes the damaged vessel cross sectional region and the resolved vessel cross sectional region by morphological operations.

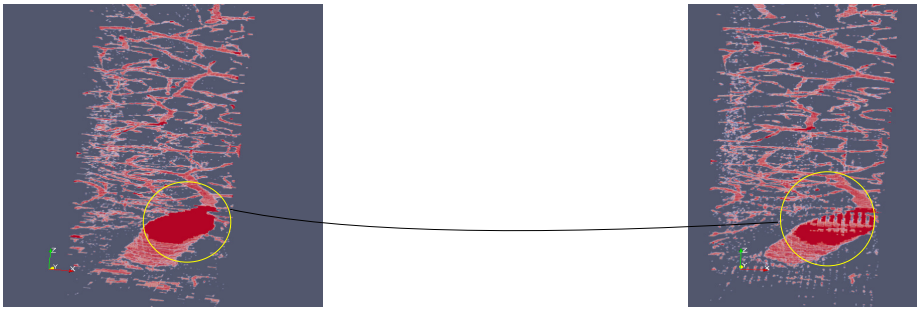


Figure 3.4: The figure describes how 2D closing and hole fillings resolve the gap problem in 3D vasculature visualization. The area in the yellow (left) is the vasculature with the gap filled by the hole fillings operation. They are in the yellow (right) is the vasculature with the gap problem.

3.4 Noise removal by the vessel connectivity property

After the morphological operations, the images may contain the noise that is binarized as false positive (background). Because some images may contain many noises, an expensive method is not promising to remove the noise. However, using vertical adjacency information of the vessel cross sectional regions, the noise can be removed simply. For example, it has been empirically observed that the vessel cross sectional regions in a binarized image have their matching vessel cross sectional regions in a next image. In other words, this observation explains that stack of those vessel cross sectional regions make up 3D vasculature.

In contrast, the noise shows the opposite observation to the vessel cross sectional regions; The noise less likely has its matching region in subsequent binarized images. Therefore, any suspicious vessel cross sectional regions that do not show the correct observation is discarded in the binarized images.

To gain the information of matching vessel cross sectional regions, a guidance map is introduced. The guidance map is generated using the following formula: $\min(I_x I_y)$ where I_x and I_y are two vertically adjacent images. In the guidance map, every regions with pixel intensity 255 represents the vessel cross sectional regions. Using Connected Component labeling technique [25], every connected component regions in the guidance map is recognized. Figure 3.6 illustrates how the connected labeling algorithm works on the guidance map. For each connected component recognized, its single pixel coordinate is only considered; By connected component labeling on the considered pixel coordinates, only vessel cross sectional regions are remained in the image that contributes to the guidance map generation. In the end, except for noise, only true vessel cross sectional regions remain in the image. Figure 3.5 illustrates how the noise removal is operated.

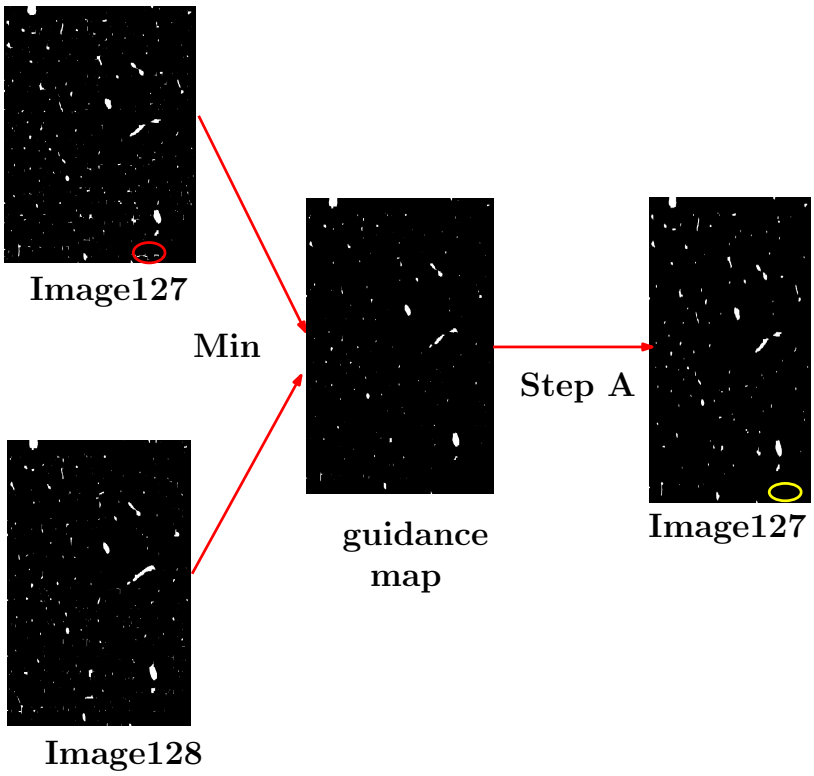


Figure 3.5: *min* method of Image127 (the image from depth 127) and Image128 (the image from depth 128) results in a guidance map. Step A is as follows: Based on connected components in the guidance map, every vessel cross sectional region is retrieved in Image127 removing unnecessary noise. Noise in a red circle is removed as shown in a yellow circle.

3.5 Tracing vascular structure and its visualization

To trace vascular structure from the clean binarized image sequence, the connectivity of vessel cross sectional regions through the image sequence needs to be defined first. The connectivity of vessel cross sectional regions is defined as follows: two vessel cross sectional regions are connected if they are adjacent through the z-axis. The guidance map introduced earlier is fully exploited to define the connectivity.

Similar to the noise removal step, the connected components in the guidance map are scanned by the connected component labeling technique. For each recognized connected components, the single pixel coordinate is recognized and used in two susbseqent binarized

images that are contributed to generate the guidance map; In those binarzed images, any connected component around the recognized pixel coordiante is labeled. Figure 3.6 shows the guidance map and the connected component labeling technique on it.

Besides generating the pair of connected vessel cross sectional regions, centroids of the connected vessel cross sectional regions are computed for tracing the vascular structure. The centroids of the vessel cross sectional regions are computed as follows: $\frac{\sum_{i=1}^{n} x_i}{n}$ where $x_1 ... x_n$ are pixel locations of the extracted vessel cross sectional [3]. Figure 3.12 illustrates how the centroids would be marked on every vessel cross sectional regions. Finally, the centroids of the connected vessel cross sectional regions are paired together and collected to be visualized. Figure 3.7 shows how the guidance map pairs two connected vessel cross sectional regions.

To visualize the image sequences and the traced vascular structure, ParaView [1] is used. ParaView is a software that allows scientistis to visualize large 3D datasets. Prior to the visualization in ParaView, the image sequence needs to be converted to RAW file format; for the conversion step, Fiji software [32] reads in the image sequence, stacks it, and convert it to a RAW file. Once ParaView imports a RAW file, it can visualize the stack of the image sequence. Figure 3.8 shows the visualization of the RAW file of the image sequence.

While two centroids are paired to show their connectivity, the connected centroids are recorded in Visualization Toolkit (VTK) [33] file for visualizatoin in ParaView. In terms of the VTK file syntax, the VTK file contains the coordinates of the centroids and their connectivity information so that they can be exploited for its visualization. Finally, Paraview visualizes the traced vascular structure using the centroids information described in the VTK file. It can overlay the stack of the binarized image sequence with the traced vascular structure. Figure 3.9 shows the traced vascular structure overlaid with the stack of the binarized image sequence.

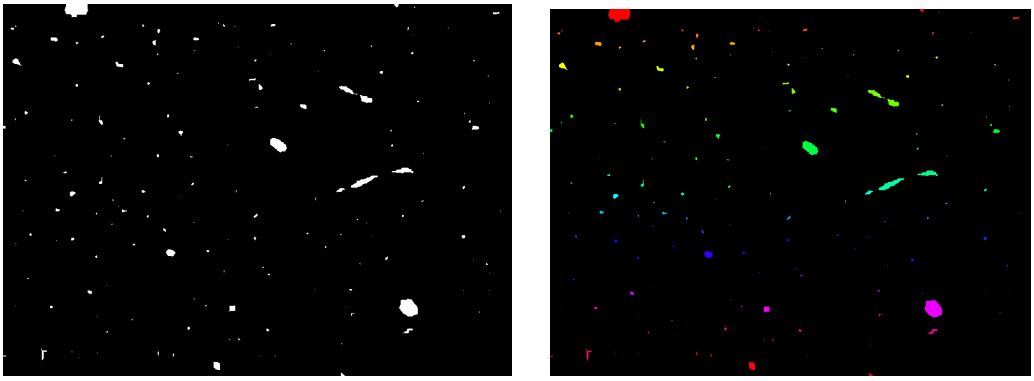


Figure 3.6: The image(left) is the guidance image, and the other image(right) shows the objects colored by the connected labeling algorithm.

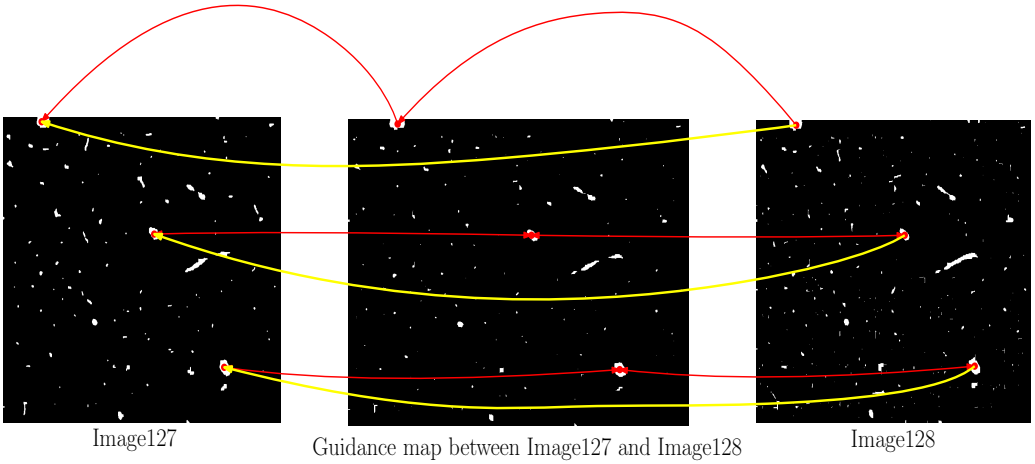


Figure 3.7: Two different images (from depth 127 and 128) are operated by *min* method to result in the guidance map. Red curves show how the guidance map connects the vessel cross sectional regions from the two different vertically adjacent images. Yellow curves define final connectivity between two vessel cross sectional regions.

