

AUTOMATIC SEEDPOINT SELECTION AND TRACING OF
MICROSTRUCTURES IN THE KNIFE-EDGE SCANNING MICROSCOPE
MOUSE BRAIN DATA SET

A Thesis

by

DONGKUN KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

August 2011

Major Subject: Computer Science

AUTOMATIC SEEDPOINT SELECTION AND TRACING OF
MICROSTRUCTURES IN THE KNIFE-EDGE SCANNING MICROSCOPE
MOUSE BRAIN DATA SET

A Thesis

by

DONGKUN KIM

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

MASTER OF SCIENCE

Approved by:

Chair of Committee, Yoonsuck Choe
Committee Members, Eun Jung Kim
 Takashi Yamauchi

Head of Department, Valerie E. Taylor

August 2011

Major Subject: Computer Science

ABSTRACT

Automatic Seedpoint Selection and Tracing of
Microstructures in the Knife-Edge Scanning Microscope Mouse Brain Data Set.

(August 2011)

Dongkun Kim, B.S., Korea Military Academy

Chair of Advisory Committee: Dr. Yoonsuck Choe

The Knife-Edge Scanning Microscope (KESM) enables imaging of an entire mouse brain at sub-micrometer resolution. By using the data sets from the KESM, we can trace the neuronal and vascular structures of the whole mouse brain. I investigated an effective method for automatic seedpoint selection on 3D data sets from the KESM. Furthermore, based on the detected seedpoints, I counted the total number of somata and traced the neuronal structures in the KESM data sets.

In the first step, the acquired images from KESM were preprocessed as follows: inverting, noise filtering and contrast enhancement, merging, and stacking to create 3D volumes. Second, I used a morphological object detection algorithm to select seedpoints in the complex neuronal structures. Third, I used an interactive 3D seedpoint validation and a multi-scale approach to identify incorrectly detected somata due to the dense overlapping structures. Fourth, I counted the number of somata to investigate regional differences and morphological features of the mouse brain. Finally, I traced the neuronal structures using a local maximum intensity projection method that employs moving windows.

The contributions of this work include reducing time required for setting seedpoints, decreasing the number of falsely detected somata, and improving 3D neuronal reconstruction and analysis performance.

To my family

ACKNOWLEDGMENTS

I am indebted greatly to my advisor Dr. Yoonsuck Choe for his continuous support of my M.S. degree research, valuable comments, and advice. Besides my advisor, I would also like to thank Dr. Eun Jung Kim and Dr. Takashi Yamauchi for their contribution as committee members. Also, thanks to many Korean students in Department of Computer Science and Engineering at Texas A&M University: Chul Sung, Suinn Park, Yong Song, Jungwhan Kim, Baeksong Ahn, and Dr. Gangman Lee. I am also thankful of Timothy Mann, Daniel Evan Miller, and Roozbeh Daneshvar. Most importantly, I am grateful to my wife Eunmi Bae and my son Daewhi Kim for cheering me up. Also, I would like to thank the Republic of Korea Army for their generous funding. Part of this research was funded by NSF CRCNS grant #0905041 (PI: Yoonsuck Choe).

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
	A. Motivation	1
	B. Goal of the research	1
	C. Approach	2
	D. Significance	3
	E. Outline of the thesis	3
II	BACKGROUND AND PREVIOUS WORK	5
	A. Background	5
	1. Automatic seedpoint selection	5
	2. Interactive 3D seedpoint validation	7
	3. Multi-scale approach	8
	4. Soma counting	9
	B. Prior work	9
	1. Imaging with the Knife-Edge Scanning Microscope . .	10
	2. Tracing vasculature of the KESM image data sets . .	10
III	METHODS	14
	A. Preprocessing	15
	1. Data acquisition	15
	2. Inverting image	15
	3. Noise filtering	17
	a. Median filtering	18
	4. Contrast enhancement	19
	5. Merging image columns and 3D image stacking	20
	B. Automatic seedpoint selection	21
	1. Deciding the threshold value	22
	2. Morphological object detection (opening and closing) .	22
	a. Dilation	25
	b. Erosion	26
	3. Opening and closing	28
	4. Interactive 3D seedpoint validation	31
	C. Multi-scale approach	34

CHAPTER	Page
a. Principle of resolution	34
D. Automated tracing	36
E. Counting soma	37
IV RESULT AND ANALYSIS	41
A. F-measure	41
B. Automatic seedpoint selection	42
C. Tracing results by automatic seedpoint selection	43
D. Manual and automatic seedpoint selection comparison	46
E. Soma counting result	48
V DISCUSSION AND CONCLUSION	52
A. Contributions	52
B. Open issues and future work	52
1. Redundant seedpoints	53
2. Loss of seedpoints	53
3. Limitations of the interactive 3D validation method	54
4. Experimenting with the entire KESM data set	54
C. Conclusion	55
REFERENCES	56
VITA	61

LIST OF TABLES

TABLE	Page
I F-measure comparison result from the three different hippocampus regions of the KESM data sets ((a), (b), and (c)). Each corresponding result is shown in Fig. 29. Each data sample is applied to different seedpoint selection methodologies. In general, the recall value is higher than the precision value. F-measure is increased when the multi-scale approach was applied to the same sample.	45
II Comparison result of manual and automatic seedpoint selection . . .	48

LIST OF FIGURES

FIGURE		Page
1	The Knife-Edge Scanning Microscope (KESM) : The KESM is composed of 3D precision stage, granite bridge, knife/illuminator assembly, and linescan/microscope.	11
2	(a), (b) A specimen of mouse brain is cut by diamond knife as the illumination passes through the knife. Images are taken while the tissue is being sectioned.	11
3	Adjusting seedpoint to the center. (a) The procedure of adjusting seedpoint. (b) An example of screen shot of a real seedpoint adjustment. (<i>x</i> denotes a seedpoint selected by manual method and <i>o</i> shows an adjusted seedpoint).	12
4	Procedure for tracing. (a) Three local axes boundary detection. (b) Calculated local axis lengths. (c) Local volume estimation based on the medium axis length. (d) Local MIP result (the MIP of the longest local axis was ignored) and 2D tracing. (e) 3D vessel direction estimation and adjustment using the eigenvalue of the Hessian matrix.	12
5	The overall steps for experiment.	14
6	Coronal sections of Golgi-stained mouse cortex can be viewed from the side of the stack (resectioning) (a). Because Golgi-stain has sparse data, neuronal structures can be seen easily by stacking several slices (b). A single neuron can be seen by overlaying 300 slices (c).	16
7	Inverting image. (a) Cropped (128 × 128 pixels) KESM image, (b) Inverted image of (a). Background and foreground color was switched.	17

	FIGURE	Page
8	Median filtering principle. (a) Image with a noise in the center with pixel value (150), surrounded by correct neighboring values. (b) The median filtering result. The noise pixel value 150 is replaced with 124, so that the distinctive pixel value is smoothed to be similar to neighboring pixel values.	18
9	Median filtering result. (a) Before median filtering. (b) After median filtering. Median filtering removed unwanted noise and smoothed the original image to be useful for experiment.	19
10	Contrast enhancement. (a) An original image before contrast stretching, (b) The result of contrast stretching. The brightness of (b) is much stronger than (a) and objects can be detected easier than (a).	21
11	Stair step cutting.	22
12	Each column (1, 2, 3, 4, and 5) was stacked to construct a 3D volume data set.	23
13	Stacked 3D image from 2D slices.	24
14	Deciding threshold value. (a) A result of global threshold, (b) An adaptive local threshold result. The segments in (b) are observed to maintain better details than (a).	24
15	Principle of dilation. $A = \{(0, 1), (1, 1), (2, 1), (2, 2), (3, 0)\}$, $B = \{(0, 0), (0, 1)\}$, $A \oplus B = \{(0, 1), (1, 1), (2, 1), (2, 2), (3, 0), (0, 2), (1, 2), (2, 2), (2, 3), (3, 1)\}$	26
16	Dilation example. (a) The broken character, (b) Dilated by structure elements. Dilation makes characters more clearer than (a).	27
17	Principle of erosion. $A = \{(1, 0), (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (2, 1), (3, 1), (4, 1), (5, 1)\}$, $B = \{(0, 0), (0, 1)\}$, $A \ominus B = \{(1, 0), (1, 1), (1, 2), (1, 3), (1, 4)\}$	27

FIGURE	Page
18 Erosion example. (a) An original image. There are thin and long lines between surface object and inner rectangular object. By applying the erosion technique, the lines can be removed. (b) A result of erosion with a structure element (disk of radius 10). The thin lines were eliminated and other objects were preserved.	28
19 Morphological Object Detection. The figures show the result of applying the opening and closing image processing method. By opening, small protruding objects were removed; also by closing, small holes of the raw images were eliminated. (b) The result of opening. The thin protrusions and outward pointing boundary irregularities of (a) were removed. Also, the thin bridge and small isolated objects were eliminated. In the case of (c), the thin gulf, the inward-pointing boundary irregularities, and the small holes were removed. (d) The result of combining a closing and an opening for removing noise.	29
20 (a) An original KESM image. (b) A result of opening. Dendrites (cylinder shape) is eliminated. The sphere shapes are somata. (c) Illustrations of the result of soma detection applying opening. Somata (red dot) are detected , however, there are redundant segments due to the overlapping dendrites and some artifacts noise. (d) Combination opening and closing. The redundant segments were removed.	30
21 In 3D data set, in terms of one object, there are three different orthogonal (xy , yz , and zx) planes depending on perspective. In the KESM, the original soma is overlapped and shaded by other somata and dendrites. (a) A case of partially overlapped structure; upper neuron's dendrite shades on top of the lower neuron. (b) Two overlapped somata. (c) Overlapped dendrites that can be detected as a soma.	32
22 $f(x, y, z)$ is calculated by referencing the $s(x', y', z')$ and $r(x'', y'', z'')$. The shared coordinate is the z axis ($z = z' = z''$). From the shared value of the three equations, the x and y are extracted from the relationship.	33

FIGURE	Page
23 z is a shared coordinate between two plane coordinates. The z value was removed as a bridge to complement the missed (x, y) coordinate. The simple mathematical equation results in (x, y)	33
24 Interactive 3D seedpoint validation. SliceXY is for frontal, SliceXZ is for trasversal, and SliceZY is for coronal view. Each orthogonal view of the KESM data set is displayed with detected colorized seedpoint. The colors are marked on the center point of the segment. The blue cross bar indicates the individual seedpoint following mouse movement. The blue crossbar can be moved by clicking on the orthogonal views. In the xy orthogonal slice view, one segment below the crossbar was not detected as a seedpoint. By referencing the other slice views' detected seedpoint coordinates, the missed segment was validated as a seedpoint.	35
25 Samples of multi-scale. (a) An original image (128×128 pixel). (b) Resolution is decreased to fifty percent compared to original image. (c) Resolution is decreased 75 percentage compared to original image. (d) Resolution is decreased 88 percentage compared to original image. The edge and shape is not the same as original, also, the information of the pixel is missing along with the resolution decline.	36
26 Multi-scale approach. By changing the image resolution, somata were discriminated from densely overlapping dendrites. (a) The original image. (b) 40% resolution decreased from the original image. The unwanted overlapping dendrites were not detected by decreasing the image scale.	37
27 Soma counting. (a) A regional annotations of mouse brain, (b) The perpendicular to tangential directions for counting somata from surface (cortex) to deeper regions (hippocampus and colliculi).	39
28 Overview of part of the mouse brain. The shaded area is selected for counting.	40
29 Soma counting direction. Blue arrow represent the counting direction, which is perpendicular to the xy projection.	40

FIGURE	Page
30 Seedpoint selection result. (a), (b), and (c) Different regions of the KESM data set ($128 \times 128 \times 128$ each). Morphological object detection, interactive 3D validation method, and multi-scale approach were applied to select somata as seedpoints. Pink dots illustrate selected seedpoints. Blue segments represent the underlying neuronal structures.	44
31 Tracing result. (a) and (b) The tracing results. The green curves represent the tracing result. The red objects are the underlying data showing somata and dendrites.	47
32 (a), (b), and (c) Tracing result of manual and automatic seed-point selection in different regions. Manual seedpoint selection chose fifty seedpoints. Automatic selection 1 chose seedpoints between fifty and sixty, similar to the manual condition. Automatic selection 2 selected seedpoints as much as possible. The green curves are the tracing results based on the acquired seedpoints.	49
33 Somata counting result. (a) and (b) The number of somata from cortex to hippocampus, respectively. Regionally, distribution of somata is different.	50
34 Visualization using MeVisLab. This figure represents a portion of a visualization of soma distribution at a point in the middle of a counting direction (oriented from cortex to hippocampus). There are different densities of neuronal structures.	51

CHAPTER I

INTRODUCTION

A. Motivation

A neuron is a fundamental element of the neuronal system. The neuron transmits signals in response to input stimulus and sends them out to other neurons via neuronal connections [1]. The complex neuronal circuits give rise to functions such as perception and cognition. Alterations of the architecture and morphology also contribute to neurological diseases as well. In the case of Alzheimer's disease, there is a deficiency in the number of neurons and dendritic regressions in some neuronal structures [2]. Research about the neuronal structure can also help understand how the neuronal structure processes information. In order to investigate neuronal structures, researchers at the Brain Network Lab (BNL) at Texas A&M University (TAMU) developed techniques for obtaining microscopic data from whole mouse brains [3] and algorithms for tracing these data sets [4, 5].

The Knife-Edge Scanning Microscope (KESM) enables imaging of an entire mouse brain at sub-micrometer resolution to investigate the neuronal structure. Using the data sets, Han et al. could trace the neuronal and vascular structures of selective parts of the entire mouse brain based on manually selected seedpoints [4, 5]. However, to trace all fibers and neuronal structures in the brain, an automatic technique is required to set the seedpoints.

B. Goal of the research

The goal of this research is (1) to develop a method for automatic seedpoint selection, extending Han et al.'s previous work [4, 5], (2) to count the number of somata for

investigating regional differences and morphological features of the mouse brain, and (3) to trace the neuronal structures in the KESM data sets.

C. Approach

In order to investigate the neuronal structure, I followed six main steps: preprocessing, morphological object detection, interactive 3D seedpoint validation, multi-scale method, automatic seedpoint selection, soma counting, and tracing.

The preprocessing step after Golgi-stained data acquisition by KESM involved image inverting, noise filtering and contrast enhancement, and the merging of image columns and 3D image stacking. After obtaining the Golgi-stained data sets, the images were inverted to negative images to clearly separate somata and dendrites. Noise was removed by median filtering [6]. Additionally, after the noise filtering, I used contrast enhancement to make reconstruction easier. I stacked the merged images into 3D volumes.

After preprocessing the data sets, I proceeded to set seedpoints automatically. First, to implement automatic seedpoint selection, I applied an adaptive threshold method for discriminating the foreground and the background. Second, I used a morphological object detection (opening and closing) method, to remove small redundant objects and holes in the data sets [7]. Third, when validating the detected seedpoints, I employed an interactive 3D seedpoint validation method. This method displayed information from the volumetric KESM data sets by using three different orthogonal slices centered around the detected seedpoint. Fourth, I generated multi-scale representations from the raw images from the KESM. In the native resolution of the raw KESM data, it was difficult to obtain correct coordinates of a seedpoint due to the complex neuronal structures. By reducing the resolution of the raw data, I was

able to determine whether a portion of the complex structure was a seedpoint or not. Fifth, based on the morphological object detection method, I counted the somata from the superficial layer of the brain (cortex) through the deep regions (colliculi and hippocampus) perpendicular to the tangential direction to determine the quantity and distribution of neurons in different areas of the mouse brain. Finally, I used the obtained seedpoints to run Han et al.'s tracing algorithm [4, 5]. After neuronal structure tracing was complete, I investigated morphological features of the neurons in the mouse brain using image visualizing software: MeVisLab¹ and ParaView².

D. Significance

My automatic seedpoint selection method extended previously used manual seedpoint selection methods [4, 5]. By combining morphological object detection processing, interactive 3D seedpoint validation, and multi-scale method, I reduced the processing time and decreased the error in seedpoint selection. In addition, I automatically traced the neuronal structure of the mouse brain based on these seedpoints. This thesis is expected to contribute to a broader understanding of brain function based on the quantitative distribution of neurons.

E. Outline of the thesis

This thesis is organized as followings. In Chapter II, backgrounds and previous works related to automatic seedpoint selection and neuronal structure tracing will be explained. In Chapter III, I will present methods which were employed to implement the automatic seedpoint selection algorithm. Next, the experimental results of the

¹<http://www.mevislab.de>

²<http://www.paraview.org>

automatic seedpoint selection method and quantitative analyses will be presented in Chapter IV. Finally, in Chapter V, discussion about the experimental result, open issues, future work, and conclusion will be presented.

CHAPTER II

BACKGROUND AND PREVIOUS WORK

A. Background

In order to investigate neuronal structures, researchers developed methods for obtaining the data sets and algorithms for tracing and analysis. Particularly, somata are an important part of the neuronal structure because neurites are extended from the somata and the main functional process of the neuron occurs in them. The somata are the starting and ending points of the neuronal signal transmission, thus, they serve the core function [8]. Here, I will review existing data acquisition techniques and analysis approaches.

