Imperial College of Science, Technology and Medicine
Department of Bioengineering

Generalizable Automated Pixel-level Structural Segmentation of Medical and Biological Data

Shearin Shuoying Cao

First Supervisor: Dr. Anil Anthony Bharath
Second Supervisor: Prof. Kim Howard Parker

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Bioengineering of Imperial College London
and the Diploma of Imperial College London, December 2012

Abstract

Over the years, the rapid expansion in imaging techniques and equipments has driven the demand for more automation in handling large medical and biological data sets. A wealth of approaches have been suggested as optimal solutions for their respective imaging types. These solutions span various image resolutions, modalities and contrast (staining) mechanisms. Few approaches generalise well across multiple image types, contrasts or resolution.

This thesis proposes an automated pixel-level framework that addresses 2D, 2D+t and 3D structural segmentation in a more generalizable manner, yet has enough adaptability to address a number of specific image modalities, spanning retinal funduscopy, sequential fluorescein angiography and two-photon microscopy.

The pixel-level segmentation scheme involves: *i*) constructing a phase-invariant orientation field of the local spatial neighbourhood; *ii*) combining local feature maps with intensity-based measures in a structural patch context; *iii*) using a complex supervised learning process to interpret the combination of all the elements in the patch in order to reach a classification decision. This has the advantage of transferability from retinal blood vessels in 2D to neural structures in 3D.

To process the temporal components in non-standard 2D+t retinal angiography sequences, we first introduce a co-registration procedure: at the pairwise level, we combine projective RANSAC with a quadratic homography transformation to map the coordinate systems between any two frames. At the joint level, we construct a hierarchical approach in order for each individual frame to be registered to the global reference intra- and inter- sequence(s). We then take a non-training approach that searches in both the spatial neighbourhood of each pixel and the filter output across varying scales to locate and link microvascular centrelines to (sub-) pixel accuracy. In essence, this "link while extract" piece-wise segmentation approach combines the local phase-invariant orientation field information with additional local phase estimates to obtain a soft classification of the centreline (sub-) pixel locations.

Unlike retinal segmentation problems where vasculature is the main focus, 3D neural segmentation requires additional flexibility, allowing a variety of structures of anatomical importance yet with different geometric properties to be differentiated both from the background and against

other structures. Notably, cellular structures, such as Purkinje cells, neural dendrites and interneurons, all display certain elongation along their medial axes, yet each class has a characteristic shape captured by an orientation field that distinguishes it from other structures. To take this into consideration, we introduce a 5D orientation mapping to capture these orientation properties. This mapping is incorporated into the local feature map description prior to a learning machine.

Extensive performance evaluations and validation of each of the techniques presented in this thesis is carried out. For retinal fundus images, we compute Receiver Operating Characteristic (ROC) curves on existing public databases (DRIVE & STARE) to assess and compare our algorithms with other benchmark methods. For 2D+t retinal angiography sequences, we compute the error metrics ("Centreline Error") of our scheme with other benchmark methods. For microscopic cortical data stacks, we present segmentation results on both surrogate data with known ground-truth and experimental rat cerebellar cortex two-photon microscopic tissue stacks.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Declaration of Originality

The work presented in this thesis was carried out between October 2008 and September 2012 in the Department of Bioengineering at Imperial College London under the supervision of Dr. Anil Anthony Bharath and Prof. Kim Howard Parker. The material herein is original and has not been submitted for a degree or diploma at any other university.

Retinal fluorescein angiograms were kindly provided by the Mr. John Arnold, an imaging consultant from Charing Cross Hospital, Imperial College Healthcare NHS Trust. Retinal fundus images were obtained from two open-access sources: the "Digital Retinal Images for Vessel Extraction" (DRIVE) online repository that is currently hosted by the Image Sciences Institute at University Medical Center Utrecht, and the "STructured Analysis of the Retina" (STARE) online repository that is currently hosted by College of Engineering and Science at Clemson University. The two-photon microscopy image stacks were kindly provided by Dr. Simon Schultz at Schultz Laboratory, Department of Bioengineering at Imperial College London.