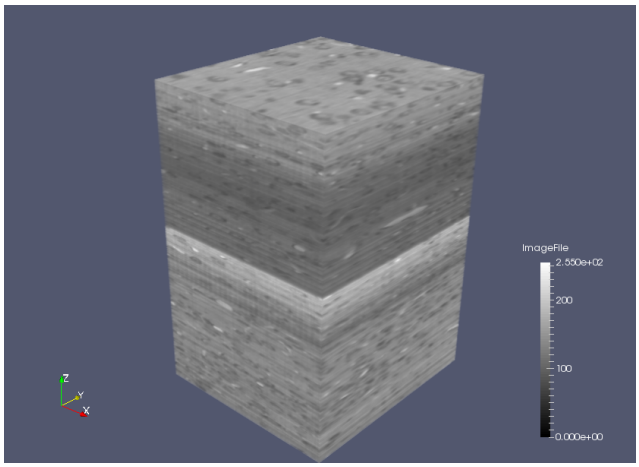


Figure 3.8: The figure represents an exemplary raw file visualization in 3D via Paraview.

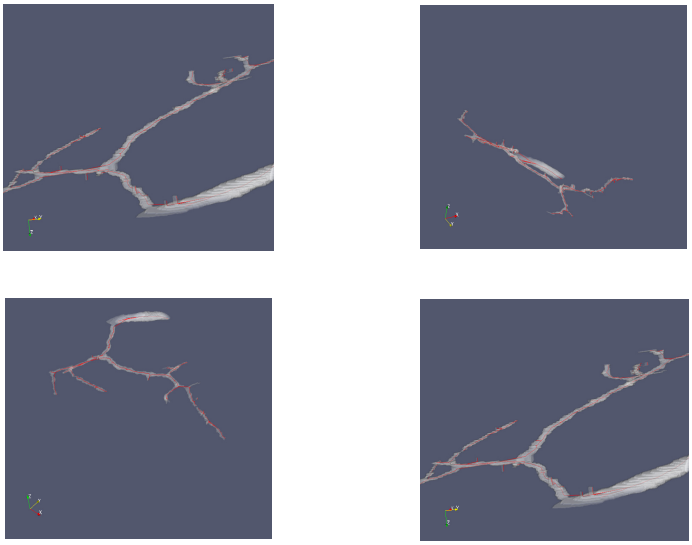


Figure 3.9: The figure shows the stack of the binarized image sequence which overlays with the traced vascular structure in Paraview.

3.6 Binarized image validation

For validation purpose, the ground truth image is generated by a human operator. Every vessel cross sectional region is marked by the human operator in every raw image. Figure 3.10 shows the raw image and the ground truth image. The ground truth image is then binarized by the pixel intensity '255'; in the binarized ground truth images, true vessel cross sectoinal regions consist of the pixel intensity '255', otherwise '0'. Pixel intensities in binarized ground truth images are reduced by 30%, added into the raw images, and the added images are subtracted by the binarized images. As a result, the ground truth image, the binarized image, and the raw image are superimposed into a new image; the superimposed result image represents both the ground truth and the binarized vessel cross sectional regions. Besides, the pixelwise reduced ground truth images are added into the raw images for a human operator's manual work which is described next.

To prevent the human operator from being error prone on manual work, the superimposed image and the renovated ground truth images are stitched together so that when the human operator manually makrs noise (true negatives) on the superimposed image,

the human operator can refer to the ground truth image at the same time. Finally, the human operator makes less mistakes while marking on the superimposed images. Figure 3.11 shows the stitched image where the superimposed images and the renovated ground truth images exist side by side. To validate binarized image quantitively, precision, recall, and F2measure are introduced. Precision is known as the proportion of retrieved positive cases which is correctly real positives. In other words, it is True Positive Accuracy (tpa). It is defined as the following

$$Precision = \frac{TP(TruePositive)}{PP(PredictedPositive).}$$

Recall is known as the the proportion of Real Positive that are Predicted Positive. It also shows how much results are relevant. In other words, it is True Positive Rate. It is defined as the following

$$Recall = \frac{TP(TruePositive)}{RP(RealPositives).}$$

In this thesis, F2 measure is used, which sets $\beta = 2$. By adjusting of β, either recall or precision is emphasized in the equation [30]. F2 measure puts more emphasis on recall. Precision, recall, and F2 measure are defined as follows:

$$True\ positive(tp) = The\ number\ of\ true\ binarized\ vessel\ cross\ sectional\ region$$

$$False\ positive(fp) = The\ number\ of\ noise\ binarized\ as\ the\ vessel\ cross\ sectional\ region$$

$$False\ negative(fn) = The\ numberof\ vessel\ cross\ sectional\ regions\ that\ have\ not\ been\ binarized$$

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn.}$$

$$F2measure = \frac{(1+\beta)^2\ precision\ recall}{(\beta^2\ precision) + recall}$$

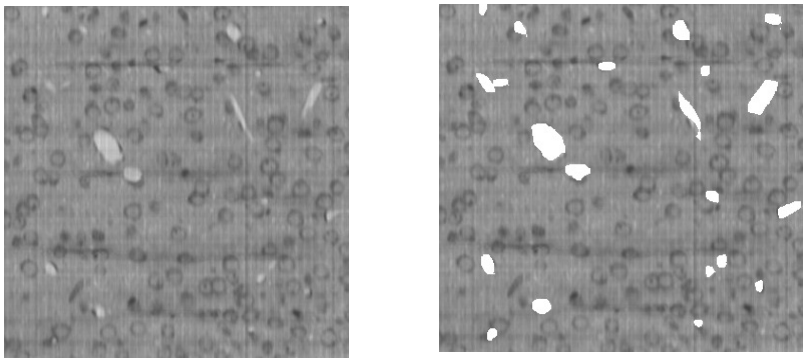


Figure 3.10: The image (left) is the raw image. The other image (right) is the image with the ground truth. The ground truth is the white colored region.

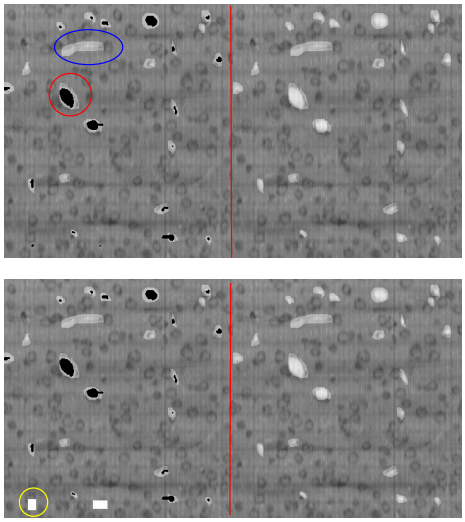


Figure 3.11: The image (top) is the stitched image not marked by the human operator. The other image (bottom) is the stiched image marked by the human operator. The black colored region marked by the red circle is the binarized vessel cross sectional region. The white blended region inside the blue circle is the ground truth. The white rectangular region inside the yellow circle is the true negative the human operator marked.

3.7 Matlab parallel for loop (parfor) and computing cluster utilization

The methods for this research are implemented in MATLAB. Because the size of the image sequence can increase enormously, it causes more likely out-of-memory error and a slow computation on the image sequence. To increase the performance of the implementation, parallel for loop (parfor) is used. Parfor allows loop iterations from the MATLAB client to be run in parallel on a parallel pool where consists of MATLAB workers. MATLAB workers handle iterations of parfor in out-of-order fashion [16]. In this sense, the MATLAB workers can run a main loop concurrently so that both low latency and high throughput of overall process are plausible. Ada cluster, a computing cluster at Texas A&M, has 32 licensed CPUs with MATLAB so that any MATLAB code with parfor can be executed in it [9].

3.8 Data analysis and quantitive volume reconstruction

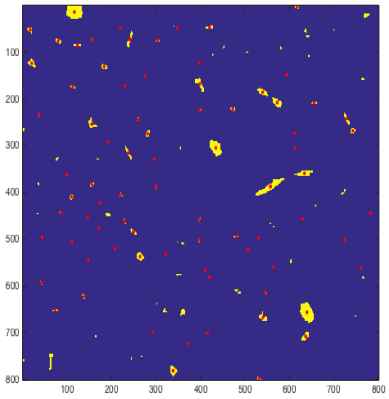
From the stack of the binarized image sequence, anatomical statistics of the vasculature are calculated such as a diameter average of all vessel cross sectional regions, a diameter standard deviation, a maximum diameter, a median diameter, and a vasculature length. To measure the diameter, the vessel radius is calculated by a 6-neighborhood exhaustive search around the centroid on the vessel cross sectional region. The mean of the radii is added to the standard deviation that is empirically reduced. Figure 3.13 shows how both the mean of the radii and the standard deviations are measured.

While the diameter is being estimated, every estimated diameter is collected. The following statistics are calculated based on the collection of the diameters. An average diameter is calculated by the following equation: average($x_1 ... x_n$) where $x_1 ... x_n$ are the collection of diameters. A maximum diameter is calculated by the following equation: $\max(x_1 ... x_n)$ where $x_1 ... x_n$ are the collection of diameters. A median diamter is measured by the following equation: median($x_1 ... x_n$) where $x_1 ... x_n$ are the collection of diameters. A minimum diameter is measured by the following equation: min($x_1 ... x_n$) where $x_1 ... x_n$ are the collection of diameters.