1. Automatic seedpoint selection

Han et al. experimented with manual and automatic seedpoint selection methods. First, they experimented with manual seedpoint selection and adjusted the selected point to correct the position [4]. However, as the number of fibers and their complexity grew, fully automatic method to set the seedpoint became necessary. For more efficiency, Han et al. used a semi-automatic seedpoint selection method [5]. With a global threshold value, the foreground and background were determined by the value. If the seedpoint was detected previously, the algorithm ignored the point. After the tracing has been initiated, the seedpoint was relocated by a momentum operator.

Can et al. [9] experimented with an automatic estimation method for setting the initial seedpoints and initial tracing directions based on grid analysis on retinal fundus images. An initial seedpoint was defined by setting $2N$ line searches: vertical and horizontal. The gray-level value on each line was low-pass filtered using a discrete

approximation to 1D Gaussian kernels. By employing a 1D sliding window, a local intensity minima was searched. Some detected local minima were noise and had to be removed to avoid spurious tracing. Additionally, a directional template provided an extensive method for eliminating noise and estimating an initial direction for tracing. With these mechanisms, seedpoint selection was performed effectively on straight segments of all the fibers and branching points.

Zhang et al. [8] developed a method to automatically select seedpoints in microscopic neuron images. The seedpoint selection method was fully automatic without human intervention. A seedpoint was selected on or near the center line of a dendritic or axonal segment. The seedpoint detecting algorithm was composed of two steps. First, by examining the local maxima on a set of grid lines of vertical and horizontal, the candidate seedpoint was automatically detected. The crossing pixel on the vertical and horizontal was low-pass filtered by employing 1D Gaussian kernels. After the filtering, candidate seedpoints were identified on the local intensity maxima. Second, signal-to-noise ratio analysis was used to remove a seedpoints that were falsely detected due to noise or artifacts. The result was good enough to extract seedpoints in poor quality, low resolution, and low contrast data.

Al-Kofahi et al. [10] traced neurons from 3D fluorescence confocal microscopy data. In order to select a seedpoint, they proceeded in two steps. In the first step, by employing maximum intensity projection, the 3D image was projected onto the xy plane. On the resulting 2D image, a grid of N horizontal lines and M vertical lines were superimposed. The gray-level of each line was low-pass filtered by the 1D Gaussian Kernel. Many of the seedpoint candidates had to be eliminated because they were generated due to noise. In the second step, the shift-and-correlate method was applied to filter the result of the first step. By using entire left and right templates lying on the xy plane, they validated the accuracy of the seedpoints. For selecting

the seedpoint on 3D, the z -coordinate was defined on the plane with local maximum intensity in the neighbourhood of detected 2D coordinates.

In summary, various methods have been developed for automatic seedpoint selection for tracing of neuronal structures. However, the methods may not be suitable for data containing multiple complex 3D structures, such as the KESM data sets. Also, the resolutions for showing morphological details of neurons or microvascular segments were not enough for tracing the entire span of the data.

2. Interactive 3D seedpoint validation

In order to ascertain the correctness of automatically selected seedpoints, interactive 3D seedpoint validation is needed. Interactive 3D seedpoint validation can ensure segmentation quality that simple visualizations in 2D can not.

Kang et al. [11] developed interactive 3D editing tools to correct or improve results of initial automatic segmentation. By employing an interactive 3D approach, they examined the segmentation result from computed tomography (CT) data sets of the proximal femur. They used three types of editing tools: hole-filling, point-bridging, and surface-dragging. The tools were evaluated by different level of noise on CT data sets. The tools corrected inadequate automatic segmentation results efficiently, as shown by quantitative results.

Varandas et al. [12] developed an interactive 3D ultrasonic system, which enabled acquisition and visualization of anatomy from different perspectives and measures of organ size and volume. The system helped reveal high contrast structures in freehand ultrasonic images. Also, multi-planar slice of the reconstructed grey level volume was useful for distinguishing genuine structure and was desirable in regions that were difficult to segment using thresholding. By applying VOLumetric Ultra-Sound (VOLUS), the 3D ultrasonic system could reduce the error rate caused by shade,

speckle, and reverberation in 2D images.

Westover [13] developed an interactive volume rendering algorithm to investigate an ambiguous volume data with enough flexibility. The algorithm was comprised of three parts: viewing transformation, signal reconstruction, and converting an input sample into a shaded intermediate sample. Since the renderer was interactive, users were able to specify application-specific mapping functions: processing speed of oil exploration, light wavelength data for galaxies, and stacks of computed tomography scans for medical imaging.

In sum, in order to investigate complex 3D structures, applying interactive 3D exploration has been found to be an effective technique.

3. Multi-scale approach

Multi-scale approach was inspired by the fine-to-coarse feature hierarchy in the human visual system [14].

Zhang et al. [14] experimented with a multi-scale technique using the object detecting strategy in humans. By using the fine-to-coarse detection strategy, many redundant objects could be excluded solely based on the low-resolution data, where they could detect features of objects with shorter computational time and minimal training time.

Murtagh et al. [15] used a multi-scale approach to detect globular cluster systems around elliptical galaxies. In order to investigate the environment, they applied pyramidal median image transform. Based on a 512×512 resolution image, other lower resolution images (50% decrease per each step) were generated. Multi-resolution transform provided several useful properties: multi-scaling of object resolution, automatically ignoring unwanted background object, and detecting specific objects. They concluded that multi-scale approaches improve the generality of the treatment and

the accuracy of the object detection results.

4. Soma counting

Soma count serves as an important metric for investigating brain function and disfunction, since the loss of neurons could cause physical and mental disorders [16].

D'Souza [17] developed an algorithm for counting the cell bodies in Nissl-stained cross sectional images. The algorithm consisted of five major steps: image acquisition from KESM, preprocessing, postprocessing, analysis and refinement, and visualization. D'Souza applied two different methodologies depending on the cell density, either sparse or high density. In the case of sparse population regions, he combined connected component labelling and template matching. In dense areas, he used the watershed algorithm. This research enabled segmentation, counting, and visualization of cells in the KESM Nissl data.

Rapp et al. [18] analyzed the relationship between the loss of hippocampal neurons and changing in mouse behaviour. By using the optical fractionator technique, they counted the entire number of neurons in the principal cell layers of the dentate gyrus and hippocampus. They demonstrated the connection between deficit learning and memory, and decrease in neuronal population.

Brizzee [2] elucidated the relationship between number of neurons and Alzheimer's disease (AD). In case of AD, there was serious decrease in the number of neurons and dendritic regression in the affected regions.

B. Prior work

The KESM can obtain data sets of whole mouse brains at a sub-micrometer resolution [3, 19, 20, 21]. Also, Han et al. applied a neuronal vascular structure tracing technique

to reconstruct the mouse brain vascular network [4, 5].

1. Imaging with the Knife-Edge Scanning Microscope

The KESM [3, 19, 20, 21] was invented by Bruce H. McCormick and improved by researchers at the Brain Network Laboratory (BNL) at Texas A&M University (TAMU) to scan entire mouse brains at a sub-micrometer resolution (down to 300 nm). This instrument consists of a high-precision stage, a diamond knife and an illuminator, a modified microscope, and a high-speed line-scan camera as shown in Fig. 1. A specimen of mouse brain (stained in Golgi, Nissl, or Indian ink) was mounted on top of a three-axis precision positioning stage, then cut by the diamond knife illuminated by white light from the rear of the diamond knife. The newly cut sections of the specimen were sampled by a high-sensitivity line-scan camera. The obtained digital images were transmitted to the two image acquisition boards and saved in a cluster computing system at the BNL [3].

2. Tracing vasculature of the KESM image data sets

In order to extract geometric descriptions from the raw data sets of the KESM, a Maximum Intensity Projection (MIP) based algorithm was developed for 3D vector tracing [4, 5]. First, manually selected seedpoints were required to start tracing. The manually selected seedpoints were not correctly centered in the vessels. So, by employing ray casting and momentum operator, the off-center seedpoints were adjusted to the center of the vessels as shown in Fig. 2.

Next, in order to determine the local volume size, boundaries were detected along the three axes (x , y , and z). The three local axis lengths were calculated by distance from the min and max-value along each axis ($x-$ to $x+$, $y-$ to $y+$, and $z-$ to $z+$) as shown in Fig. 3(a), (b). The obtained three local axis lengths were used to generate a

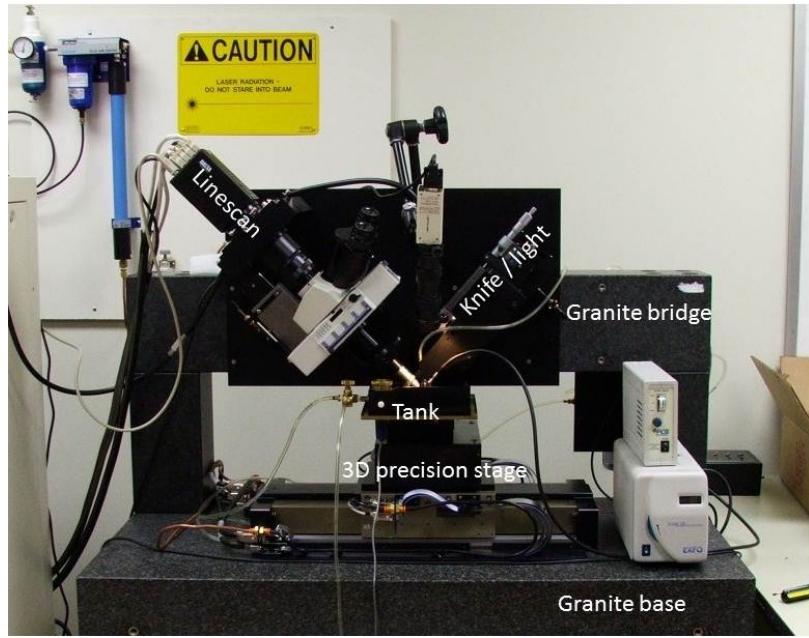


Fig. 1. The Knife-Edge Scanning Microscope (KESM) : The KESM is composed of 3D precision stage, granite bridge, knife/illuminator assembly, and linescan/microscope. Adapted from [3].

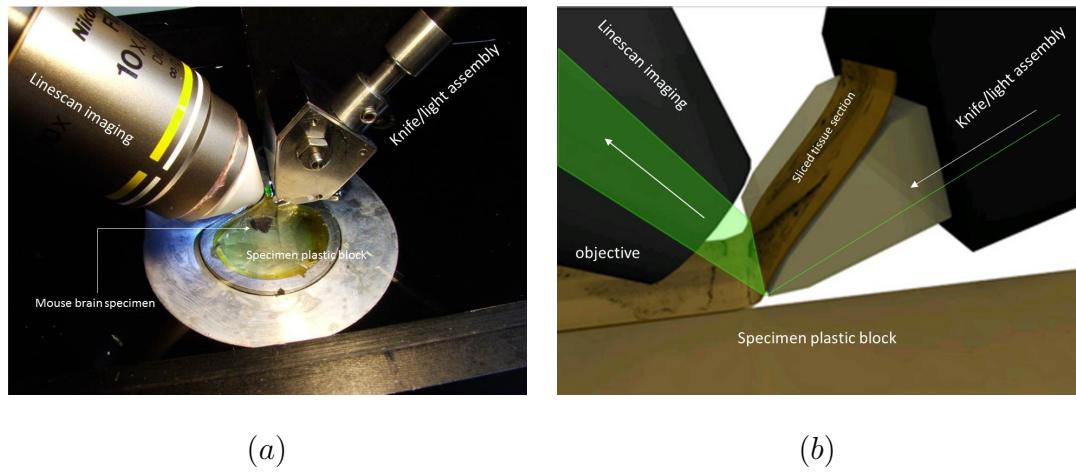


Fig. 2. (a), (b) A specimen of mouse brain is cut by diamond knife as the illumination passes through the knife. Images are taken while the tissue is being sectioned. Adapted from [3].

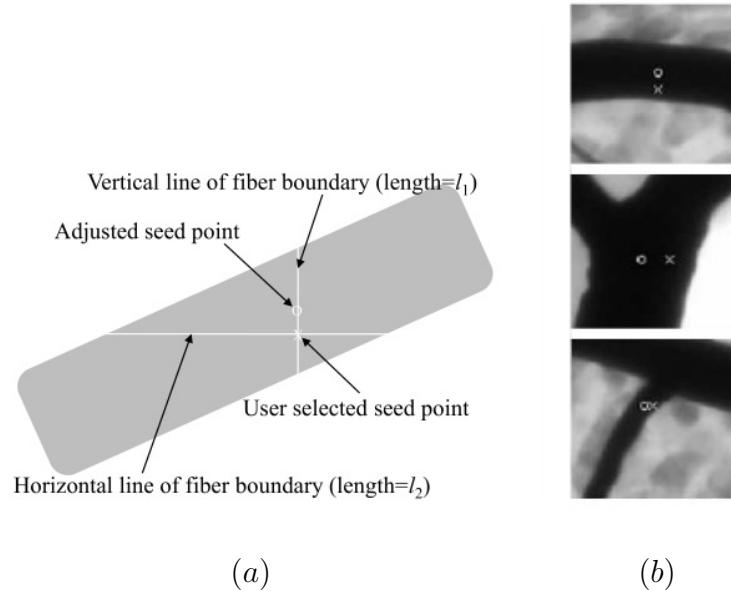


Fig. 3. Adjusting seedpoint to the center. (a) The procedure of adjusting seedpoint. (b) An example of screen shot of a real seedpoint adjustment. (x denotes a seedpoint selected by manual method and o shows an adjusted seedpoint). Adapted from [4].

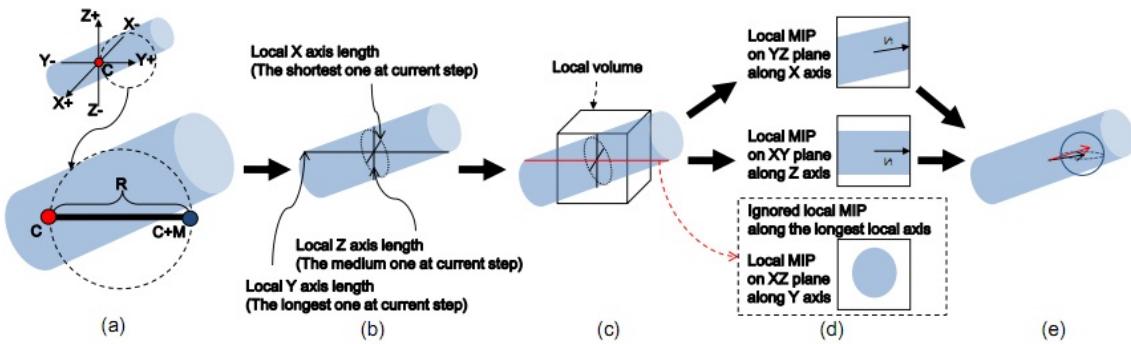


Fig. 4. Procedure for tracing. (a) Three local axes boundary detection. (b) Calculated local axis lengths. (c) Local volume estimation based on the medium axis length. (d) Local MIP result (the MIP of the longest local axis was ignored) and 2D tracing. (e) 3D vessel direction estimation and adjustment using the eigenvalue of the Hessian matrix. Adapted from [5].

fitting 3D volume. Based on the three local MIPs of this volume, fiber direction was estimated. The local MIP along the longest local axis was ignored due to the lack of information about the fiber direction. Therefore, only two local MIPs were used to determine fiber direction. A multi-scale filter applying the eigenvalue of the Hessian matrix was used to estimate fiber direction based on the two local MIPs [22]. The tracing was stopped when two conditions were met: maximum intensity of the next segment was less than a predefined threshold and the next candidate center point located on the traced axes.

CHAPTER III

METHODS

In this chapter, I elaborate on the methodologies for the research. Broadly, there are six main steps for conducting the experiment: preprocessing, morphological object detection, interactive 3D seedpoint validation, multi-scale method, automatic seed-point selection, somata counting, and tracing as shown in Fig. 4. Each procedure is described in detail in the following sections.