The author has developed different algorithms for automated analysis based on code developed by herself and standard and modified libraries of Matlab, unless otherwise stated.

Shearin Shuoying Cao

Publication List

Publications that have been produced based on the material in this thesis are as follows:

- **Microvascular Segmentation of Co-Registered Retinal Angiogram Sequences**
  Shuoying Cao, Anil A. Bharath, Kim H. Parker, Jeffrey Ng, John Arnold, Alison McGregor, Adam Hill, *In The Annals of The British Machine Vision Association (BMVA) (Roy Davies, ed.)*, 2012.

- **Patch-Based Feature Maps For Pixel-Level Image Segmentation**
  Shuoying Cao, Saadia Iftikhar, Anil A. Bharath, *In 20th European Signal Processing Conference (EUSIPCO)*, pp. 2263-2267, 2012.

- **Patch-based Automatic Retinal Vessel Segmentation in Global and Local Structural Context**
  Shuoying Cao, Anil A. Bharath, Kim H. Parker, Jeffrey Ng, *In 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4942-4945, 2012.

- **Spatio-temporal registration and microvasculature segmentation of retinal angiogram sequences**
  Shuoying Cao, Anil A. Bharath, Kim H. Parker, Jeffrey Ng, John Arnold, Alison McGregor, Adam Hill, *In Proceedings of Medical Image Understanding and Analysis 2011 (Bill Crum, Graeme Penney, eds.)*, pp. 293-297, 2011.

- **Joint Spatio-temporal Registration and Microvasculature Segmentation of Retinal Angiogram Sequences**
  Shuoying Cao, Anil A. Bharath, Kim H. Parker, Jeffrey Ng, John Arnold, Alison McGregor, Adam Hill, *In Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2618-2621, 2011.

- **Joint Spatial - Temporal Registration of Retinal Angiograms**
  Shuoying Cao, Anil A. Bharath, Kim H. Parker, Jeffrey Ng, John Arnold, Alison McGregor, Adam Hill, *In Proceedings of Medical Image Understanding and Analysis 2010 (Abhir Bhalerao, Nasir Rajpoot, eds.)*, pp. 179-184, 2010.

Awards and Bursary

- "IPEM Bursary Scheme" awarded by the the Institute of Physics and Engineering in Medicine (IPEM) in 2012, restricted to one award each institution per annual.
- "I J Shelley Travel Grant" awarded by the Institution of Engineering and Technology (IET) in 2012, only six awarded per annual worldwide.
- "Best Poster Award" jointly awarded by the King's Centre for Excellence in Medical Engineering and the MIUA chairing committee at the "Medical Image Understanding and Analysis (MIUA) 2011".
- "OC Trust Postgraduate Travel Award" awarded by City & Guilds College Assocation (CGCA) Old Centralians' Trust in 2011.
- Awarded full-sponsorship from the British Machine Vision Association (BMVA) to attend the "Medical Image Understanding and Analysis (MIUA) 2009".
- Awarded full-sponsorship from IDEA League (five leading European universities in Science and Technology) and Imperial Graduate School at intercollegiate exchange summer schools: ParisTech 2010, Tsinghua University 2010, TU Delft 2009 and Imperial College London 2009.
- Research bursary awarded by the Hammersmith Hospital Trustees, Imperial College Healthcare NHS Trust and the Royal Brompton & Harefield NHS Foundation Trust.

Acknowledgements

I would like to thank all the following mentors and funding bodies, without whom this would not have been possible:

- Dr. Anil A. Bharath for his supervision. He financially supported my training and attendance in various conferences, workshops and summer schools. He was also supportive of my extra-curricular responsibilities at the Imperial Graduate School.

- Prof. Kim H. Parker for his supervision. He taught me how to write in simple and efficient English, how to grasp the essence of the arguments without having to use big words, and most importantly, how to sprinkle humour onto everything in life.

- Dr. Jeffrey Ng for his supervision in the first year of my PhD. He was my mentor from the start, and nurtured my coding capabilities and the "can-do" attitude.