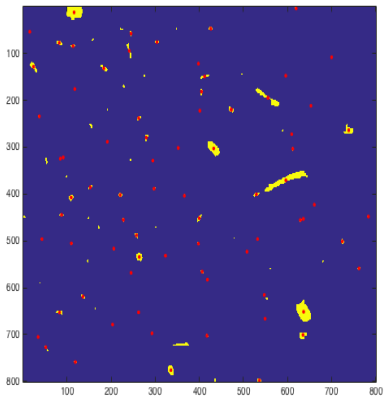
The vasculature length is estimated using the following formula: $\sum_{i=1}^{n-1} \|C_{i+1} - C_i\|$ where C_i is the centroid of the connected vessel cross sectional region, and n is the last vessel cross sectional region of the vasculature. With all the vasculature length estimations from the vasculatures, their average, standard deviation, maximum, median, and minimum are calculated.

Using estimated diameters, the volume of the traced vascular structure is quantitatively reconstructed. To quantitatively reconstruct the volume, every vessel cross sectional regions make up the volume of the traced vascular structure. In other words, the volume is the summation of the areas of all vessel cross sectional regions. The volume formula is defined as follows: Let the area of each vessel cross sectional region be $S_i = \pi\ r^2$ where r is the estimated radius; The volume $\sum_{i=1}^{m} \sum_{j=1}^{n} S_j$ is estimated, where n is the number of vessel cross sectional regions in an image, and m is the depth of the volume (the number of images in the image sequence). To validate the reconstructed quantitive volume, the ratio of the binarized image volume to the reconstructed volume is estimated. In the end, the higher the ratio of the volumes, the more accurate the diameter estimation is.

Besides estimating the length and the diameter, the number of bifurcating vessel cross sectional regions where the blood could flow in more than one direction is estimated. To mark the bifurcation point, the number of adjacenct vessel cross sectional regions in a subsequnt image is considered. If the number of adjacent vessel cross sectional regions is more than one, the vessel cross sectional region is considered as the bifurcation point. Figure 3.14 shows how the bifurcating vessel cross sectional region is adjacent with other vessel cross sectional regions in the subsequent image.



Centroids extracted in (X,Y,127)



Centroids extracted in (X,Y,128)

Figure 3.12: Centroids extracted for every vessel cross sectional region on two different images. Red dots represent the centroids.

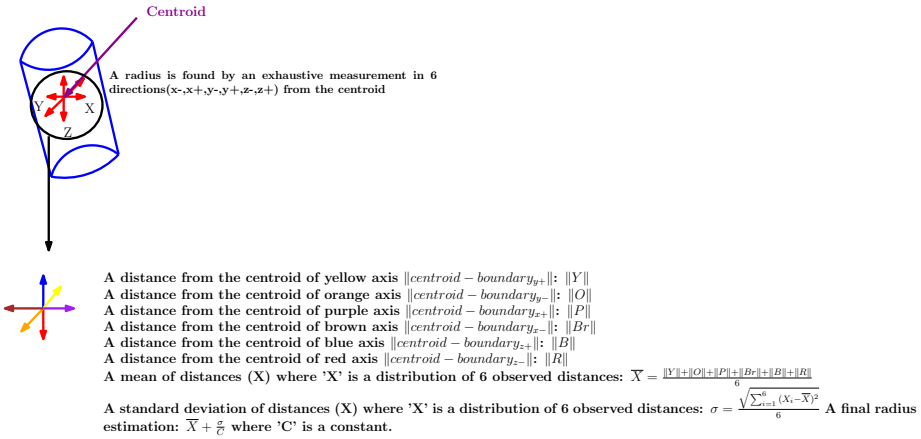


Figure 3.13: The figure describes how a mean of radii is calculated and adjusted by its standard deviation.

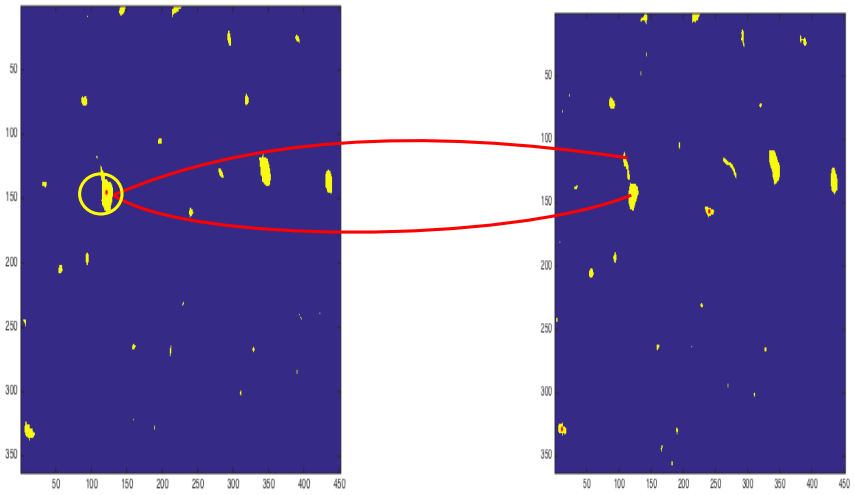


Figure 3.14: The figure describes the bifurcating vessel cross sectional regions through two subsequent images. The vessel cross sectional region inside the yellow circle represents the bifurcation point since it is adjacent to two vessel cross sectional regions in the subsequent image (presented by two red curves).

4. RESULTS AND ANALYSIS

Previously, the methods for binarization, vascular structure tracing, and quantitive analysis were introduced. In this section, using the described methods, raw images are enhanced and binarized. In the end, the binarized images' centerlines are traced. Additionally, the anatomical statistics are also extracted from the binarized images. Speedups are also compared using execution times which are measured with different variations of the computing cluster resources. In this section, five image sequences are tested and analyzed: Image sequence1 (376 images in their width 240 and height 264), Image sequence2 (324 images in their width 320 and height 268), Image sequence3 (427 images in their width 615 and height 615), Image sequence4 (652 images in their width 750 and height 750), and Image sequence5 (652 images in their width 1434 and height 1434). An image sequence (100 images in their width 363 and height 367) is quantitively validated for the binarization method by precision, recall, and F2-measure.

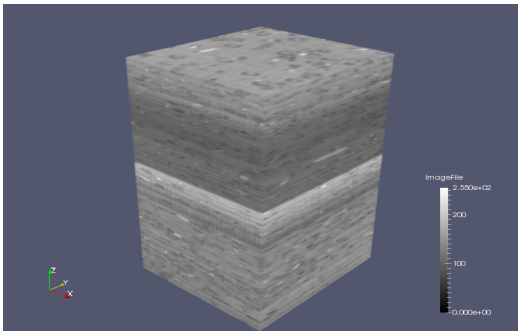
4.1 Results of binarization

As discussed earlier, the vessel cross sectional regions are binarized from the raw 2D images. The segmentation method was tested on 50 raw images. Each 2D image has its width 363 and height 367. The raw images are binarized by the previously propsed methods. For validation purpose, the ground truth is marked by a human operator. For every 2D raw images, precision, recall, and F2measure are used. Table 4.1 shows the average of precision, recall, and F2-measure from the binarized result images.

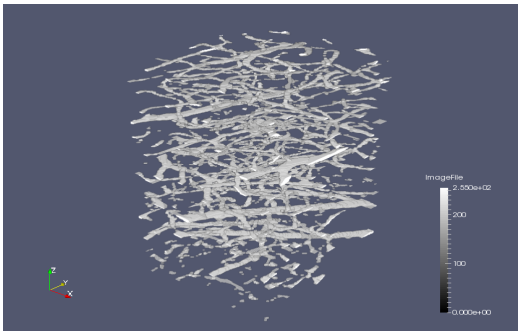
Finally, three results reflect that the binarizatoin method works well. Granted that the binaraization method works well, five raw image sequences are binarized and visualized via Paraview. Figure 4.1 to 4.5 illustrates the visualization of the averaged image sequence, the binarized vessel cross sectional regions, and the overlayed image sequences via Paraview.

Precision, Recall, and F2-measure		
Precision	Recall	F2-measure
0.9235	0.9389	0.9828

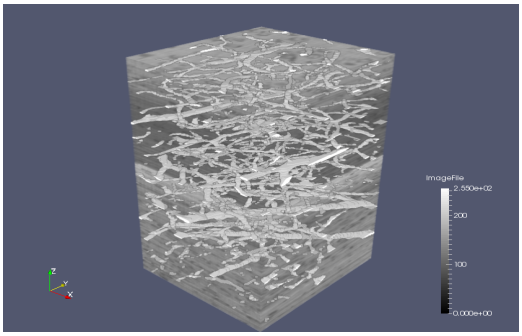
Table 4.1: The table shows the precision, recall, and F2-measure of the binarized result images.



(a) Image A

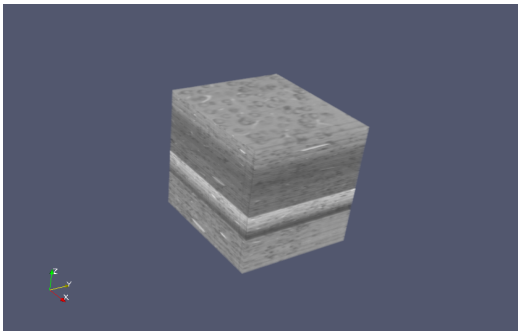


(b) Image B



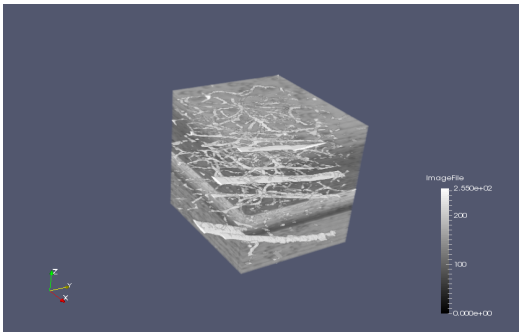
(c) Image C

Figure 4.1: Image A represents an averaged image sequence (376 images in their width 240 and height 264). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