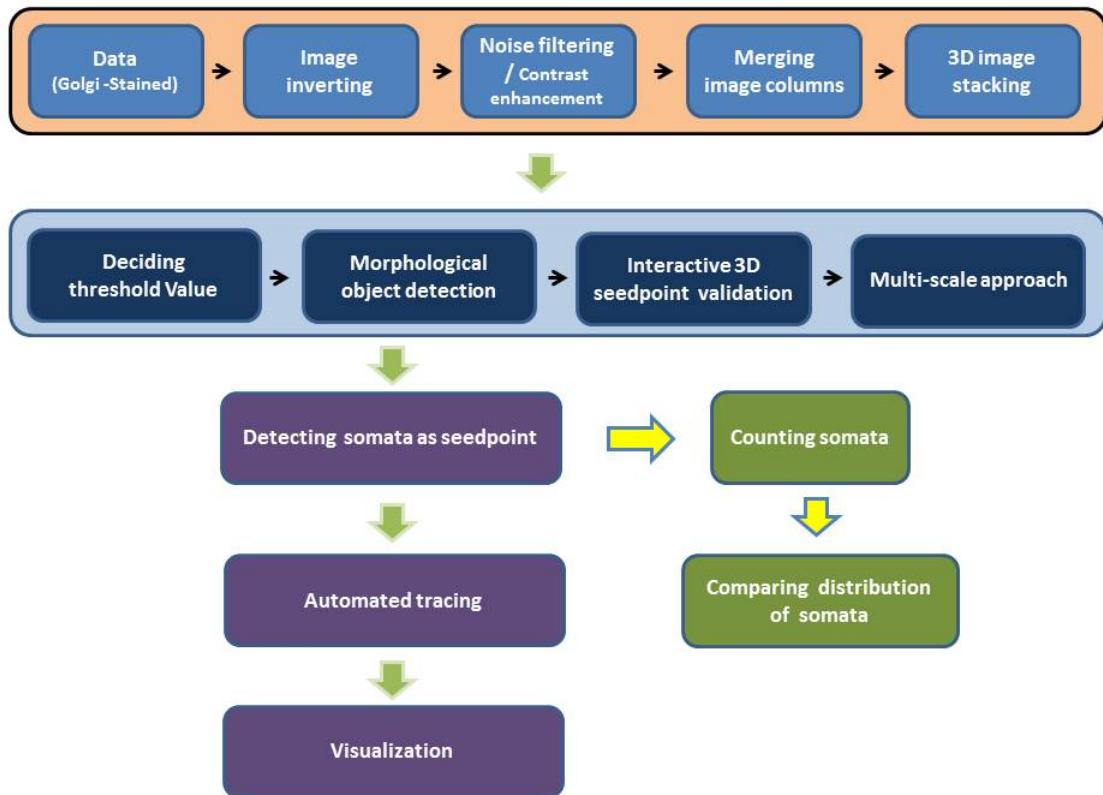


Fig. 5. The overall steps for experiment.

A. Preprocessing

In this section, I describe preprocessing steps for the experiment. Initially, the obtained KESM data sets have various sources noise, which need to be removed. Without preprocessing, seedpoint selection algorithms could select a noise as a seedpoint. In order to scale the method from small regional samples to whole of a mouse, preprocessing is an essential step. The preprocessing stage consists of five steps: data acquisition, image inverting, noise filtering and contrast enhancement, merging image columns, and 3D image stacking.

1. Data acquisition

The Golgi-stained data obtained from the KESM was from a whole mouse brain (2TB, 60,000+ images). The Golgi method stains somata, dendrites, and axons of neurons, so the full neuronal structure could be investigated [23]. Each image had the dimension of $600 \times 3,000$ pixels with size of ~ 317 KB (compressed). These images were cropped down to 128×128 pixels to insure quick rendering. Fig. 5 shows an example of the Golgi-stained coronal section of a mouse cortex.

2. Inverting image

The Golgi-stained images reveal the whole structure of the neurons while only staining about one percent of the neurons in the tissue [24]. In the raw image data sets, somata and dendrites were not clearly separated due to other random cells and noise. Those raw images were first inverted to negative images. The inverting function of Matlab¹ was used to change the raw images. Algorithm 1 explains the step of image inverting. Fig. 6 shows an original image (a) and its inverted version (b).

¹<http://www.mathworks.com>

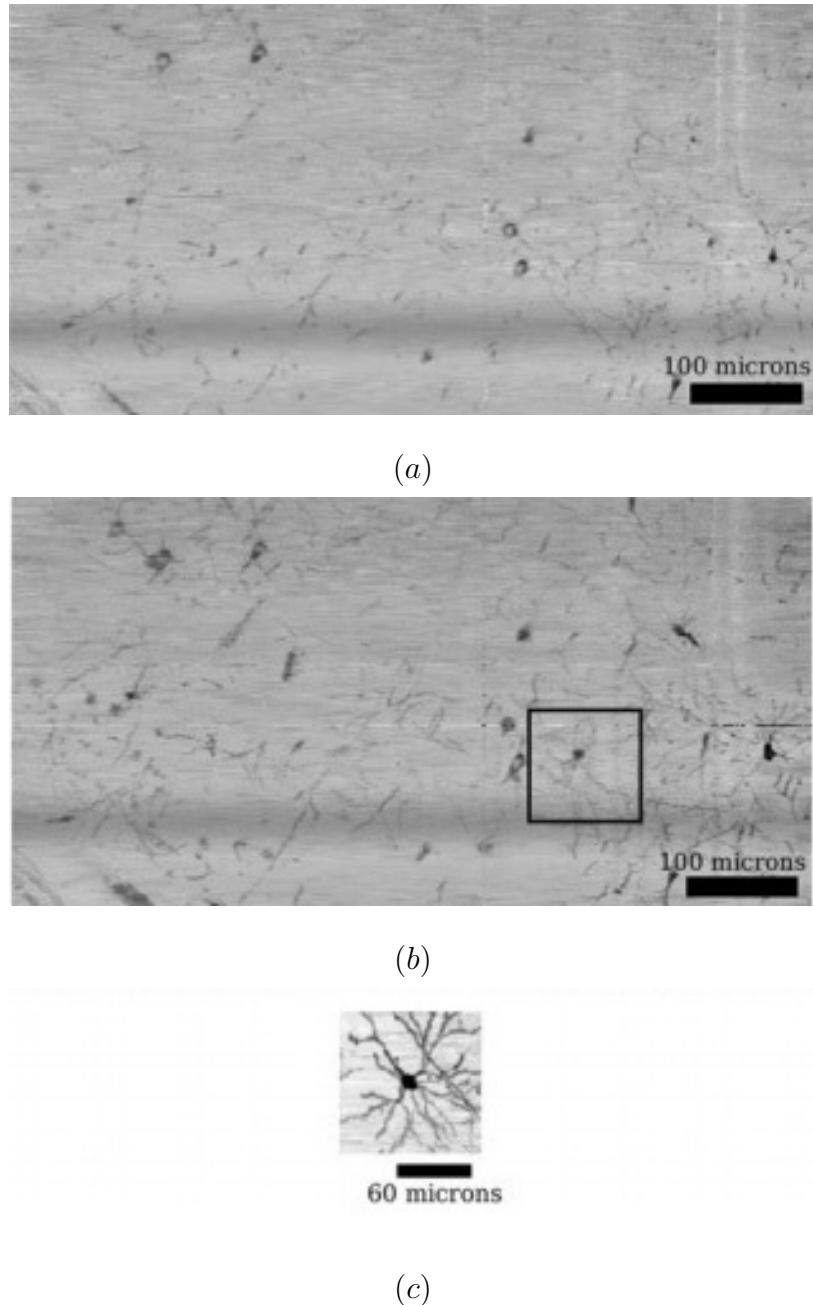


Fig. 6. Coronal sections of Golgi-stained mouse cortex can be viewed from the side of the stack (resectioning) (a). Because Golgi-stain has sparse data, neuronal structures can be seen easily by stacking several slices (b). A single neuron can be seen by overlaying 300 slices (c). Adapted from [3].

- 1: convert the image matrix to double data type
- 2: find the max value in this new array and take its inverse
- 3: Multiply the image matrix by the factor of negImageScale and subtract the total from 1

Algorithm 1: Inverting algorithm.

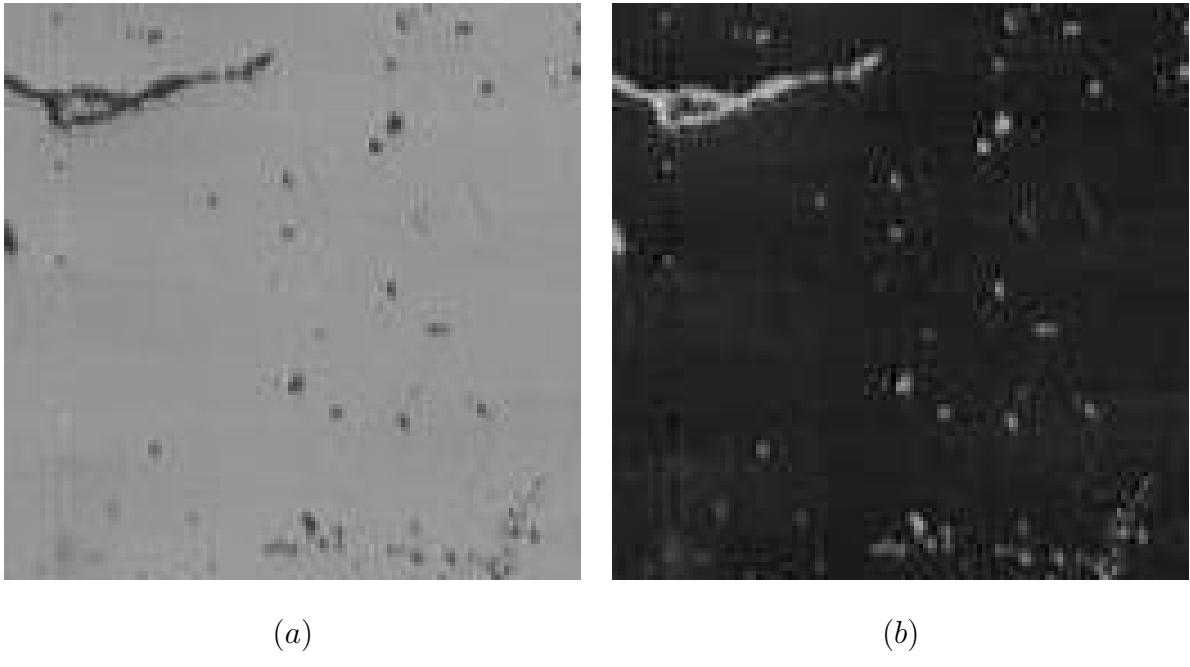


Fig. 7. Inverting image. (a) Cropped (128×128 pixels) KESM image, (b) Inverted image of (a). Background and foreground color was switched.

3. Noise filtering

The raw KESM images contained background noise caused by the sectioning process and certain lighting artifacts [3]. The noise must be eliminated from the images for the data sets to be useful. Noise was eliminated by median filtering [6]. Median filtering not only removes extreme values, but also preserves other pixel values to retain its original image.

Figure 8 consists of two 5x5 grids labeled (a) and (b). Grid (a) contains the following values:

123	125	126	130	140
122	124	126	127	135
118	120	150	125	134
119	115	119	123	133
111	116	110	120	130

Grid (b) contains the following values:

123	125	126	130	140
122	124	126	127	135
118	120	124	125	134
119	115	119	123	133
111	116	110	120	130

Fig. 8. Median filtering principle. (a) Image with a noise in the center with pixel value (150), surrounded by correct neighboring values. (b) The median filtering result. The noise pixel value 150 is replaced with 124, so that the distinctive pixel value is smoothed to be similar to neighboring pixel values. Adapted from [6].

a. Median filtering

Fig. 7(b) displays a result of median filtering. All surrounding pixel values of a target pixel to be denoised are reordered as an ascending sequence. By calculating the median pixel value among the ascending sequence, the target pixel value is adjusted. Equation from (3.1) to (3.3) explains the median filtering mathematically. The pixel values are stored in an $N[0 \dots n-1]$ array in an ascending manner.

$$R_{x,y} = N\lfloor n/2 \rfloor \quad (3.1)$$

where,

$R_{x,y}$: median pixel value.

For example, in Fig. 7(a), there are $5 \times 5 - 1$ neighbors. Suppose the center pixel value 150 is regarded as noise and needs to be smoothed out using these surrounding pixels.

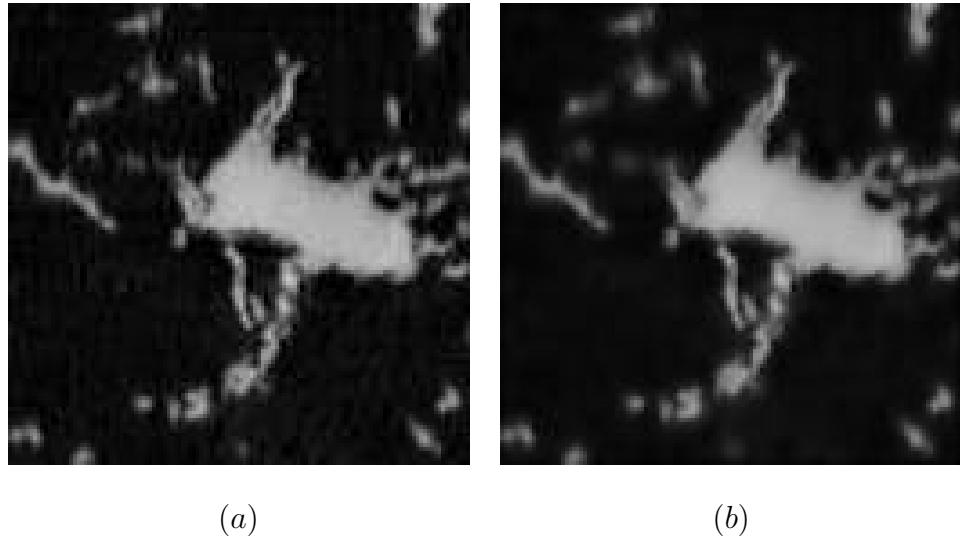


Fig. 9. Median filtering result. (a) Before median filtering. (b) After median filtering. Median filtering removed unwanted noise and smoothed the original image to be useful for experiment.

By ascending sorting, the sequence of immediate neighbor pixel values is reordered as

$$N[0\dots8] = \{115, 119, 120, 123, 124, 125, 126, 127, 150\} \quad (3.2)$$

The median is 124, so the noise pixel value is changed to

$$R_{x,y} = N\lfloor 9/2 \rfloor = N[4] = 124 \quad (3.3)$$

Fig. 8 illustrates a median filtering result. Ambiguous background noise was eliminated and overall image was smoothed so that seedpoint can be clearly identified.

4. Contrast enhancement

Additionally, with the median filtered image, I applied contrast enhancement. Contrast is the difference in brightness of region compared to neighboring regions [17].

The KESM image had various brightness and gray-scale range due to uneven illumination during imaging. Such irregularities could hide and obscure an object that is to be detected as a seedpoint. Contrast enhancement technique provided an important role in image processing to clear the noise. In order to enhance the contrast, a contrast stretching technique was applied [17]. Based on the original contrast and brightness of the image, a linear mapping function increased the contrast and brightness level of the image to clarify the original image [25].

$$P_k = \frac{(max - min)}{f_{max} - f_{min}} (q_k - f_{min}) + min \quad (3.4)$$

where,

P_k : contrast enhanced pixel value

q_k : raw input pixel value

f_{max} : maximum pixel value of the raw input image

f_{min} : minimum pixel value of the raw input image

$max & min$: target maximum and minimum pixel value of the contrast enhanced image.

5. Merging image columns and 3D image stacking

In order to make 3D volume data sets, image merging was needed. The whole mouse brain slices were cut as eight deep columns using a stair-stepping method shown in Fig. 10 to reduce knife chatter and increase the quality of the images [24]. By using the obtained 3D data sets, whole neuronal structures in the mouse brain could be traced. Fig. 11 illustrates how image stacking is done to obtain the 3D data sets. Whenever a slice of a specimen was cut and scanned, the whole image was saved [24]. Part of the images were cropped to render small 3D volumes. For efficiency of

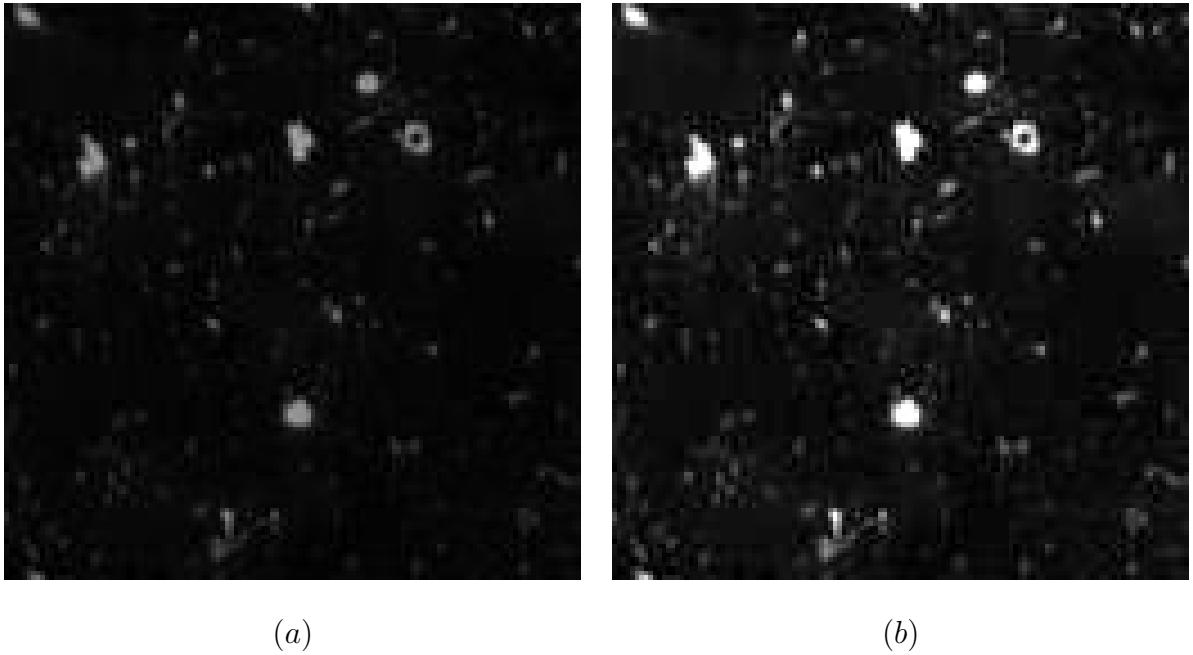


Fig. 10. Contrast enhancement. (a) An original image before contrast stretching, (b) The result of contrast stretching. The brightness of (b) is much stronger than (a) and objects can be detected easier than (a).

computation, I chose $128 \times 128 \times 128$ volume size as shown in Fig. 12. Also, I applied alpha channel to all of the images to observe soma and dendrite so that seedpoints can be easily selected [7].