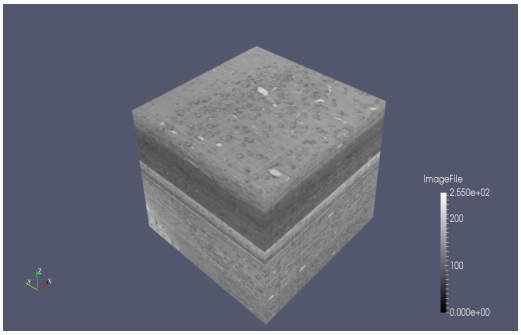
(a) Image A

(b) Image B

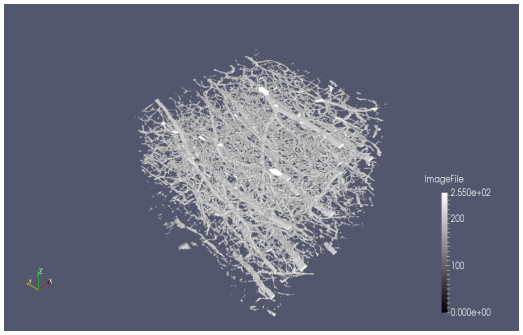


(c) Image C

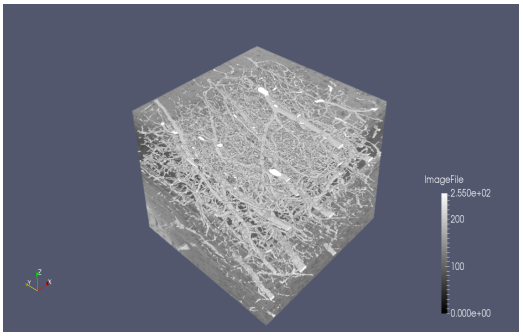
Figure 4.2: Image A represents an averaged image sequence (324 images in their width 320 and height 268). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

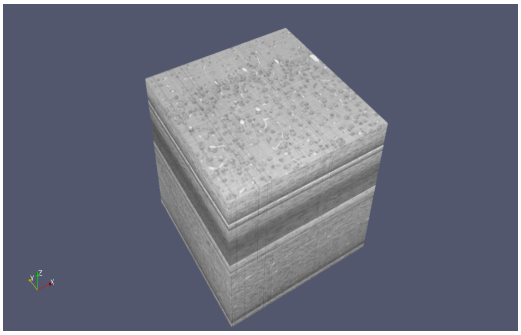


(b) Image B

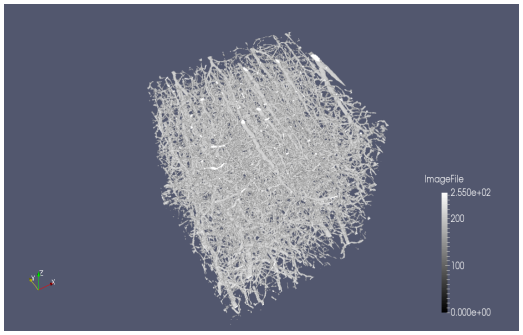


(c) Image C

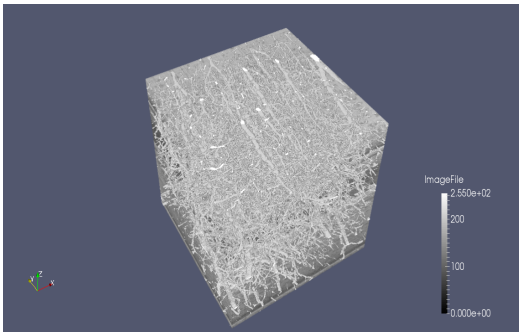
Figure 4.3: Image A represents an averaged image sequence (427 images in their width 615 and height 615). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

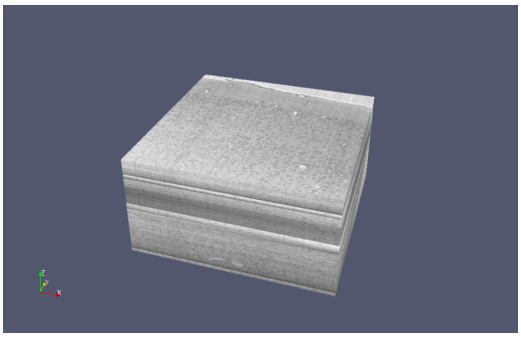


(b) Image B

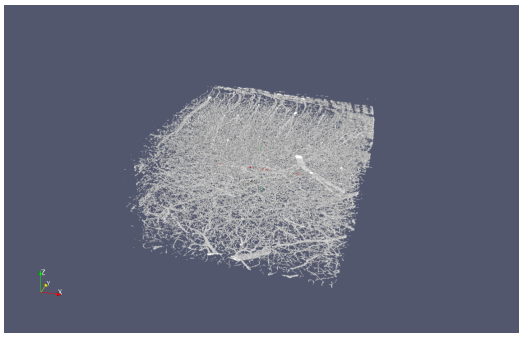


(c) Image C

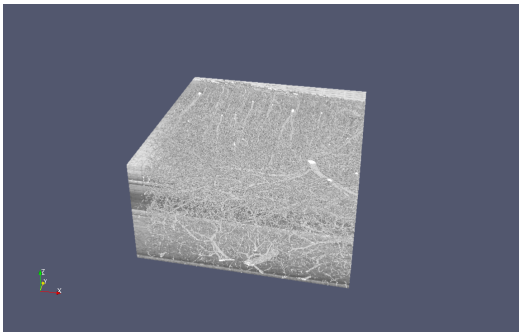
Figure 4.4: Image A represents an averaged image sequence (652 images in their width 750 and height 750). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A



(b) Image B



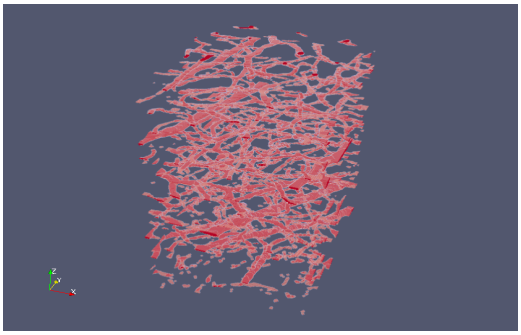
(c) Image C

Figure 4.5: Image A represents an averaged image sequence (652 images in their width 1434 and height 1434). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.

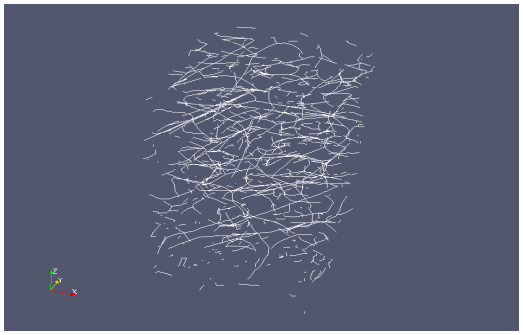
4.2 Results of vascular structure tracing and anatomical statistics

The vascular tracing method is tested on Image sequence 1 to 5. First, raw input images are binarized; every centroids of vessel cross setional region is detected; And based on them, vascular structure information is generated in VTK file format. Generated VTK files are visualized by Paraview. Finally, the visualized vascular structure is overlaid with the stack of the binarized image sequence for its visual inspection purpose. Figure 4.6 to 4.10 show visualizations of both the stack of the binarized image sequence and of the traced vascular structure. Overall, the manual visual inspection of the traced vascular structures concludes that all the stacks of the binarized image sequence is traced well. In addition to tracing the vascular structure based on the stack of the binarized image sequnence, in this section, several anatomical statistics of the binarized image sequence 1 to 5 are estimated:

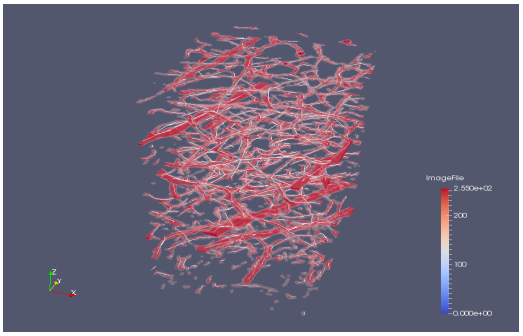
1. A diameter average of all vessel cross sectional regions
2. A diameter standard deviation
3. A maximum diameter
4. A median diameter
5. A vasculature length average
6. A vasculature length standard deviation
7. A vasculature length median
8. A vasculature length maximum
9. A vasculautre length minimum
10. A number of bifurcation points



(a) Image A

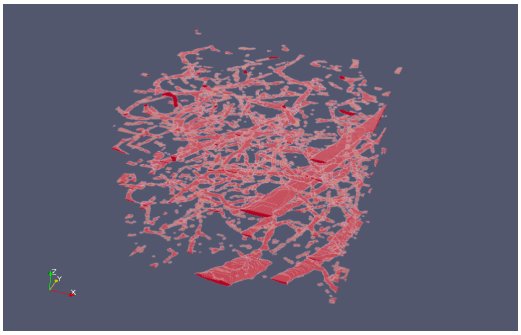


(b) Image B

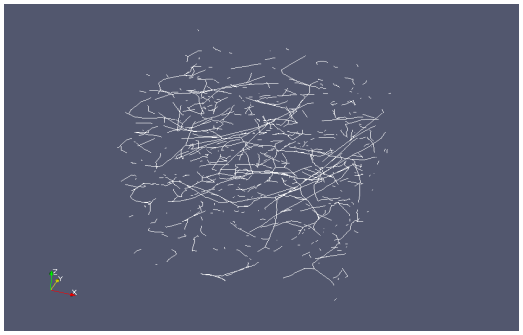


(c) Image C

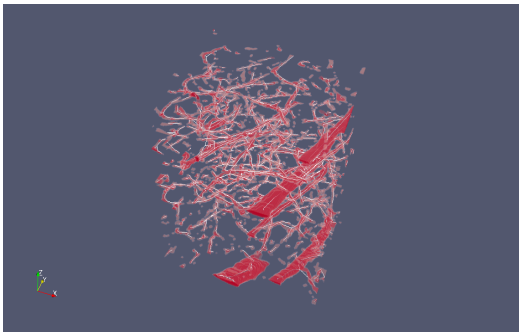
Figure 4.6: Image A represents an vessel cross sectional region image sequence (376 images in their width 240 and height 264). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

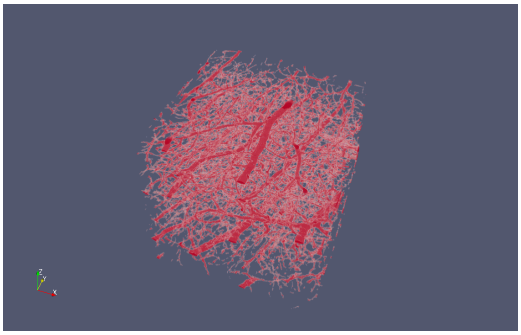


(b) Image B

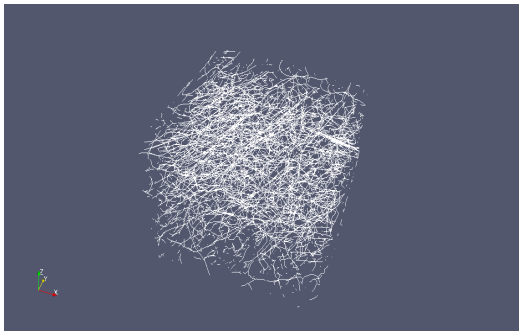


(c) Image C

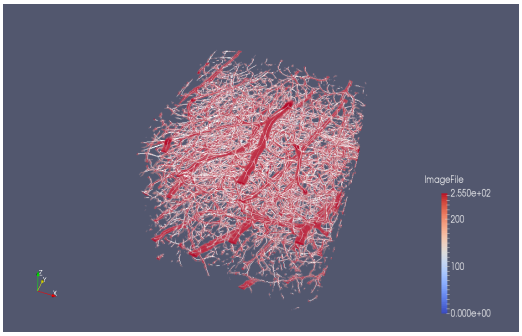
Figure 4.7: Image A represents an vessel cross sectional region image sequence (324 images in their width 320 and height 268). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

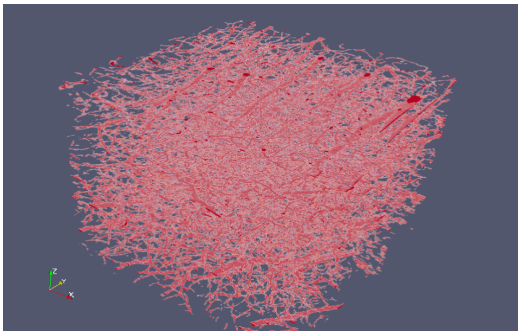


(b) Image B

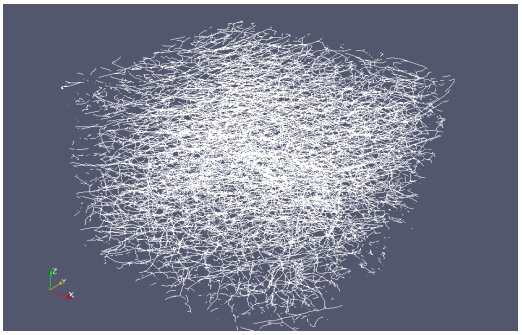


(c) Image C

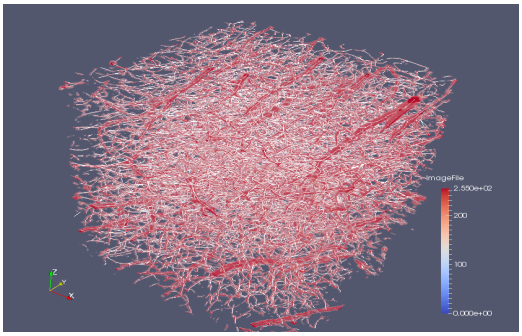
Figure 4.8: Image A represents an vessel cross sectional region image sequence (427 images in their width 615 and height 615). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

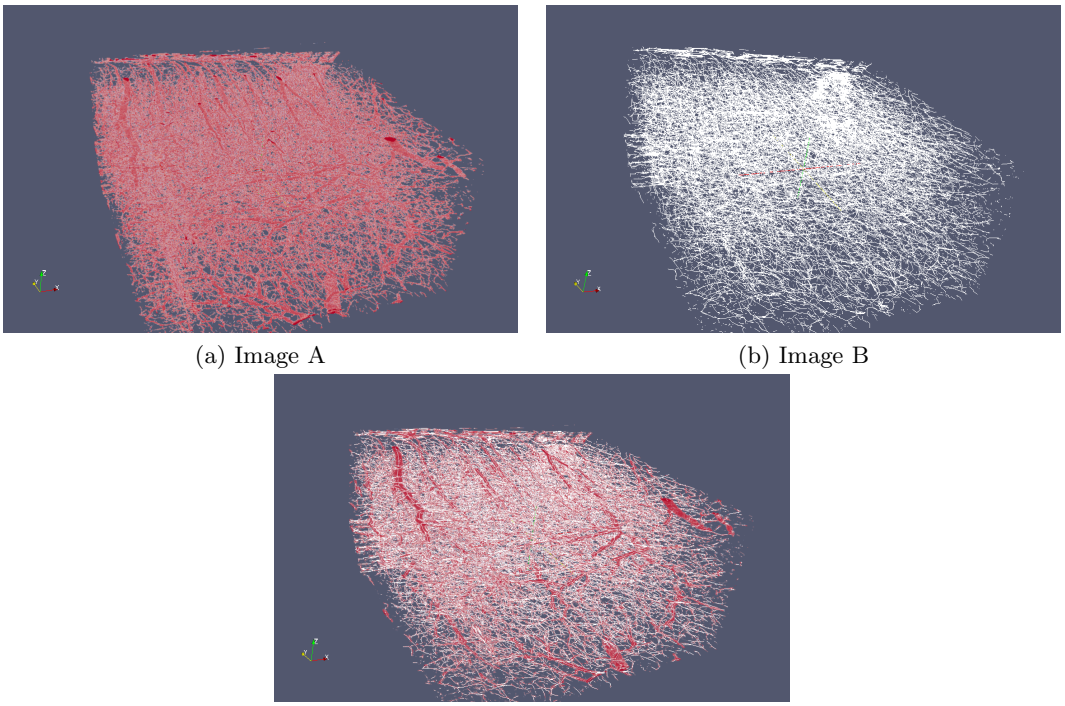


(b) Image B



(c) Image C

Figure 4.9: Image A represents an vessel cross sectional region image sequence (652 images in their width 750 and height 750). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.



(a) Image A

(b) Image B

(c) Image C

Figure 4.10: Image A represents an vessel cross sectional region image sequence (652 images in their width 1434 and height 1434). Image B represents a binarized image sequence of Image A. Image C represents an overlayed image sequence of both Image A and Image B.

Anatomical statistic in micron(μm) using diameters							
AxBxC(Width:A Height:B Depth:C)	constant variable for diameter adjustment	Diameter average	Diameter standard deviation	Maximum diameter	Minimum diameter	Median diameter	the ratio of the volumes
240x264x376	2.2	6.0546	2.9632	36.5394	1.2	5.2278	0.98485
320x268x324	2.2	5.8432	3.6533	44.4314	1.2	4.8907	0.98365
615x615x427	2.2	5.5626	3.5015	68.258	1.2	4.6817	0.99779
750x750x652	3	5.3022	3.0589	76.0219	1.1633	4.5347	0.90559
1434x1434x652	3	5.2128	3.4615	102.0666	1.1633	4.3155	0.92807

Table 4.2: The table shows the statistics in vessel cross sectional region's diameters. The second column (Constant variable for diameter estimation) represents the constant value that is used to adjust the mean of the radii which is previously discussed

Anatomical statistic in millimeter using vasculature length per cubic millimeter of tissue					
AxBxC(Width:A Height:B Depth:C)	A vasculature length average	A vasculature length standard deviation	A vasculature length median	A vasculature length maximum	A vasculautre length minimum
240x264x376	5.3519	14.8973	1.3652	136.5272	0.046281
320x268x324	2.8426	7.9084	0.96287	75.7943	0.015233
615x615x427	1.2765	30.2859	0.1522	1263.8416	0.00053512
750x750x652	0.58518	18.247	0.077599	1137.4818	0.0010536
1434x1434x652	0.10949574	4.8468264	0.01778614	600.77978144	0.000574542

Table 4.3: The table shows the statistics in vascular structure's length

Anatomical statistic using the number of branch points per cubic millimeter of tissue	
AxBxC(Width:A Height:B Depth:C)	the number of branch points
240x264x376	10094.1
320x268x324	4969.9
615x615x427	12634.3
750x750x652	13425.5
1434x1434x652	9673.0

Table 4.4: The table shows the statistics of the number of branch points

ALL STATISTICS ARE SPECIFIED PER CUBIC MILLIMETER OF TISSUE.						
Region	Segments	Length (mm)	Branches	Surface (mm^2)	Volume (mm^3)	Volume (% of total)
Neocortex	11459.7	758.5	9100.0	10.40	0.0140	1.4%
Cerebellum	34911.3	1676.4	19034.4	20.0	0.0252	2.5%
Spinal Cord	36791.7	1927.6	26449.1	22.2	0.0236	2.4%

Table 4.5: Computed statistics from Mayerich et al. [18]. Reprinted from [18]

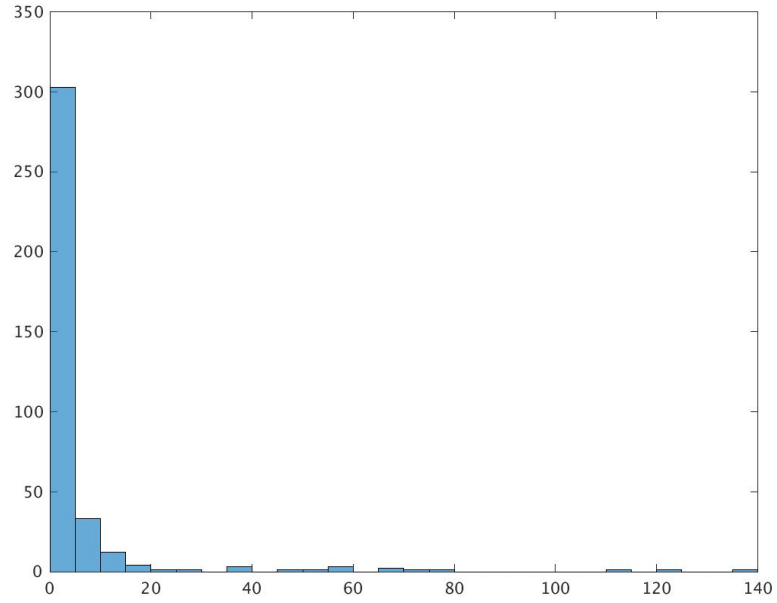


Figure 4.11: A histogram of vessel segment lengths in 240x264x376.