B. Automatic seedpoint selection

Based on the preprocessed KESM data sets, image processing techniques were employed to implement automatic seedpoint selection. The KESM data included objects of multiple shapes and density depending on the region. So, in order to investigate the neuronal structure within the KESM data, several image processing techniques were required: adaptive thresholding, morphological object detection, interactive 3D seedpoint validation, and multi-scale approach.

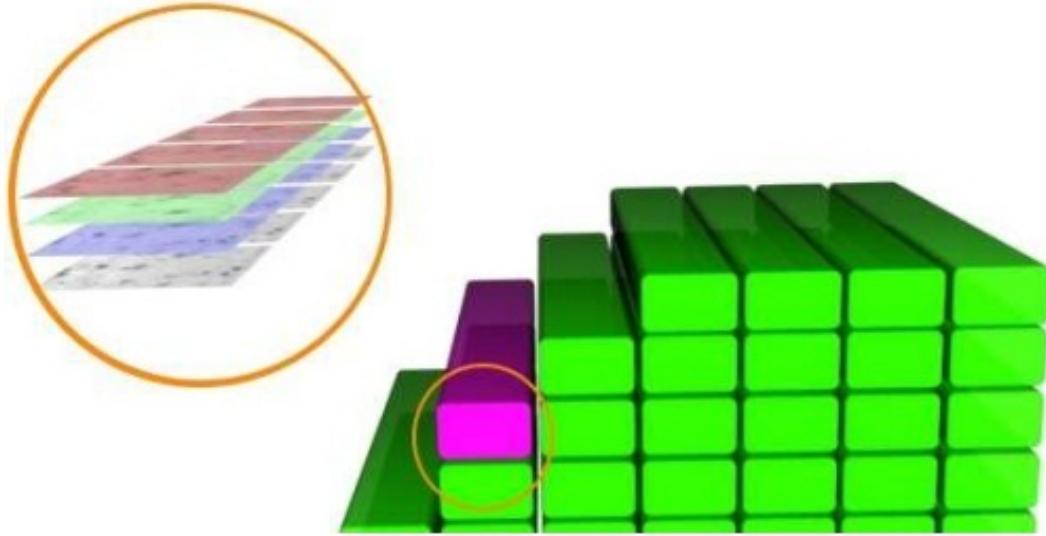


Fig. 11. Stair step cutting. Adapted from [24].

1. Deciding the threshold value

To implement an automatic seedpoint selection method, an optimal threshold was required to segment the foreground and the background based on the intensity value of the pixels. The acquired images have various structures and complex shapes. Using only one global threshold for all pixels was not suitable for segmentation. I applied an adaptive threshold method for discriminating the foreground and the background [26]. Adaptive thresholding changes the threshold value dynamically over the complex images. Based on the range of intensity values in its local neighbourhood, the adaptive thresholding method selects an individual threshold for each pixel. Fig. 13 illustrates different results of image segmentation with global and adaptive local threshold.

2. Morphological object detection (opening and closing)

I employed a morphological object detection approach in order to define seedpoints. Most of the objects obtained from the KESM locally resemble spheres (somata) and

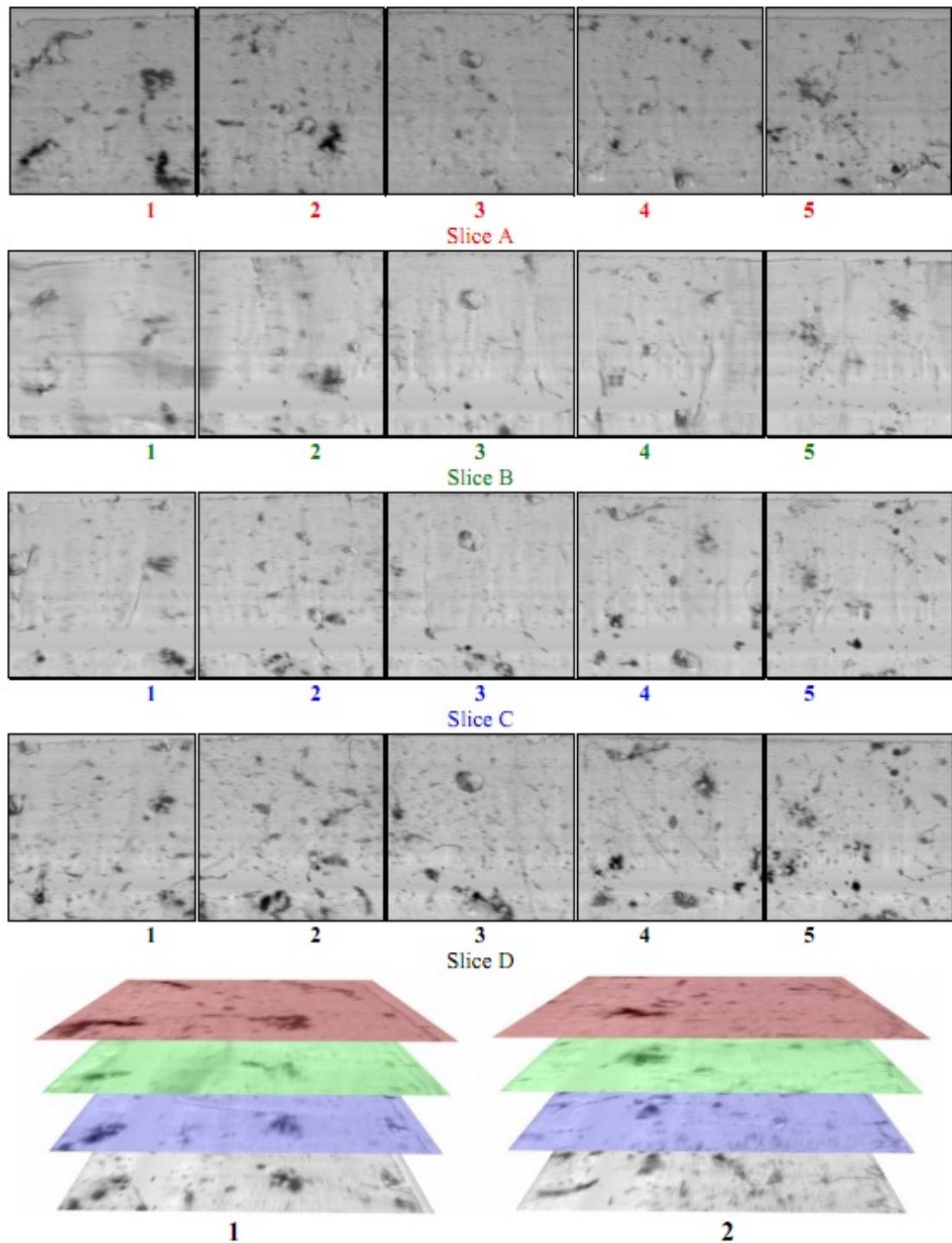


Fig. 12. Each column (1, 2, 3, 4, and 5) was stacked to construct a 3D volume data set. Adapted from [24].

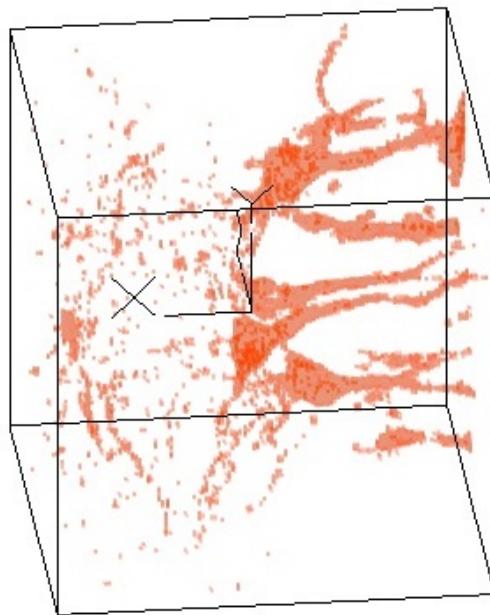


Fig. 13. Stacked 3D image from 2D slices.

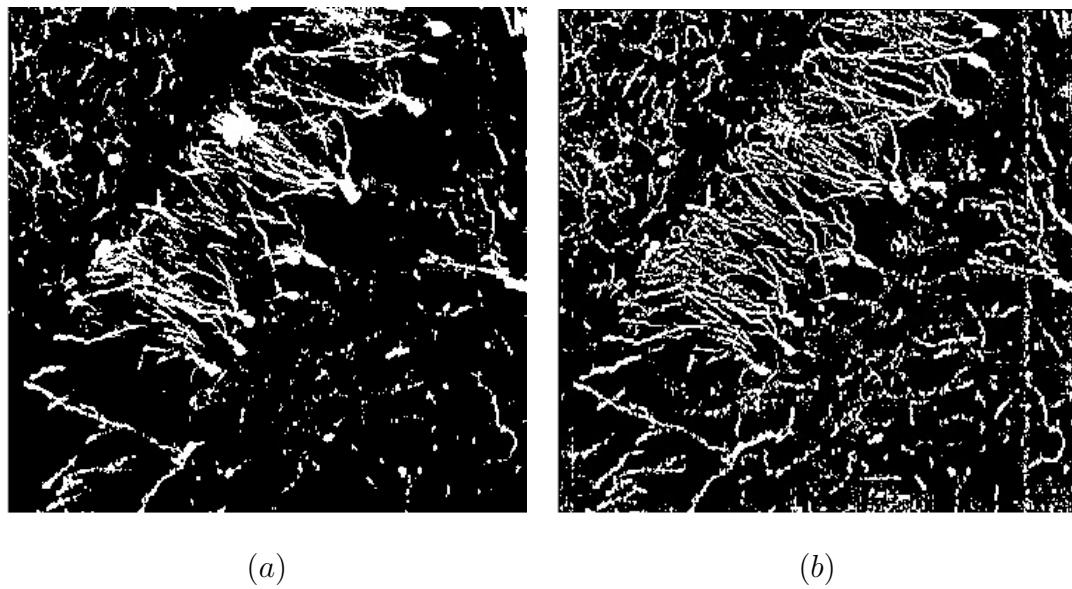


Fig. 14. Deciding threshold value. (a) A result of global threshold, (b) An adaptive local threshold result. The segments in (b) are observed to maintain better details than (a).

cylinders (dendrites) connected to form complex structures. To extract useful geometric components of the KESM data sets (somata and dendrites), morphological object detection processing was needed to preserve shapes and remove irrelevant components in the images.

By applying the opening and closing method [7], I identified specific objects as somata. Fig. 18 illustrates the morphological object detection steps. Fundamentally, morphological image processing (opening and closing) is implemented by dilation and erosion to remove or to add a segment.

a. Dilation

Dilation transforms a binary image to grow or become thicker by adding pixels to the boundaries of the existing foreground object. Depending on the structure element, additional number of pixels are determined. The structure element is a basic shape for adding and eliminating pixel shapes in a given image. By employing the structure element, shapes are extracted from the image depending on the structure element's morphological feature [7]. If A is dilated by B (structure element), the resulting set is defined as follows [27]

$$A \oplus B = \{c \in E^N \mid c = a + b \text{ for some } a \in A \text{ and } b \in B\} \quad (3.5)$$

where,

E^N : Euclidean N-space

$a = (a_1, \dots, a_N)$, elements of A

$b = (b_1, \dots, b_N)$, elements of B

All possible combinations of elements of the sets A and B generate the dilation of A by B [27]. Fig. 14 illustrates the principle of dilation and Fig. 15 shows an example

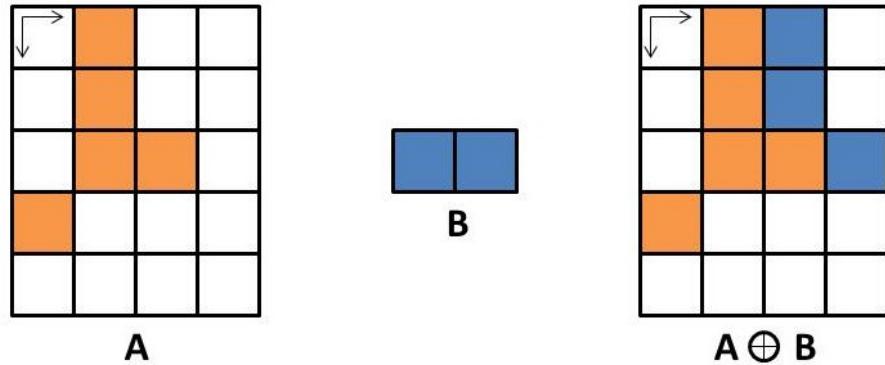


Fig. 15. Principle of dilation.

$$\begin{aligned} A &= \{(0,1), (1,1), (2,1), (2,2), (3,0)\}, B = \{(0,0), (0,1)\}, \\ A \oplus B &= \{(0,1), (1,1), (2,1), (2,2), (3,0), (0,2), (1,2), (2,2), (2,3), (3,1)\}. \end{aligned}$$

of dilation result.

b. Erosion

Erosion transforms a binary image to shrink or become thinner. Similar to dilation, the structure elements control how the original image should be reduced. Fig. 17 shows an example of erosion. If A is eroded by B (structure element), the resulting set is defined as follows [27]

$$A \ominus B = \{x \in E^N \mid x + b \in A \text{ for every } b \in B\} \quad (3.6)$$

where,

E^N : Euclidean N-space

$a = (a_1, \dots, a_N)$, elements of A

$b = (b_1, \dots, b_N)$, elements of B

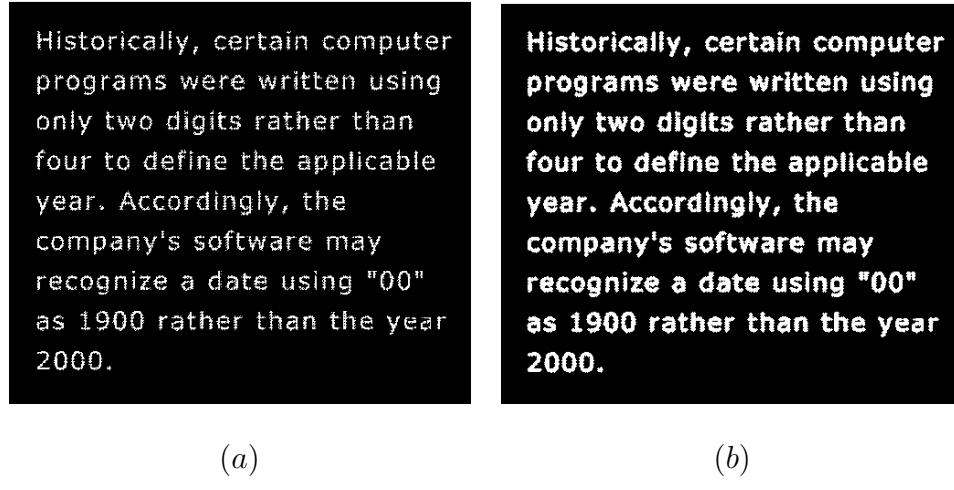


Fig. 16. Dilation example. (a) The broken character, (b) Dilated by structure elements. Dilation makes characters more clearer than (a). Adapted from [7].