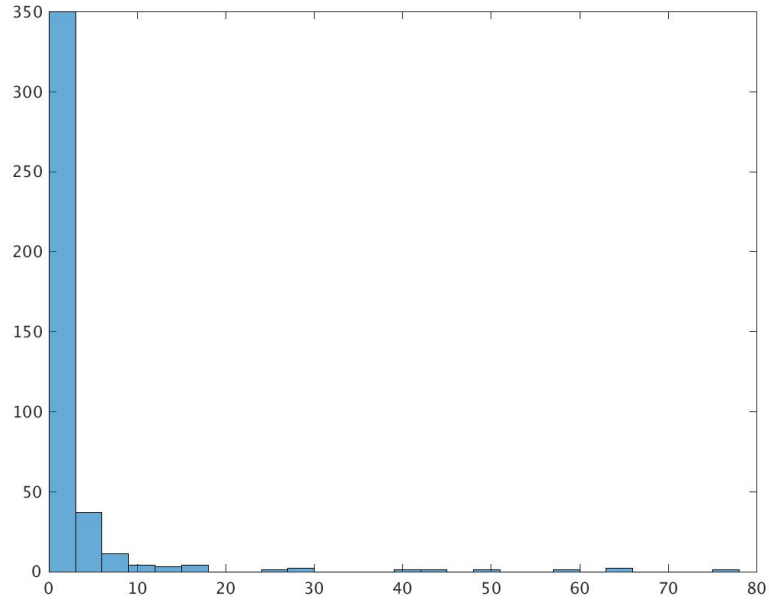


Figure 4.12: A histogram of vessel segment lengths in 320x268x324.

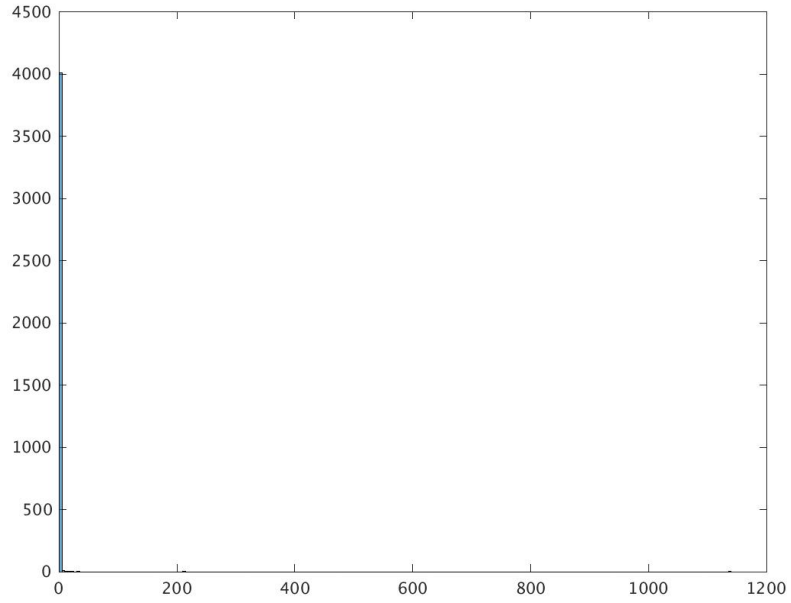


Figure 4.13: A histogram of vessel segment lengths in 750x750x652.

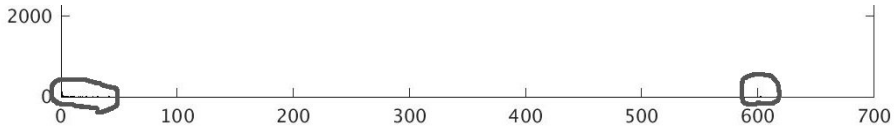


Figure 4.14: A histogram of vessel segment lengths in 1434x1434x652. Most of distribution occurs inside gray circles.

Table 4.2 shows the statistics in the vessel cross sectioanl region's diameters. One interesting thing observed from the table 4.2 is the ratio of the volumes. Since the ratios are all above 90%, the results of the diameter estimation seem good. The diameter of microvasculature is known to have a diameter of around 5 µm [18], so the average diameter from the table seems reasonable. Table 4.3 shows the anotomical statistics using the vascular lengths. Refering to figures 4.11 to 4.14, the histograms show skewness in their distribution. Therefore, the median values don't seem close to the average values in table 4.3. Table 4.4 shows the number of branch points. Mayerich et al. [18] quantified the number of the branch points from parts of a mouse brain. Table 4.5 shows the results from Mayerich et al. Because the number of branch points from Mayerich et al. was measured per cubic millimeter, the number of branch points of mine is also scaled in terms of cubic millimeter. Since the KESM rat Nissl data is sectioned from the somatosensory cortex, we focus on Neocortex of the mouse brain stained with India ink from Mayerich et al. The average of the number of branch points from Table 4 is 10159.36. Comparing branch number statistic of Mayerich et al. to that of mine, his result and my result seem to obtain a number close to 9000. In the end, the number of branch points from the rat and the mouse seem similar each other in their particular regions.

4.3 Analysis of tracing speed in the computing cluster

In this thesis, instead of using a local computer, Ada, a computing cluster in Texas A&M, is used to increase the throughput of tracing method. Ada is a 17,5000-core IBM commodity cluster with nodes based mostly on Intel's 64-bit 10-core IvyBridge processors; Its memory size is 811 nodes with 64 GB/node and 34 nodes with 256GB (DDR3 1866 MHZ) [8]. Fortunately, Ada offers a simple program matlabsubmit for batch jobs in Matlab. Following options are available for matlabsubmit to run batch jobs with different resources:

1. -h Shows this message
2. -m set the amount of requested memory
3. -t sets the walltime; form hh:mm:ss

4. -w sets the number of additional workers

5. -n sets the number of nodes to assign the workers to

6. -g indicates script needs GPU

7. -b sets the billing account to use

8. -q sets the queue to be used

In this section, only '-m','-t','-w', and '-n' are considered for the processing time experiment [9].

Paralell for loop (PARFOR) in MATLAB is also fully exploited in the computing cluster. Since PARFOR allows a main iteration of the tracing method to be run in parallel on a cluster of workers, the throughput of tracing method improves. With different variations of the options, different processing times are observed from Image sequence1 to 5 mentioned in section 4.

To observe the performance difference according to the variations of the options, 5 different cases of both '-w' and '-n' are observed as follows: for both '-w' and '-n', '1','2','4','8','16', and '31'. To present better comparison, speedups are measured as follow: $speedup = \frac{A}{B}$ where A presents the computation time with a node and worker, and B presents the computation time with more than a node and worker. In other words, the speedups reflect the performance improvement as more resources in the cluster are consumed. However, since the network overhead between a client and workers happens when the client distributes the processing raw images equitably into workers and combines partial results from them, speedup results turn out to be lower than '30'.

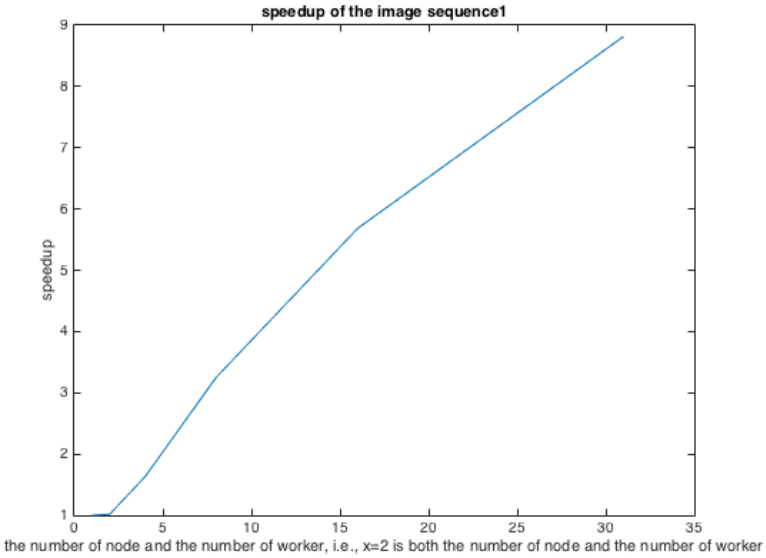


Figure 4.15: The figure describes how the speedup of the image sequence1 changes as the number of nodes and the number of workers change. For further description on the plot, when x = 2, it represents that both the number of nodes and the number of workers are 2.

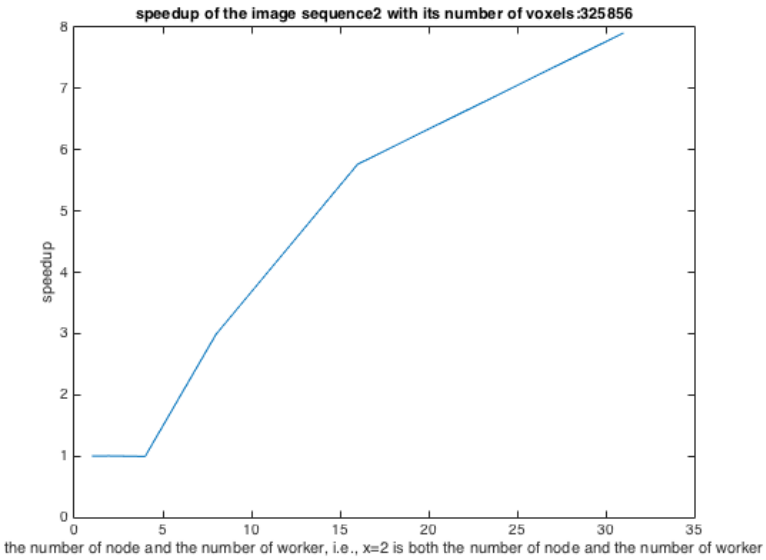


Figure 4.16: The figure describes how the speedup of the image sequence2 changes as the number of nodes and the number of workers change. For further description on the plot, when x = 2, it represents that both the number of nodes and the number of workers are 2.