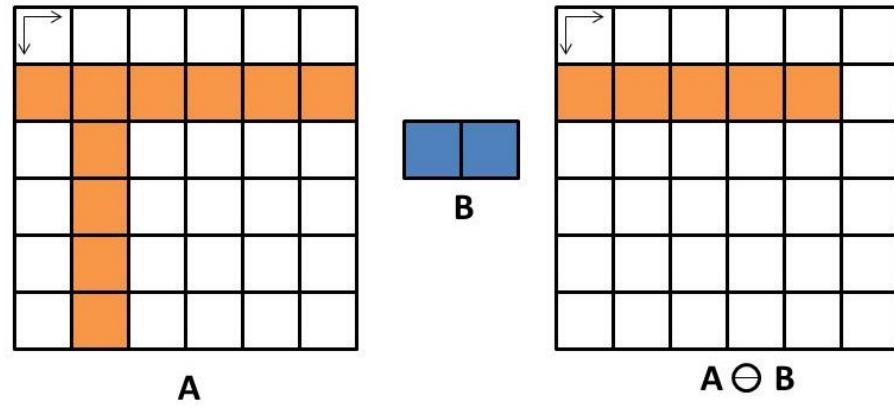


Fig. 17. Principle of erosion.

$$\begin{aligned}
 A &= \{(1,0), (1,1), (1,2), (1,3), (1,4), (1,5), (2,1), (3,1), (4,1), (5,1)\}, \\
 B &= \{(0,0), (0,1)\}, A \ominus B = \{(1,0), (1,1), (1,2), (1,3), (1,4)\}.
 \end{aligned}$$

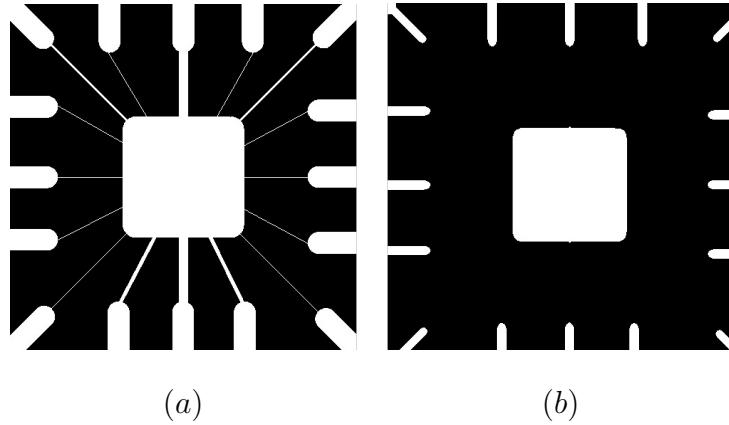


Fig. 18. Erosion example. (a) An original image. There are thin and long lines between surface object and inner rectangular object. By applying the erosion technique, the lines can be removed. (b) A result of erosion with a structure element (disk of radius 10). The thin lines were eliminated and other objects were preserved. Adapted from [7].

3. Opening and closing

Dilation and erosion are combined to identify a specific component without image distortion. Opening removes sharp peaks and thin connections with various structure elements (diamond, disk, rectangle, line, and square) to smooth out the contour. Closing links narrow breaks, fills long thin gulfs, and fills holes smaller than the structure elements. The morphological opening of A by a structure element K is defined as

$$A \circ K = (A \ominus K) \oplus K \quad (3.7)$$

Morphological closing of A by structure element K is defined as

$$A \bullet K = (A \oplus K) \ominus K. \quad (3.8)$$

Fig. 18 shows an example of morphological object detection procedure. Based on the morphological opening and closing method, seedpoints were selected on the

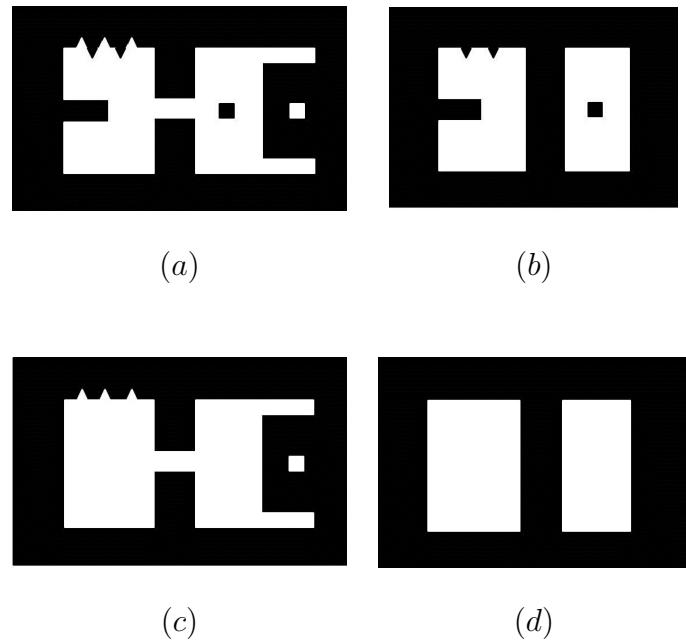


Fig. 19. Morphological Object Detection. The figures show the result of applying the opening and closing image processing method. By opening, small protruding objects were removed; also by closing, small holes of the raw images were eliminated. (b) The result of opening. The thin protrusions and outward pointing boundary irregularities of (a) were removed. Also, the thin bridge and small isolated objects were eliminated. In the case of (c), the thin gulf, the inward-pointing boundary irregularities, and the small holes were removed. (d) The result of combining a closing and an opening for removing noise. Adapted from [7]

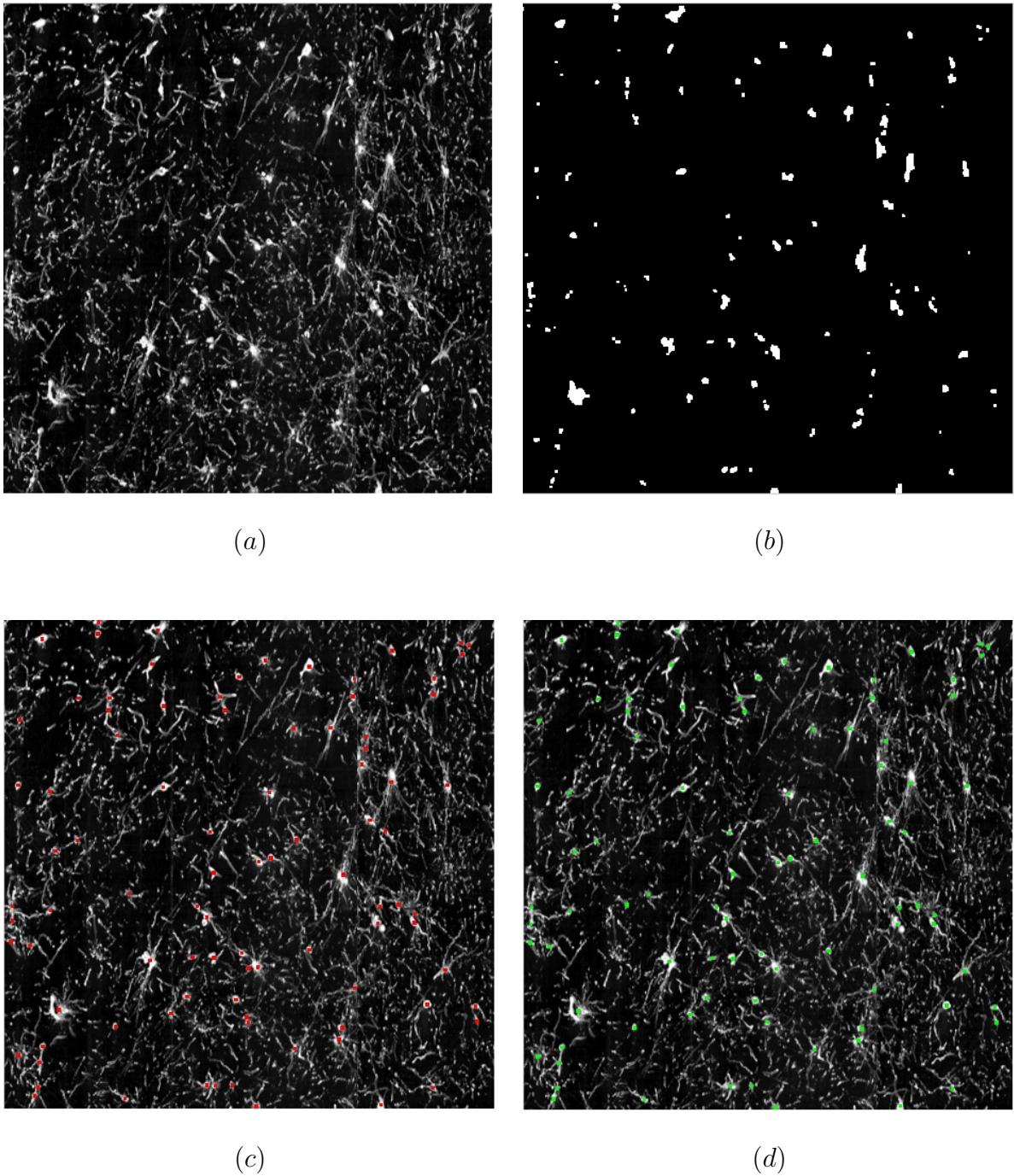


Fig. 20. (a) An original KESM image. (b) A result of opening. Dendrites (cylinder shape) is eliminated. The sphere shapes are somata. (c) Illustrations of the result of soma detection applying opening. Somata (red dot) are detected , however, there are redundant segments due to the overlapping dendrites and some artifacts noise. (d) Combination opening and closing. The redundant segments were removed.

KESM data sets to trace. As can be seen in Fig. 19, the KESM data contains complex structure. By employing morphological image processing, somata were reliably selected as seedpoints. Because opening and closing method cuts dendrite segment and adds more pixels around somata to be selected as seedpoints. In case of regions with low density of structures, the technique extracted seedpoints correctly. However, as the complexity grew, there were overlapping objects that were incorrectly marked as soma as shown in Fig. 19(c). For example, when the dendrites are overlapping in stacked slices, the region had enough clustered pixels to be confused as soma. So, a validation procedures for morphological object detection approach was required. In the next section, I will explain the interactive 3D seedpoint validation method.

4. Interactive 3D seedpoint validation

With the morphological object detection method, there was limitation about validating whether the detected object was soma or not. The KESM image structure was not easy to validate with single layer images. In order to validate the detected seedpoints, I employed an interactive 3D seedpoint validation method. This method extracted information from the volumetric KESM data sets by using three different orthogonal slices corresponding to the detected seedpoint, manipulating, and rendering with enough flexibility to validate the detected objects.

In order to validate a detected seedpoint in different planes, this method analyzes not only each 2D slice of a specific region, but also a 3D rendering of the spatial information of the area related to the 2D image. In short, interactive 3D seedpoint validation could ensure segmentation quality that simple visualizations in 2D can not. Han et al. [4] set the seedpoints manually based on xy plane images. However, his work was labor-intensive, since it had to set all initial points in a complex and massive structure. This was because the morphology of neuronal objects has different shapes

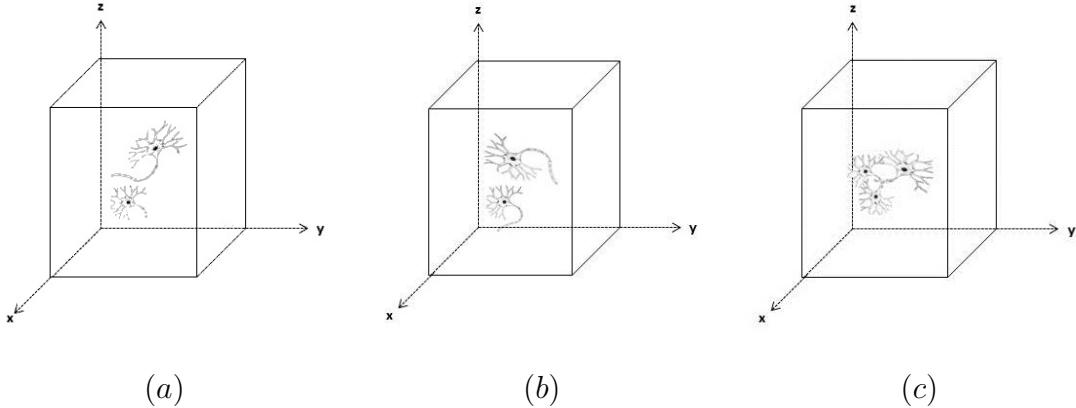


Fig. 21. In 3D data set, in terms of one object, there are three different orthogonal (xy , yz , and zx) planes depending on perspective. In the KESM, the original soma is overlapped and shaded by other somata and dendrites. (a) A case of partially overlapped structure; upper neuron's dendrite shades on top of the lower neuron. (b) Two overlapped somata. (c) Overlapped dendrites that can be detected as a soma.

depending on the perspective and transparency as shown in Fig. 20.

The proposed interactive viewer shows three different orthogonal slices (xy , yz , and zx) of the data sets along with the seedpoints from three different views. If one of the objects on the xy view was not selected, it can be corrected by referencing other slices' detected coordinates.

Fig. 21 illustrates the concept of the interactive 3D seedpoint validation approach. Fundamentally, the raw KESM image slices were stacked along the z axis direction (bottom to top). Thus, a voxel can have three orthogonal views (xy , yz , and zx). In order to trace the neuronal structure, Han et al.[4] set seedpoints on xy plane and the tracing was initiated on that plane.

In Fig. 21, the green dot $s(x', y', z')$ represents a voxel coordinate on the xz view and the blue dot $r(x'', y'', z'')$ denotes a voxel of the zy view. The red dot, $f(x, y, z)$, is required to be validated as soma. If one of the seedpoints was not detected on the

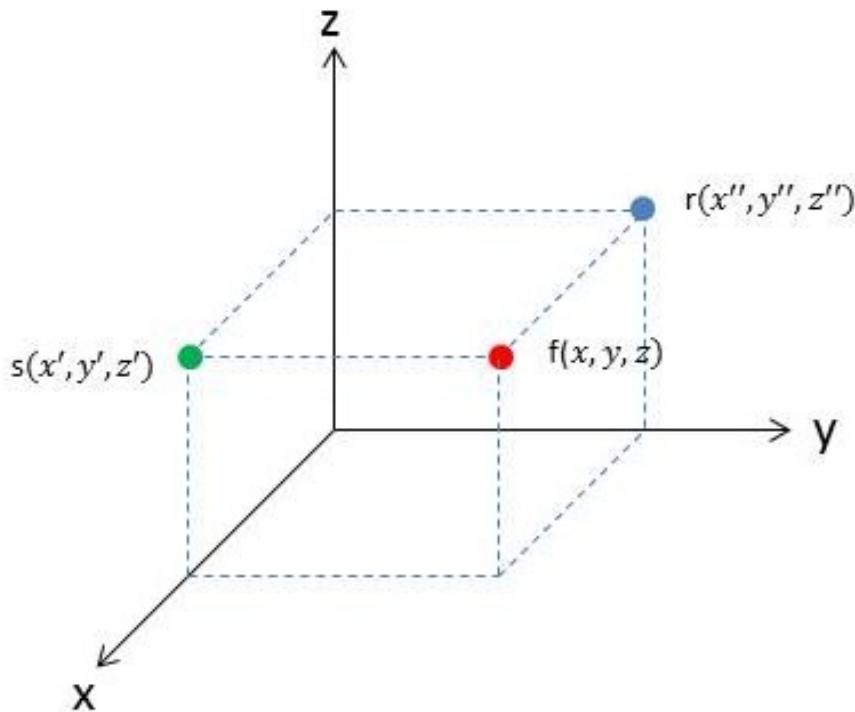


Fig. 22. $f(x, y, z)$ is calculated by referencing the $s(x', y', z')$ and $r(x'', y'', z'')$. The shared coordinate is the z axis ($z = z' = z''$). From the shared value of the three equations, the x and y are extracted from the relationship.

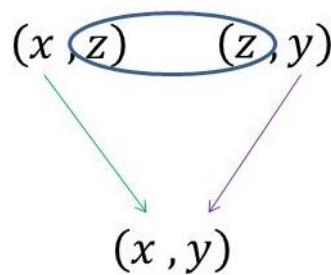


Fig. 23. z is a shared coordinate between two plane coordinates. The z value was removed as a bridge to complement the missed (x, y) coordinate. The simple mathematical equation results in (x, y) .

xy orthogonal view, while it was detected on the other slices, the missed coordinate of the object could be calculated by referencing the intersection coordinate of (x, y) of xz and zy planes as shown in Fig. 22. In the example, the z , z' , and z'' have the same value. Fig. 22 shows an simple relationship between two coordinates which include same z axis coordinate. Fig. 23 shows an example of validating ambiguous seedpoints using the interactive 3D seedpoint validation.

C. Multi-scale approach

a. Principle of resolution

Scale (resolution) means the level of detail in image information. Resolution is defined as the number of pixels per inch. High resolution image contains more details in each pixel, so it provides detailed information about the image. Pixels are the basic element for digital image. When the total number of pixels in the image is fixed, the increase of resolution results in the change of the physical dimension of the image and vice versa. Multi-scale methods enable the analysis of image with different scales. Fig. 24 displays an example of multi-scale image and the loss of information as the resolution is decreased. The purpose of using multi-scale approach was to discriminate multiple overlapping segments from true seedpoints.

In the previous section, the morphological object detection technique provided a method to define somata as seedpoints and interactive 3D seedpoint validation method validated the detected somata. However, in the native resolution of the raw image, it was difficult to obtain correct coordinates of a seedpoint due to the complex neuronal structures in the KESM data. In case of a high resolution image, some parts of the detailed object parts could be falsely identified as a seedpoint. On the other hand, in low resolution images, legitimate seedpoints could be missed. By changing

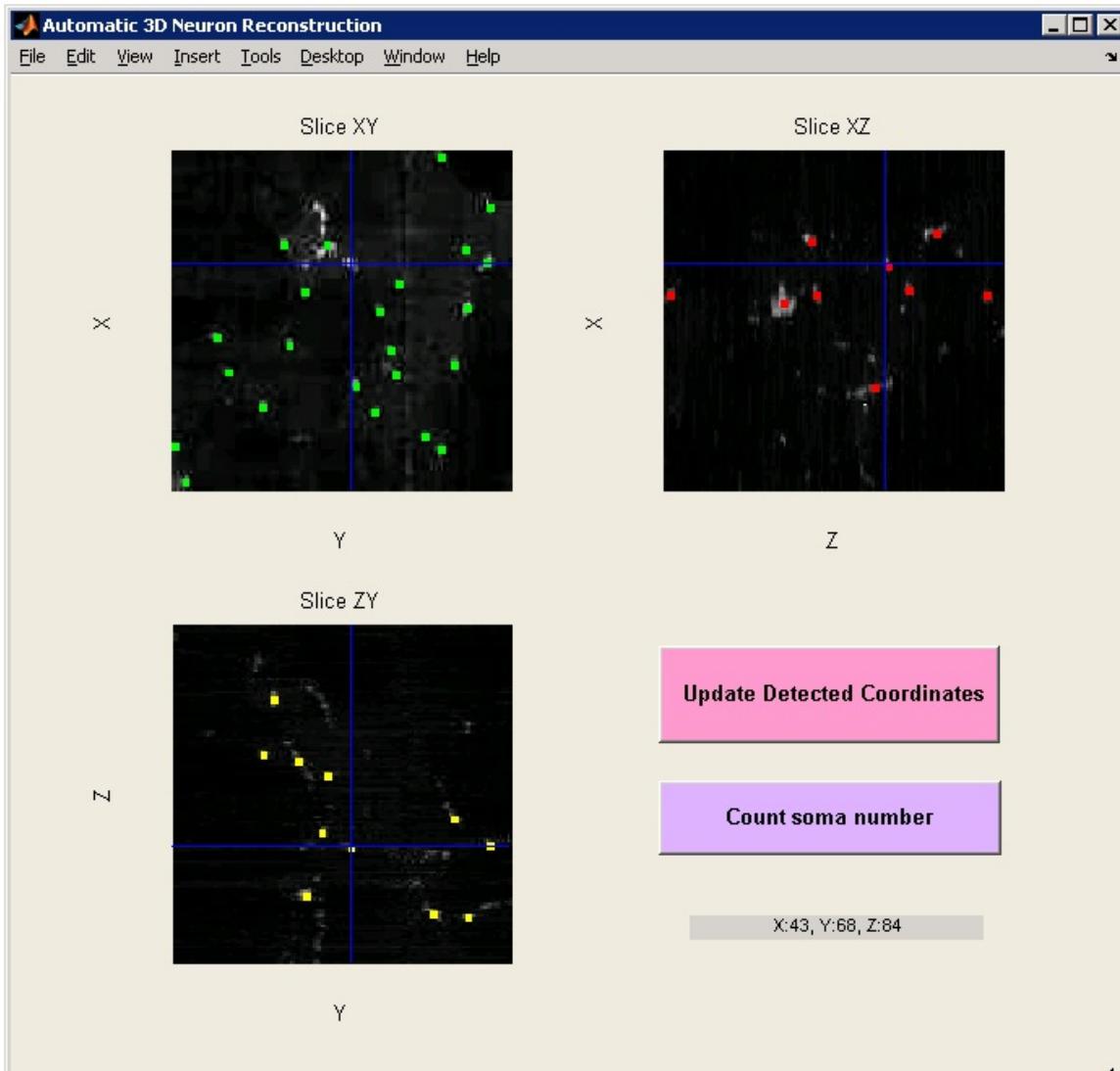


Fig. 24. Interactive 3D seedpoint validation. SliceXY is for frontal, SliceXZ is for trasversal, and SliceZY is for coronal view. Each orthogonal view of the KESM data set is displayed with detected colorized seedpoint. The colors are marked on the center point of the segment. The blue cross bar indicates the individual seedpoint following mouse movement. The blue crossbar can be moved by clicking on the orthogonal views. In the *xy* orthogonal slice view, one segment below the crossbar was not detected as a seedpoint. By referencing the other slice views' detected seedpoint coordinates, the missed segment was validated as a seedpoint.

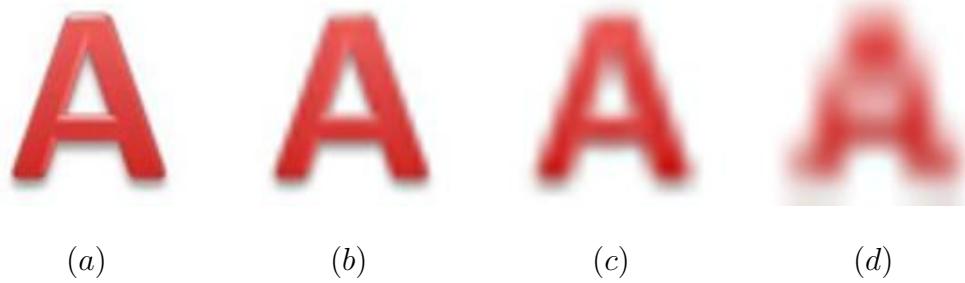


Fig. 25. Samples of multi-scale. (a) An original image (128×128 pixel). (b) Resolution is decreased to fifty percent compared to original image. (c) Resolution is decreased 75 percentage compared to original image. (d) Resolution is decreased 88 percentage compared to original image. The edge and shape is not the same as original, also, the information of the pixel is missing along with the resolution decline.

the resolution of the raw image to higher or lower, I determined whether an object in the complex structure was a seedpoint or not. Fig. 25 shows a screen shot of a complex data set with multiple overlapping dendrites that could be misidentified as cell bodies.

D. Automated tracing

After obtaining the seedpoints, I used the seedpoints to run Han et al.'s tracing algorithm [4, 5]. After neuronal structure tracing, I analyzed morphological features of neurons and other structures of the mouse brain with image visualization programs: MeVisLab and ParaView. Also, I validated the seedpoint selection algorithm with the tracing result. The two coordinate system used by the seedpoint selection algorithm and the tracing algorithm were different, so synchronizing the two coordinates was needed. After modifying the seedpoint coordinate system, the coordinates were transferred to a seed-list file for use with the tracing algorithm. Finally, the tracing algorithm was initiated to reconstruct the KESM data sets.

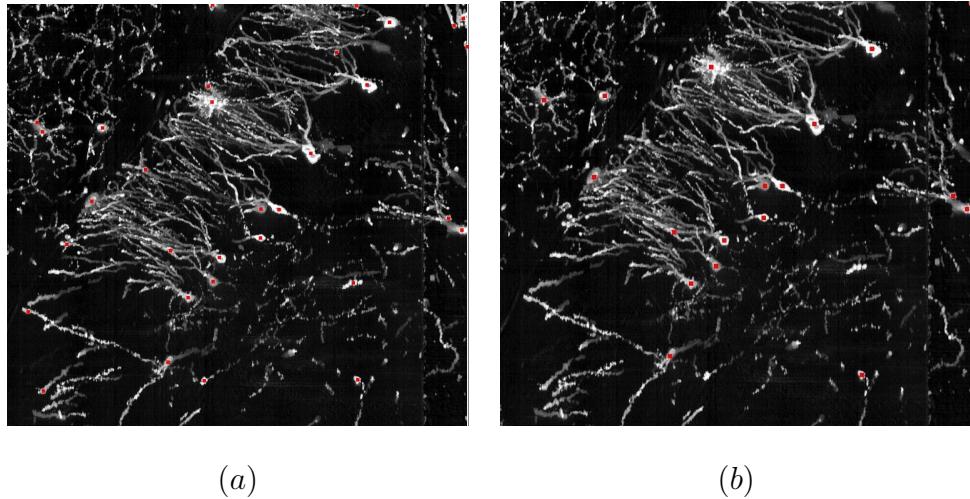


Fig. 26. Multi-scale approach. By changing the image resolution, somata were discriminated from densely overlapping dendrites. (a) The original image. (b) 40% resolution decreased from the original image. The unwanted overlapping dendrites were not detected by decreasing the image scale.

E. Counting soma

In this section, I investigate regional differences in neuronal distribution [28, 29]. The brain is divided into several parts related to different functions. For example, different parts of the brain are clearly distinguishable: the cortex, hippocampus, colliculi, and cerebellum as shown in Fig. 26(a). Each region has different populations of neurons and quantitative evaluation of the functional capacity of neuronal systems [30].

I counted somata to estimate the distribution of neurons in different areas in the KESM Golgi data sets. The procedure was as follows. From the superficial layer of the brain (cortex) to the deep regions (colliculi and hippocampus), the somata along the perpendicular to tangential directions were counted as shown in Fig. 26(b). Using the previous sections' methods (combining opening and closing), I implemented an algorithm for automatically counting somata on three different orthogonal directions (xy , yz , and zx), and measured the number of somata across different regions. Neu-

ronal cell bodies are spherical or nearly spherical structures with a diameter of about $6 - 12 \mu m$. So, soma can be identified as a counting unit within them [31].

In order to count the number of somata, I started with a small sample of $128 \times 128 \times 128$ voxels as shown in Fig. 28. Depending on the regions, I experimented with various thickness and multiple scanning directions. The experimental regions were cropped as rectangular blocks for counting as shown in Fig. 27. The counting direction is illustrated in Fig. 28. The xy , yz , and zx projections were extracted using the squeeze function in Matlab to count the number of somata. One projection, the xy , was chosen for counting. With only a single projection, parts of neuron were too small to distinguished. To detect the shape of neuronal features, I stacked 12 projections as shown in Fig. 28. The counting results will be shown in the following chapter.

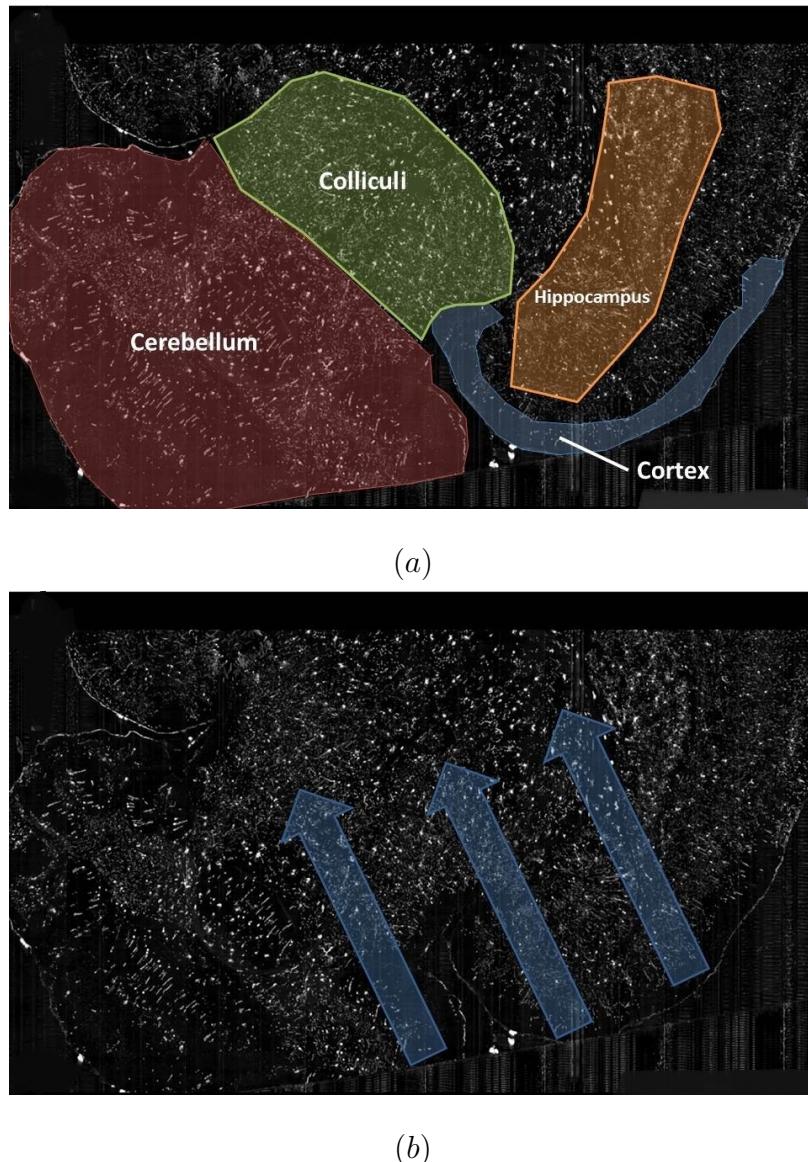


Fig. 27. Soma counting. (a) A regional annotations of mouse brain, (b) The perpendicular to tangential directions for counting somata from surface (cortex) to deeper regions (hippocampus and colliculi).

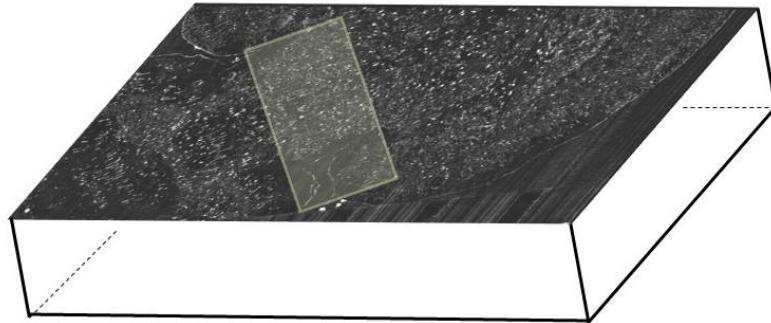


Fig. 28. Overview of part of the mouse brain. The shaded area is selected for counting.

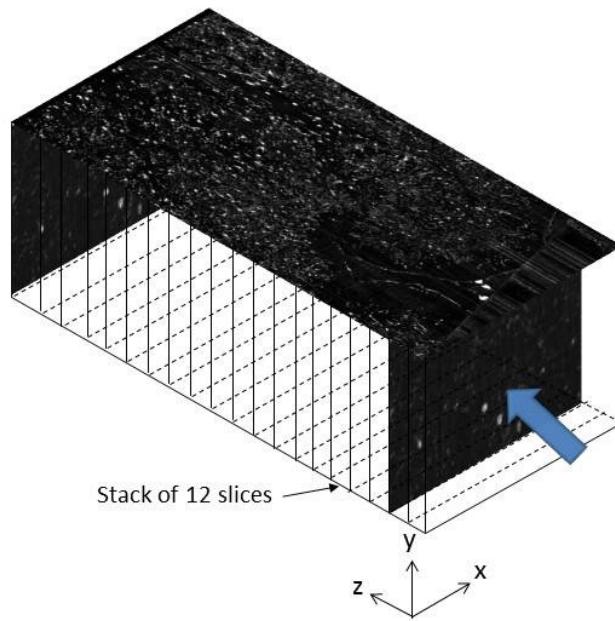


Fig. 29. Soma counting direction. Blue arrow represent the counting direction, which is perpendicular to the xy projection.

CHAPTER IV

RESULT AND ANALYSIS

In the previous chapter, several methodologies for setting seedpoints in the KESM data sets were described. In this chapter, through analysis and comparison of all the methodologies, the efficiency and accuracy of each method is analyzed. Also, I compared the performance between manual and automatic method for seedpoint selection.

In order to evaluate the effectiveness of the methods; a comprehensive experiment on various KESM data sets is necessary. The experimental volume size was $128 \times 128 \times 128$ (rows \times columns \times images) and the color depth was 256 gray-levels. After obtaining the seedpoints automatically, the result was analyzed using F-measure [32]. Detail regarding F-measure will be explained in section A. In section B, the performance result of automatic seedpoint selection from three different regional data sets will be presented. In section C, tracing results based on the automatically selected seedpoints will be shown. In section D, comparison of the tracing result based on manual or automatic seedpoints will be shown. In section E, the soma counting result will be described.

A. F-measure

In order to measure the accuracy of the introduced methodologies, F-measure was employed. F-measure consists of two primary metrics: precision (p) and recall (r). Precision is the proportion of correctly selected seedpoints among all returned seedpoints. Recall is defined as the proportion of correctly selected seedpoints among all true seedpoints. Below is the equation for F-measure [32]

$$F = \frac{2pr}{p + r} \quad (4.1)$$

where,

$$p \text{ (precision)} = n_3 / n_2$$

$$r \text{ (recall)} = n_3 / n_1$$

n_1 = true number of seedpoints

n_2 = number of all selected seedpoints

n_3 = number of correctly selected seedpoints

The F-measure values range between 0 to 1 and the higher the value of F-measure, the higher the performance.