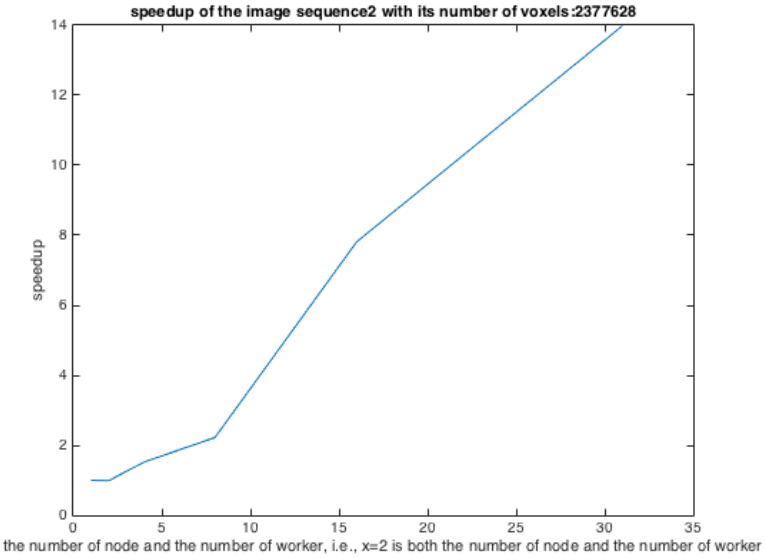


Figure 4.17: The figure describes how the speedup of the image sequence3 changes as the number of nodes and the number of workers change. For further description on the plot, when x = 2, it represents that both the number of nodes and the number of workers are 2.

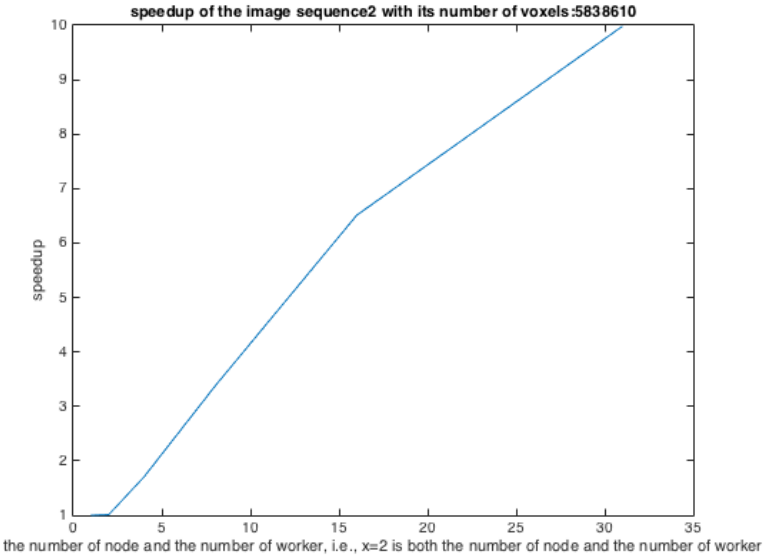


Figure 4.18: The figure describes how the speedup of the image sequence4 changes as the number of nodes and the number of workers change. For further description on the plot, when x = 2, it represents that both the number of nodes and the number of workers are 2.

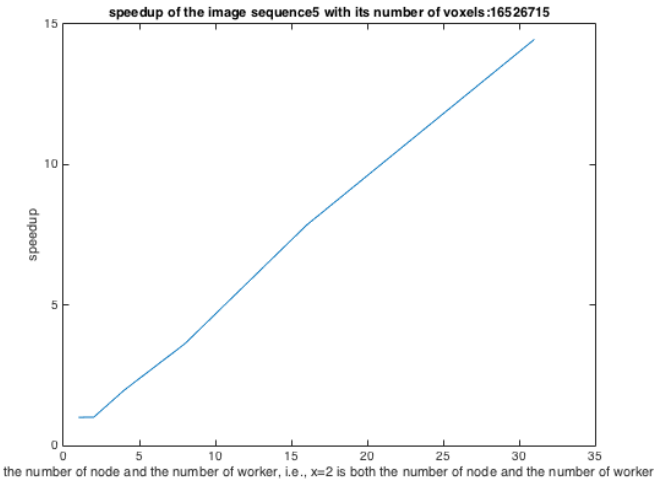


Figure 4.19: The figure describes how the speedup of the image sequence5 changes as the number of nodes and the number of workers change. For further description on the plot, when x = 2, it represents that both the number of nodes and the number of workers are 2.

In the end, figures 4.15 to 4.19 reflect that the more resources the tracing method consumes the more the tracing method performance improves. Since the performance reflects the local computer's performance when x = 1 in the plots, the speedup also reflects how much the computing cluster is contributed to the performance improvement. In other words, exploiting the computing cluster enables the tracing large vascular structure within a reasonalbe time. In the end, the usage of the computing cluster is necessary since the KESM rat Nissl data can become larger and larger to be processed. Thanks to more sophisticated algorithm compatible with the computing cluster to process the KESM rat Nissl data, the analysis of whole rat brain vasculature can become more feasible.

5. DISCUSSION

In this thesis, the novel binarization and tracing method were presented. Since they are simple and efficient, they can easily be adaptable with the large Nissl data. At the same time, the computing cluster is contributed to the increase of the throughput. In this section, contribution, open issues, and future work in this thesis are discussed.

5.1 Contribution

The binarization and tracing method in this thesis accurately extracts the vessel cross sectional regions and traces the vascular structure. Addtionally, the tracing method enables obtaining the anatomical statistics such as the vascular structure length, the number of branch points, and the diameter estimation. With the obtained statistics, the interesting comparison between the anatomical statistic of the mouse and that of the rat can be done. After the vascular structure is traced, the result is stored as VTK files to be visualized via ParaView. People interested in the vascular structure can observe the appearacne of the vascular structure.

When the size of the data becomes larger, the computing cluster plays an important role to increase the throughput. In the analysis of the tracing speed on the computing cluster, the computing cluster has shown its potential in processing the large dataset within reasonable time. In summary, the methods presented in this thesis perform well on the Nissl datasets, and the tracing methods can be applied to other KESM datasets such as the India ink dataset.

5.2 Open issues and future work

In this thesis, while the resuts look good, three issues have been found: (1) After the binarization step, noise is still present in the result image. (2) When the vessel runs along the image plane (i.e., laterally), the shape of the vessel cross sectional regions become elongated, and as a result, the bifurcation point quauntization step may become error prone. (3) The largest image sequence visualized and analyzed is 652 images, width:1434, and height: 1434. However, once the image sequence size exceeds the larget size in this thesis, it cannot be visualized via ParaView. (4) If one of two subsequent images contains noise which would overlap with the vessel cross sectional region in the other subsequent image, the guidance map would result in containing the noise coordinate information.

Although the binarization step in this thesis results in successful vessel cross sectional region extraction, the noise is still present in the result images. Figure 5.1 shows the noise still present after the binarization step. The noise will have a negative affect on tracing vascular structure and computing the anatomical statistics. To completely remove the noise, more careful observation needs to be made on the raw images, and using the observation, more sophisticated method need to be developed.

The vessel cross sectional regions that are compatible with the proposed methods are usually developing towards vertical orientation. Due to their developing orientation, they are usually reflected in the shape of circle or ellipse in the image. However, some vessel cross sectional regions are developing along an orientation other than the vertical orientation. Therefore, the proposed methods become erroneous on the misoriented vessel cross sectional regions. Figure 5.2 shows the case when the misoriented vessel cross sectional regions are misclassified as the bifurcation point. To resolve the case of the misoriented vessel cross sectional regions, more careful regional property of the vessel cross sectional regions should be observed.

Minimum intensity of more than two subsequent images would be able to result in the more accurate guidance map where only contains the coordinate information of the vessel cross sectional regions. For example, if one of two subsequent images contains the noise

which would overlap with the vessel cross sectional region in the other subsequent image, the guidance map would result in containing the noise coordinate information. To resolve such an exceptional case, more than two subsequent images can be used to result in the guidance map. For example, let three subsequent images be contributed to generate the guidance map. If image1 out of three images contained the noise, it might be overlapped with vessel cross sectional regions from image2. However, the noise from the image1 would less likely be overlapped with vessel cross sectional regions from image3. In the end, the guidance map from more than two subsequent images would resolve the exceptional case.

As already discussed in the earlier section, the visual inspection via ParaView has been proposed. Although ParaView is famous for visualizing the large datasets, it has been empirically observed that ParaView cannot visualize the data exceeding the certain large size. Since the visual inspection step is crucial to validate the vascular structure, other methods to visualize the large Nissl data need to be sought.

The averaging operation can be done onto 2D images generated from different orientations. For example, 2D raw images can be stacked in 3D, and the original 3D volume can be sectioned into 2D images in different orientations such as x-axis or y-axis. In the end, the averaging operation can be done onto new 2D images of different orientations, and result enhanced images can be used for tracing their vascular structures and anatomical statistics.

The tracing method proposed in this thesis works on any binarized vasulature data. The India ink-stained mouse brain can be also binarized, and it is adaptable with the tracing method. Afterwards, from the rat and mouse data, using the proposed methods to compute the anatomical statistics, more of their anatomical statistics can be revealed and compared each other.

A synthetic model of the KESM rat Nissl data can prove whether the KESM rat Nissl images (digital phantom) from the KESM rat Nissl brain specimen (physical phantom) is correct. For example, a synthetic model of vasculatures can be generated, and cell bodies of neurons can be modeled as a sphere shaped object and be integrated with the modeled

vasculature. In the end, a synthetic model of rat Nissl data is generated representing the relationship between the vasculature and the cells. The geometrical data from the synthetic data enables a 3D printer to obtain its physical phantom. After the 3D printer results in a physical phantom of the digital phantom, the KESM can section the physical phantom and reconstruct its digital phantom. Using methods from my thesis, vascular structure and anatomical statistic result from both synthetic data and digital phantom data can be compared to each other. If the results turn out to be matched, the reconstructed digital phantom from KESM is believed to be true. In the end, the acquired KESM image is considered true once the vascular structure and anatomical statistics from both the synthetic model and the digital phantom are matched.