B. Automatic seedpoint selection

The evaluation of the automatic seedpoint selection method was done with different regions in the Golgi-stained KESM data sets. The data sets were sampled ($128 \times 128 \times 128$ voxel) from the hippocampus region of the KESM data. Fig. 27(a), (b), and (c) had different density. The comparison of seedpoint selecting was proceeded perpendicular to xy plane at intervals of eight slices. As described in the methods chapter, each seedpoint selection algorithm was optimized for different density and neuronal morphology. Fig. 29 shows seedpoint selection results. Table I shows the precision, recall, and F-measure result of automatic seedpoint selection, respectively.

In the Fig. 29, the blue segmentation is neuronal structure (somata and dendrites) and the pink dot represents selected seedpoints, respectively. Depending on the image resolutions and validation methods, the number of detected seedpoints was different as can be seen in the Fig. 29. The left column shows the result of morphological object detection, the middle column is the visualization result of interactive 3D

validation, and the right column represents the result of multi-scale approach. The morphological object detection method and interactive 3D validation shows almost the same result. Visualization alone is not enough to show the similarity between the two conditions, however, Table I shows that they are approximately the same. Between morphological object detection and interactive 3D validation, there are only one or two seedpoints that need to be eliminated. By removing the redundant seedpoints, the precision and recall result of interactive 3D validation is changed to slightly higher than that of the morphological object detection method.

In case of sparse neuronal structure area, these two methodologies performed good enough to set seedpoints efficiently. However, Fig. 29 shows the limitations of these methodologies. There are several seedpoints that are set on the same segment of complex neuronal structures. In a densely overlapping area, the multi-scale approach eliminates such redundant seedpoint problems and improves the F-measure compared to the previous methodologies. As can be seen in Fig. 29, the right column of the figure has fewer seedpoints than the left column images. Also, Table I shows the result of improved precision, recall, and F-measure. The low resolution image could remove unnecessary noise and small segments that are not parts of the soma.

C. Tracing results by automatic seedpoint selection

In order to trace the neuronal structure with the automatically selected seedpoints, the voxel coordinates were transmitted to Han et al.’s tracing algorithm. Fig. 30 shows the result of tracing. For computational efficiency, $128 \times 128 \times 128$ volume data sets were traced. Han et al.’s original tracing algorithm was optimized for vasculature data sets. However, the algorithm worked well with the Golgi-stained data sets. In

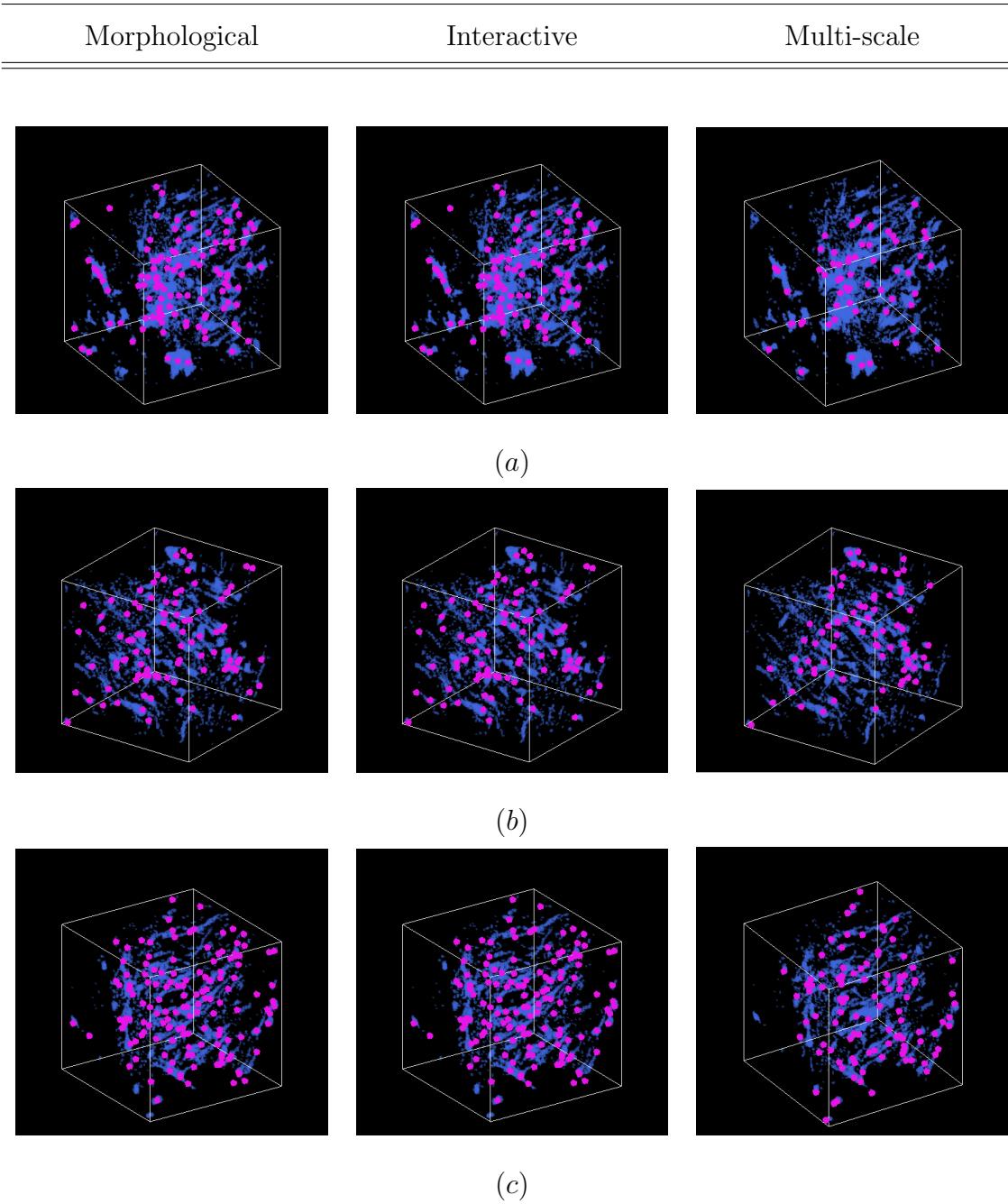


Fig. 30. Seedpoint selection result. (a), (b), and (c) Different regions of the KESM data set ($128 \times 128 \times 128$ each). Morphological object detection, interactive 3D validation method, and multi-scale approach were applied to select somata as seedpoints. Pink dots illustrate selected seedpoints. Blue segments represent the underlying neuronal structures.

Table I. F-measure comparison result from the three different hippocampus regions of the KESM data sets ((a), (b), and (c)). Each corresponding result is shown in Fig. 29. Each data sample is applied to different seedpoint selection methodologies. In general, the recall value is higher than the precision value. F-measure is increased when the multi-scale approach was applied to the same sample.

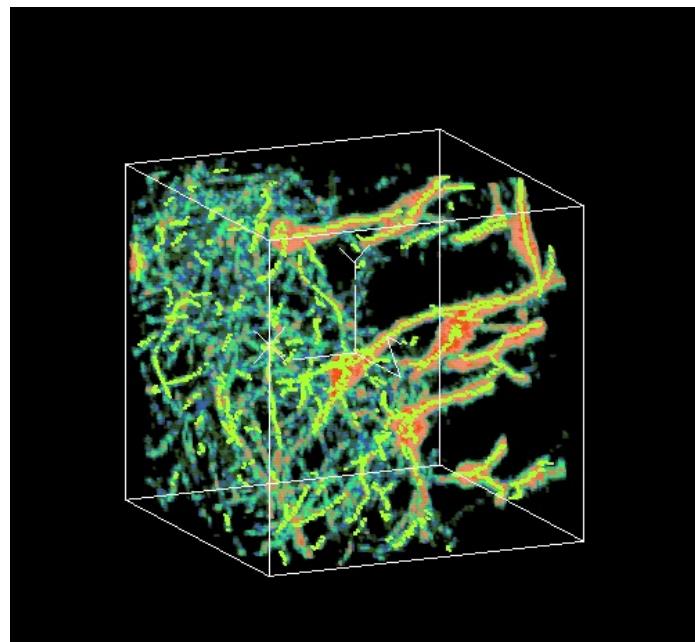
	Data sample	Methods	Precision	Recall	F-measure
(a)		Morphology	0.26	0.76	0.39
		Interactive	0.27	0.76	0.40
		Multi-scale	0.42	0.74	0.53
(b)		Morphology	0.37	0.82	0.51
		Interactive	0.38	0.84	0.52
		Multi-scale	0.44	0.87	0.58
(c)		Morphology	0.30	0.85	0.45
		Interactive	0.32	0.89	0.47
		Multi-scale	0.43	0.81	0.56

Fig. 30(a), the light green segments indicate the result of tracing. The long structures shown as red in Fig. 30(a) are somata and dendrites. The complicated structures to the left side are also neuronal structures. In dendrites, the tracing result was aligned along the structure. In case of soma, there were loopy tracing results. As described before, Han et al.'s algorithm was originally for vessel tracing, so in case of soma, the object diameter did not match that of the typical vessel segment, thus resulting in incorrect tracing, since the boundary detection and tracing direction decision in Han et al.'s algorithm did not apply well in this case.

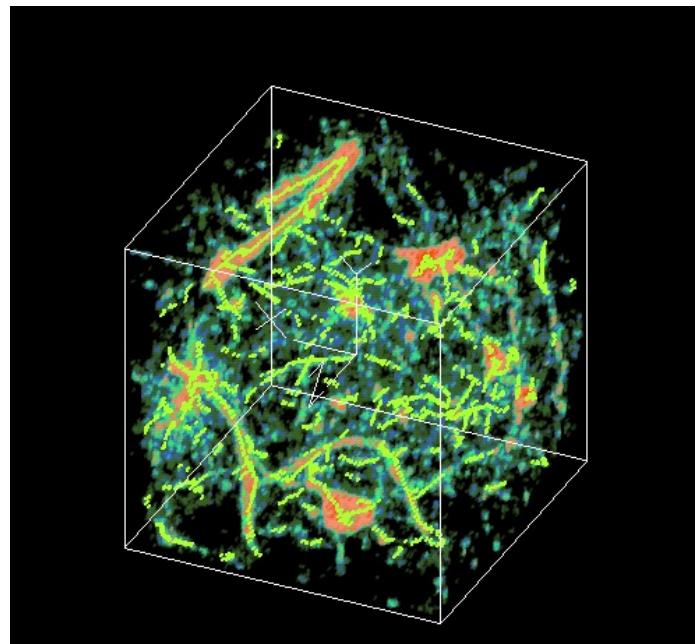
Comparing the tracing results and the seedpoint selection results showed that the total traced neuronal structure length depended upon the number of selected seedpoints. More seedpoints led to much longer traces than traces started from fewer selected somata in the same data sets. There was redundant tracing with more seedpoints, however, it led to more detailed tracing results, thus there is a trade-off.

D. Manual and automatic seedpoint selection comparison

In order to validate the performance of the automatic seedpoint selection methods, manual and automatic methods were compared. To evaluate the performance of the seedpoint selection methods, three different conditions were given: manual, automatic selection 1, and automatic selection 2. Manually, three users were asked to select fifty seedpoints each in the three data sets. Automatic selection 1 was configured to select similar number of seedpoints as that of the manual method. On the other hand, automatic selection 2 was allowed to select seedpoints as many as possible. Table II shows the results of the experiments. Automatic methods performed well in terms of time, but F-measure was not better than the manual method. When users selected a seedpoint, they checked and analyzed, rejecting small objects. However, in case of



(a)



(b)

Fig. 31. Tracing result. (a) and (b) The tracing results. The green curves represent the tracing result. The red objects are the underlying data showing somata and dendrites.

Table II. Comparison result of manual and automatic seedpoint selection

		Manual	Automatic 1	Automatic 2
	Seedpoints	50	68	301
(a)	Time (second)	95	3	6
	F-measure	0.92	0.74	0.25
	Seedpoints	50	64	269
(b)	Time (second)	105	4	7
	F-measure	0.94	0.86	0.30
	Seedpoints	50	58	320
(c)	Time (second)	125	5	7
	F-measure	0.80	0.75	0.24

automatic method, the algorithm tended to identify small particle as seedpoints. For the automatic selection 2 condition, seedpoint counts increased to almost six times the usual value, and the F-measure was improved and tracing result was much better than the automatic selection 1 condition. Fig. 31 illustrates the tracing results for the three different seedpoint selection conditions. Manual seedpoint selection identified seedpoints more correctly, but it took longer compared to automatic seedpoint selection methods. In general, automatic seedpoints performed well in terms of execution time. However, in terms of F-measure, the manual seedpoint selection was better.

E. Soma counting result

Based on the seedpoint selection methodologies, we can measure the number of somata in different area of the KESM data sets. I examined the result of the soma

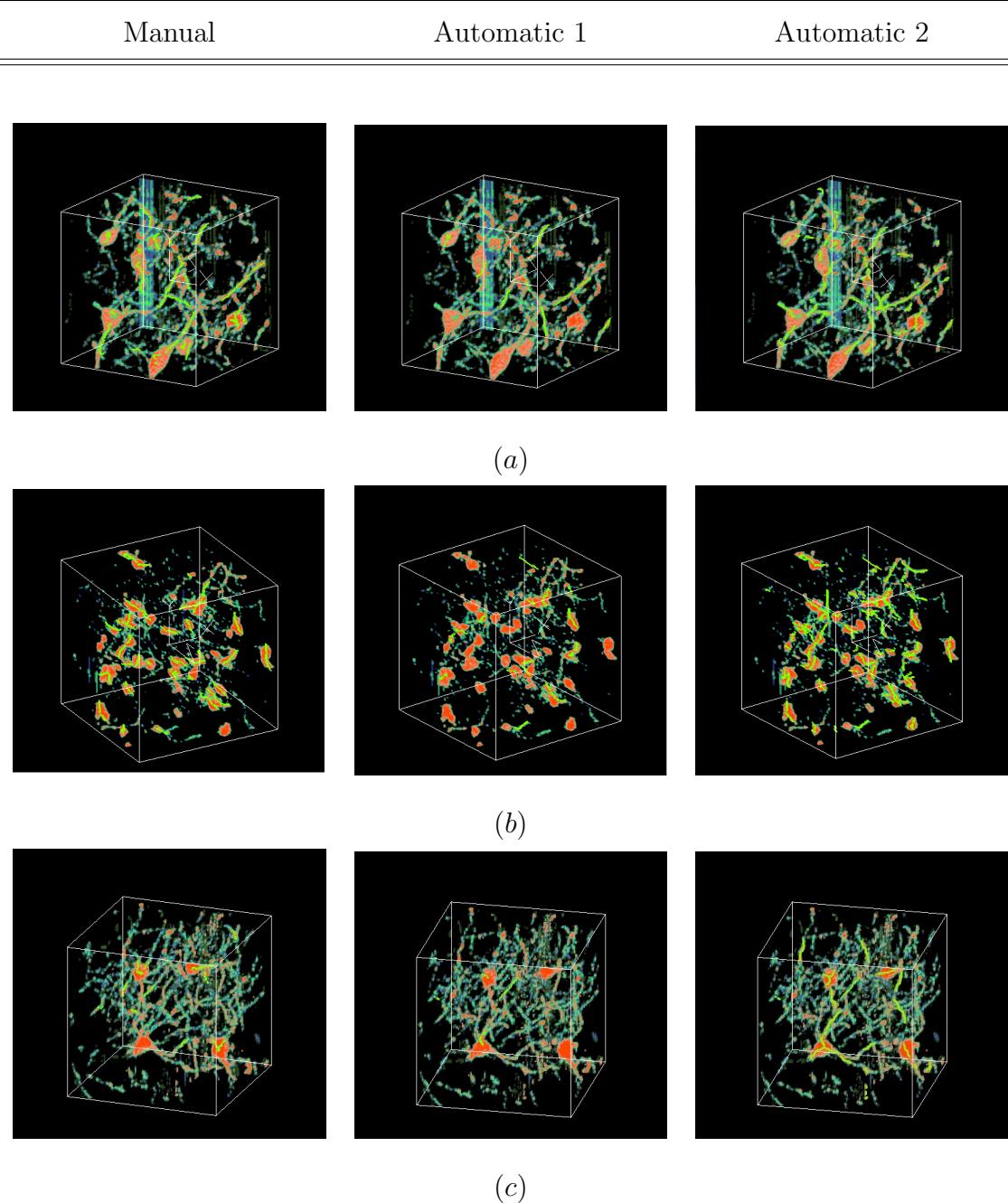


Fig. 32. (a), (b), and (c) Tracing result of manual and automatic seedpoint selection in different regions. Manual seedpoint selection chose fifty seedpoints. Automatic selection 1 chose seedpoints between fifty and sixty, similar to the manual condition. Automatic selection 2 selected seedpoints as much as possible. The green curves are the tracing results based on the acquired seedpoints.

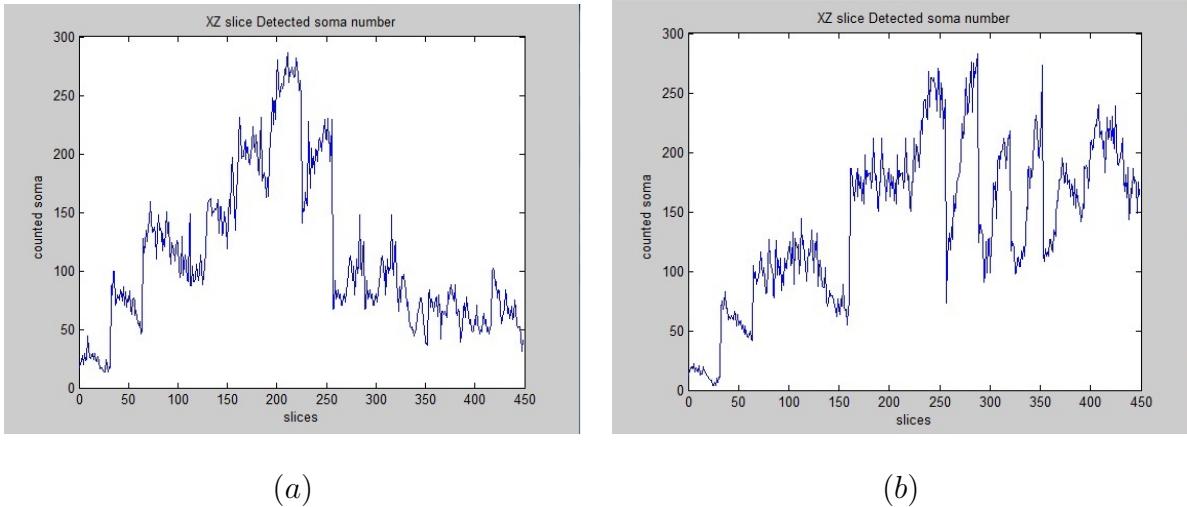


Fig. 33. Somata counting result. (a) and (b) The number of somata from cortex to hippocampus, respectively. Regionally, distribution of somata is different.

distribution to obtain a quantitative information regarding the regional variations. The distribution of soma was estimated from the superficial to the inner depth direction of the KESM data sets. To experiment, I connected $128 \times 128 \times 128$ volumes along the perpendicular to tangential direction. Regionally, the somata density is uneven. Due to different functions of brain, the neuronal density is different as shown in Fig. 32. Further visual inspection in MeVisLab shows that these gaps are areas with sparse somata and neuronal structures as shown in Fig. 33. The soma number counting algorithm can be useful in diagnosing malfunction in the brain as previously explained in the background chapter.

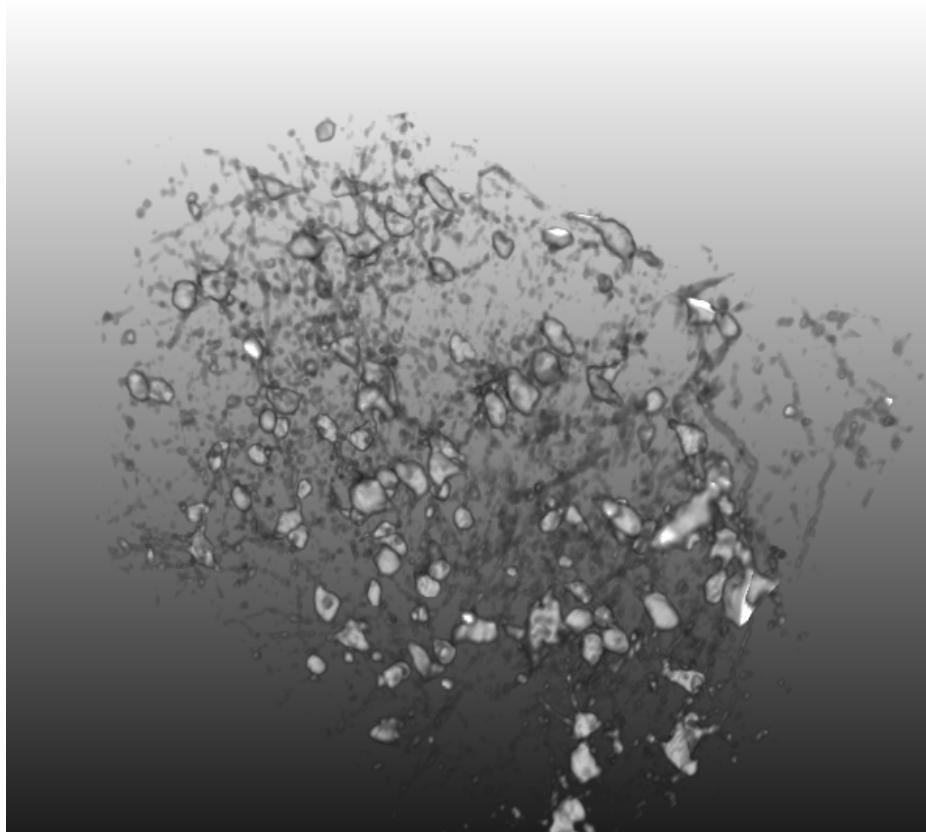


Fig. 34. Visualization using MeVisLab. This figure represents a portion of a visualization of soma distribution at a point in the middle of a counting direction (oriented from cortex to hippocampus). There are different densities of neuronal structures.

CHAPTER V

DISCUSSION AND CONCLUSION

This chapter discusses the main contribution of this research compared to related works, open issues, future directions, and conclusion.

A. Contributions

In order to overcome limitations in the related works in the background chapter, I presented and evaluated new methodologies to set seedpoints automatically. The main contributions of my methodologies are: (1) Automatic seedpoint selection: The algorithm can significantly reduce time needed to set seedpoints, without human intervention. (2) Tracing neuronal structures of the KESM data set with many seedpoints: Han et al.'s tracing algorithm can be initiated with enough seedpoints to trace large number of neuronal structures. (3) Quantitatively measured soma distribution: the number of somata was counted to investigate regional differences and morphological features. In summary, the proposed methodologies performed well on the large scale KESM data sets.

B. Open issues and future work

In this thesis, I have demonstrated that the proposed automatic seedpoint selection algorithm can complement Han et al.'s previous manual seedpoint selection method and quantitatively measure the number of somata of the mouse brain. However, there are several open issues: (1) a method to eliminate redundant seedpoints on the same segment (2) an advanced method for determining the threshold, (3) a noise filtering method (4) improvement on the interactive 3D validation framework. In this section,

I will discuss these open issues and future work to overcome these problems.

1. Redundant seedpoints

In order to trace large portions of the KESM data set, I had to set as many seedpoints as possible. As I described in the result and analysis chapter, the tracing result was better when more seedpoints were selected. However, by using the automatic method, redundant seedpoints were found, which made the tracing step inefficient. Due to the unnecessary seedpoints, the neuronal structures were traced several times depending on the number of seedpoints. Also, redundant seedpoints wandered off to a different direction: opposite direction or overlapping preexisting traces. To reduce the tracing time of the entire mouse brain and correctly trace the neuronal structures, redundant seedpoints need to be eliminated.

As I explained in the methods chapter, adaptive thresholding was used to identify seedpoints in complex structures. The adaptive thresholding method performed well, but it produced excessive seedpoints. It generated redundant seedpoints within a single object. In the future, I will apply more advanced thresholding techniques to overcome this limitation.

2. Loss of seedpoints

The KESM data sets contain neuronal structures of various shapes and sizes. Tiny dots or thin lines were removed by morphological object detection and median filtering. However, loss of essential segment happened at the same time due to the intensity values or irregular shapes. Without verification, those components would be incorrectly considered as noise, where the result of tracing would be disconnected dendrites or abnormal connections.

In order to solve the loss of seedpoints due to excessive noise removal, I will

develop methods to preserve seedpoints in significant parts of the data set. Also, I will investigate other adaptive methods to decide the threshold value and filtering conditions.

3. Limitations of the interactive 3D validation method

The interactive 3D validation method was used to discriminate overlapping seedpoints and to define part of a neuron as a seedpoint in massively complex neuronal structures in the KESM data sets. However, performance of the method was not enough to overcome all problems. The method was based on manually matching the coordinates in all three orthogonal projections. However, it was hard to match the corresponding points in the three projections. So, there was no big difference between the morphological object detection algorithm and the interactive 3D validation method. In order to enhance the interactive 3D validation method, another coordinate matching technique is needed.

In the future, I will develop an advanced technique that can discriminate overlapping seedpoints using the pixels' intensity range. By fixing a specific intensity bounds between obtained coordinates, this method will be able to pair more corresponding coordinates as matching coordinates.

4. Experimenting with the entire KESM data set

In this research, only a small portion of the KESM data set was used. Basically, I experimented with $128 \times 128 \times 128$ volume data sets. Because the KESM data sets' size was too big to execute in one single experiment with the current computer systems, small interesting portions of the KESM was sampled and processed. The neuronal structures were traced well locally by using the automatic seedpoint selection algorithm and Han et al.'s tracing algorithm. However, the ultimate goal of the BNL

is to investigate an entire mouse brain to understand the morphology and function through the reconstructed connectivity. Therefore, by utilizing increasing computer performance (CPU, GPU, memory, and programming language), I will experiment with larger chunks of the KESM data set.

C. Conclusion

The aim of this research was to develop a robust algorithm for automatically setting seedpoints for the KESM Golgi data set. The new automatic seedpoint selection algorithm was developed to complement the previous manual approach in Han et al.'s work. Through comprehensive experiments, this research improved the previous manual seedpoint selection method [4, 5], to set the seedpoints automatically. This method helped rapidly reconstruct the sub-micrometer resolution KESM data set. By combining the morphological object detection algorithm, interactive 3D seedpoint validation, and multi-scale approach, I reduced the processing time and decreased the error rate in seedpoint selection. The proposed algorithms were more robust and comprehensive than conventional approaches. Furthermore, I contributed to the analysis of quantitative distribution of neurons and to the tracing of neuronal structures in the mouse brain. The methodologies introduced in this thesis are expected to help analyze larger portions of the KESM data set.

REFERENCES

- [1] P. Dayan and L. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Cambridge, Massachusetts: The MIT Press, 2001.
- [2] K. R. Brizzee, “Neurons numbers and dendritic extent in normal aging and alzheimer’s disease.” *Neurobiol Aging*, vol. 8, no. 6, pp. 579–580, 1987.
- [3] D. Mayerich, L. Abbott, and B. McCormick, “Knife-Edge Scanning Microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain.” *Journal of Microscopy*, vol. 231, no. pt. 1, pp. 134–143, Jul 2008.
- [4] D. Han, “Rapid 3D tracing of the mouse brain neurovasculature with local maximum intensity projection and moving windows,” Ph.D. dissertation, Department of Computer Science and Engineering, Texas A&M University, College Station, 2009.
- [5] D. Han, J. Keyser, and Y. Choe, “A local maximum intensity projection tracing of vasculature in knife-edge scanning microscope volume data,” in *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI*, 2009, pp. 1259–1262.
- [6] B. I. Justusson, “Two-dimensional digital signal prcessing II,” *Topics in Applied Physics*, vol. 43, pp. 161–196, 1981.
- [7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Boston: Addison-Wesley Longman Publishing Co., Inc., 1992.
- [8] Y. Zhang, X. Zhou, A. Degterev, M. Lipinski, D. Adjeroh, J. Yuan, and S. T. C. Wong, “A novel tracing algorithm for high throughput imaging screening of

neuron-based assays.” *J Neurosci Methods*, vol. 160, no. 1, pp. 149–162, Feb 2007.

- [9] 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 Trans. Inform. Technol. Biomed.*, vol. 3, no. 2, pp. 125–138, 1999.
- [10] K. A. Al-Kofahi, S. Lasek, D. H. Szarowski, C. J. Pace, G. Nagy, J. N. Turner, and B. Roysam, “Rapid automated three-dimensional tracing of neurons from confocal image stacks,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 2, pp. 171–187, 2002.
- [11] Y. Kang, K. Engelke, and W. A. Kalender, “Interactive 3D editing tools for image segmentation,” *Medical Image Analysis*, vol. 8, no. 1, pp. 35 – 46, 2004.
- [12] J. Varandas, P. Baptista, J. Santos, R. Martins, and J. Dias, “VOLUS—a visualization system for 3D ultrasound data,” *Ultrasonics*, vol. 42, no. 1-9, pp. 689 – 694, 2004.
- [13] L. Westover, “Interactive volume rendering,” in *Proc. Chapel Hill workshop on volume visualization*, 1989, pp. 9–16.
- [14] W. Zhang, G. Zelinsky, and D. Samaras, “Real-time accurate object detection using multiple resolutions,” *IEEE 11th International Conference on Computer Vision*, pp. 1–8, 2007.
- [15] F. Murtagh, W. Zeilinger, J.-L. Starck, and Bijaoui, “Object detection using multi-resolution analysis,” in *Astronomical data Analysis Software and Systems*

- IV*, D. Shaw, H. Payne and J. Hayes, Eds. Baltimore: Astronomical Society of the Pacific, 1994, pp. 260–263.
- [16] P. J. Uhlhaas and W. Singer, “Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology,” *Neuron*, vol. 52, no. 1, pp. 155 – 168, 2006.
 - [17] A. C. D’Souza, “Automated counting of cell bodies using nissl stained cross-sectional images,” Master’s thesis, Department of Computer Science and Engineering, Texas A&M University, College Station, 2007.
 - [18] P. R. Rapp and M. Gallagher, “Preserved neuron number in the hippocampus of aged rats with spatial learning deficits.” in *Proc. Natl. Acad. Sci.*, vol. 93, no. 18, Sep 1996, pp. 9926–9930.
 - [19] J. Kwon, D. Mayerich, Y. Choe, and B. H. McCormick, “Automated lateral sectioning for Knife-Edge Scanning Microscopy,” in *Proc. 5th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI*, 2008, pp. 1371–1374.
 - [20] D. Mayerich, B. H. McCormick, and J. Keyser, “Noise and artifact removal in knife-edge scanning microscopy,” in *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, 2007, pp. 556–559.
 - [21] B. McCormick and D. Mayerich, “Three-dimensional imaging using Knife-Edge Scanning Microscope,” *Microscopy and Microanalysis*, vol. 10, Sup.2, pp. 1466–1467, 2004.
 - [22] R. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Proc. Medical Image Computing and Computer-Assisted Intervention*, vol. 1496, 1998, pp. 130–137.

- [23] E. Pannese, “The Golgi stain: Invention, diffusion and impact on neurosciences,” *Journal of the History of the Neurosciences: Basic and Clinical Perspectives*, vol. 8, no. 2, pp. 132–140, 1999.
- [24] D. M. Mayerich, “Acquisition and reconstruction of brain tissue using knife edge scanning microscopy,” Master’s thesis, Department of Computer Science and Engineering, Texas A&M University, College Station, 2003.
- [25] A. R. Weeks, *Fundamentals of Electronic Image Processing*. Bellingham, WA: SPIE Optical Engineering Press, 1996.
- [26] N. Ray and B. N. Saha, “Edge sensitive variational image thresholding,” in *Proc. IEEE Int. Conf. Image Processing*, vol. 6, 2007.
- [27] R. M. Haralick, S. R. Sternberg, and X. Zhuang, “Image analysis using mathematical morphology,” *Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532 –550, July 1987.
- [28] R. E. Coggeshall and H. A. Lekan, “Methods for determining numbers of cells and synapses: A case for more uniform standards of review,” *The Journal of Comparative Neurology*, vol. 364, no. 1, pp. 6–15, 1996.
- [29] T. Miki, Y. Fukui, M. Itoh, S. Hisano, Q. Xie, and Y. Takeuchi, “Estimation of the numerical densities of neurons and synapses in cerebral cortex,” *Brain Research Protocols*, vol. 2, no. 1, pp. 9 – 16, 1997.
- [30] M. J. West, “Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias,” *Trends in Neurosciences*, vol. 22, no. 2, pp. 51 – 61, 1999.

- [31] K. Bedi, “Lasting neuroanatomical changes following undernutrition during early life,” in *Early Nutrition and Later Achievement*, J.Dobbing Ed. London: Academic Press, 1987, pp. 1–49.
- [32] B. Larsen and C. Aone, “Fast and effective text mining using linear-time document clustering,” in *Proc. 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1999, pp. 16–22.

VITA

Dongkun Kim was born in Busan, The Republic of Korea. He received his B.S. in Weapons Engineering and Military Science from the Korea Military Academy in 2005, after which he was commissioned an officer in the ROK army. He served in the ROK Army from 2005 to 2009 as a platoon leader, staff officer, aide-de-camp of Corps Commander. He entered the Department of Computer Science and Engineering at Texas A&M University in August 2009 and graduated with his M.S. degree in August 2011.

Permanent Address:

Department of Computer Science and Engineering,
College Station, TX 77843-3112

Chair of Committee: Dr. Yoonsuck Choe

E-mail Address : dongkunkim64@gmail.com