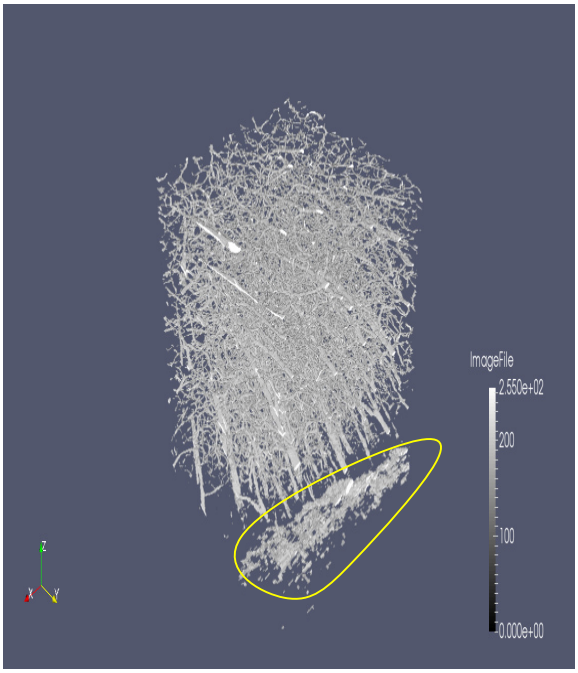


Figure 5.1: Noise is still present after the binarization step (inside yellow circle).

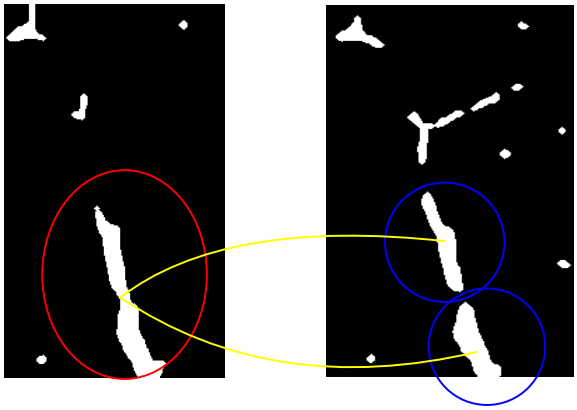


Figure 5.2: The vessel cross sectional region inside the red circle are classified as the branch point because it has two adjacent vessel cross sectional regions (inside blue circles) in the subsequent image. However, the detected branch point is misoriented so that the proposed method misclassified it as the branch point.

6. CONCLUSION

The main goal of this research was to efficiently extract vessel cross sectional regions, to trace the vascular structure, to compute the anatomical statistics, and to fully exploit the computing cluster to increase the throughput from the KESM rat Nissl-stained images. The methods in this thesis are considered novel since they are customized to the specific characteristics of the KESM data. The computing cluster demonstrated its potential for processing the large dataset within reasonable time. The comparison of anatomical statistics between rat and mouse leads to another interesting aspect of this research. The binarization step was quantitively validated by precision, recall, and F2-measure, and the diameter estimation by its quantitive volume reconstruction. The estimated structural statistics were comparable to earlier reports on a similar KESM data set. The methods proposed in this thesis can be extended to the other KESM datasets, and they will be contributed to analyze the vascular system in mouse and rat brain.

REFERENCES

[1] James Ahrens, Berk Geveci, and Charles Law. 36 paraview: An end-user tool for large-data visualization. *The Visualization Handbook*, page 717, 2005.

[2] K. Al-Kofahi, S. Lasek, D. Szarowski, C. Pace, G. Nagy, J. Turner, and B. Roysam. Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Transactions on Information Technology in Biomedicine*, 6:171–186, 2002.

[3] T Balić Žunić and Emil Makovicky. Determination of the centroid orthe best centre'of a coordination polyhedron. *Acta Crystallographica Section B: Structural Science*, 52(1):78–81, 1996.

[4] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[5] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *IEEE Transactions on Information Technology in Biomedicine*, 3:125–138, 1999.

[6] Y. Choe, D. Mayerich, J. Kwon, D. E. Miller, J. R. Chung, C. Sung, J. Keyser, and L. C. Abbott. Knife-edge scanning microscopy for connectomics research. *Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE*, 2011.

[7] P. Doddapaneni. Segmentation strategies for polymerized volume data sets. Department of Computer Science, Texas A&M University, 2004.

[8] Texas A&M Supercomputing Facility. Hpc systems. `"http://sc.tamu.edu/systems/#ada"`.

[9] Texas A&M Supercomputing Facility. Matlab faq. `"http://sc.tamu.edu/help/faq/matlab.php"`.

[10] Gargouri Fatma. Thresholding the maximum entropy. `"http://www.mathworks.com/matlabcentral/fileexchange/35158-thresholding-the-maximum-entropy"`.

MATLAB Central File Exchange.Retrieved Feb10, 2015.

[11] Allen Institute for Brain Science. Allen bran atlas. "http://www.brain-map.org/". Accessed Aug. 2007.

[12] Rafael C Gonzalez. *Digital image processing*. Pearson Education India, 2009.

[13] Dong Hyeop Han. *Rapid three-dimensional tracing of the mouse brain neurovasculature with local maximum intensity projection and moving windows*. Texas A & M University, 2009.

[14] Jagat Narain Kapur, Prasanna K Sahoo, and Andrew KC Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing*, 29(3):273–285, 1985.

[15] A. MacKenzie-Graham, E. S. Jones, D. W. Shattuck, I. D. Dinov, M. Bota, and A. W. Toga. The informatics of a c57bl/6j mouse brain atlas. *Neuroinformatics*, 1:397—-410, 2003.

[16] MathWorks. Introduction to parfor. "http://www.mathworks.com/help/distcomp/introduction-to-parfor.html".

[17] D. Mayerich, L. Abbott, and B. H. McCormick. Knife-edge scanning microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain. *Journal of Microscopy*, pages 134—-143, 2008.

[18] D. Mayerich, J. Kwon, Y. Choe, L. Abbott, and J. Keyser. Constructing high- resolution microvascular models. In *Proceedings of the 3rd International Workshop on Micro- scopic Image Analysis with Applications in Biology (MIAAB 2008)*, 2008.

[19] David Mayerich, Jaerock Kwon, Yoonsuck Choe, Louise Abbott, and John Keyser. Constructing high resolution microvascular models. In *Third Workshop on Microscopic Image Analysis with Applications in Biology*, 2008.

[20] David M Mayerich, Louise Abbott, and John Keyser. Visualization of cellular and microvascular relationships. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1611–1618, 2008.

[21] B. H. McCormick. Development of the brain tissue scanner. *Brain Networks Lab*

Technical Report.

[22] B.H. McCormick. The knife-edge scanning microscope. "http://research.cs.tamu.edu/bnl/". Department of Computer Science, Texas A&M University, Tech.Rep.,2003.

[23] M. R. Metea and E. A. Newman. Glial cells both dilate and constrict blood vessels: a mechanism of neurovascular coupling. *Journal of Neuroscience*, 26:2862—2870, 2006.

[24] S. Mikula, I. Trotts, J. M. Stone, and E. G. Jones. Internet-enabled high- resolution brain mapping and virtual microscopy. *Neuroimage*, 35:9–15, 2007.

[25] S Bryan Morse. Lecture 2: Image processing review, neighbors, connected components, and distance. *Bringham Young University, Copyright Bryan S. Morse*, 2000, 1998.

[26] Nikhil R Pal and Sankar K Pal. A review on image segmentation techniques. *Pattern recognition*, 26(9):1277–1294, 1993.

[27] Kálmán Palágyi, Juerg Tschirren, Eric A Hoffman, and Milan Sonka. Quantitative analysis of pulmonary airway tree structures. *Computers in biology and medicine*, 36(9):974–996, 2006.

[28] R. B. Panerai. System identification of human cerebral blood flow regulatory mechanisms. *Cardiovascular Engineering International Journal*, 4:59—-71, 2004.

[29] Massimo Piccardi. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.

[30] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[31] B. Rosengarten, O. Huwendiek, and M. Kaps. Neurovascular coupling and cere- bral autoregulation can be described in terms of a control system. *Ultrasound Medical Biology*, 27:189–193, 2001.

[32] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature*

methods, 9(7):676–682, 2012.

[33] William J Schroeder, Lisa S Avila, and William Hoffman. Visualizing with vtk: a tutorial. *Computer Graphics and Applications, IEEE*, 20(5):20–27, 2000.

[34] S.G.Diamond, T.J. Huppert, V. Kolehmainen, and M. A. Franceschini. A dynamic physiological modeling for functional difuse optical tomography. *Neuroimage*, 30:88–101, 2006.

[35] Juan Shan, Heng-Da Cheng, and Yuxuan Wang. A novel automatic seed point selection algorithm for breast ultrasound images. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[36] R. Xiao, J. Yang, M. Goyal, Y. Liu, and Y. Wang. Automatic vasculature identification in coronary angiograms by adaptive geometrical tracking. *Computational and Mathematical Methods in Medicine*, 2013, 2013.

[37] Yi Yin, Mouloud Adel, and Salah Bourennane. Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognition*, 45:1235–1244, 2012.

[38] Yong Zhang, Xiaobo Zhou, Ju Lu, Jeff Lichtman, Donald Adjeroh, and Stephen TC Wong. 3d axon structure extraction and analysis in confocal fluorescence microscopy images. *Neural computation*, 20(8):1899–1927, 2008.

[39] B.V. Zlokovic. Neurovascular mechanisms of alzheimer's neurodegeneration. *Trends in Neurosciences*, 28:202—208, 2005.

[40] Z.Melek, D.Mayerich, C.Yuksel, and J.Keyser. Visualization of fibrous and thread-like data. *IEEE Transactions on Information Technology in Biomedicine*, 12:1165–1172, 2006.