- Dr. Simon Schultz for the neural data and Mr. John Arnold for the retinal data. They both have endured the hardwork of sampling the high quality data, and have both sit through long meetings and patiently walked me through the technical details.

- Prof. Alison McGregor and Dr. Adam Hill for convincing the Hammersmith Hospital Trustees to fund the preliminary study. This kicked start this amazing journey at 2008 and introduced me to the wonderful world of medical image processing.

- Both the Hammersmith Hospital Trustees, Imperial College Healthcare NHS Trust and the Royal Brompton & Harefield NHS Foundation Trust for research bursary; The Institute of Physics and Engineering in Medicine, the Institution of Engineering and Technology and the City & Guilds College Assocation (CGCA) Old Centralians' Trust for travel awards; The British Machine Vision Association, the IDEA League, the Imperial Graduate School and the BICV lab for sponsorship.

- Dr. Andy Jack for his advice and support. He has been my rock for the past 3 years and always has my best interest at heart.

- All the past and current BICV group members, Zeynep, Lucas, Yannis and Jose.

Dedication

I would like to dedicate this work to my parents, Prof. Junjie Cao and Prof. Jingjuan Sun. They are both high achievers in their own rights, and I would not have come this far without their perseverance and support!

"C makes it easy to shoot yourself in the foot; C++ makes it harder, but when you do, it blows away your whole leg."
- *Bjarne Stroustrup*

"There is only one nature - the division into science and engineering is a human imposition, not a natural one. Indeed, the division is a human failure; it reflects our limited capacity to comprehend the whole."
- *Bill Wulf*

"Use orderliness and stability to confront chaos and disorder. Use calmness and steadfastness to deal with noisiness and clamour. This is control of the psychological factor. "
- *Sun Tzu*

"He who knows does not speak. He who speaks does not know. He who is truthful is not showy. He who is showy is not truthful."
- *Lao Tse*

"Not knowing that one knows is best. Thinking that one knows when one does not know is sickness. Only when one becomes sick of the sickness can one be free from sickness."
- *Lao Tse*

Contents

Abstract i

Copyright Declaration ii

Declaration of Originality iii

Publication List iv

Awards and Bursary v

Acknowledgements vi

1 Introduction 1

1.1 Theme 1

1.2 Background of Image Modalities 3

1.2.1 Retinal Imaging 3

1.2.2 Neural Imaging 15

1.3 Thesis Structure 26

2 2D Retinal Funduscopy: Segmentation 29

2.1 Introduction 29

2.2 Context 30

2.3 Literature Review on Retinal Vascular Segmentation 30

2.4 Proposed Pixel-level Patch-based Segmentation with Training 33

2.4.1 Construction of Phase-invariant Orientation Field of the Local Spatial Neighbourhood 33

2.4.2 Construction of Feature Patches from Intensity-based Measures and Local Orientation Feature Maps 36

2.4.3 Supervised Machine Learning (MLP or SVM) for "Soft" Classification . . 39

2.5 Discussion and Conclusion 45

3 2D+t Sequential Fluorescein Angiogram I: Co-Registration 48

3.1 Introduction 48

3.2 Context 49

3.2.1 Application-Based Relevance 49

3.2.2 Feature-/Area- Based Relevance 50

3.3 Literature Review on Registration in Fluorescein Angiograms 53

3.4 Proposed Co-registration Procedure: "Joint Global RANSAC" method 54

3.4.1 Pairwise (6 step process): Projective RANSAC with Quadratic Homography Transformation 55

3.4.2 Joint: Hierarchical Compound of Transformation Matrices 75

3.5 Discussion and Conclusion 79

4 2D+t Sequential Fluorescein Angiogram II: Centreline Extraction and Tracking 84

4.1 Introduction 84

4.2 Context 85

4.3 Literature Review on Retinal Microvascular Centreline Segmentation 86

4.4 Suggested Method 88

4.4.1 Proposed "Link while Extract" Piece-wise Centreline Segmentation . . . 88

4.5 Discussion and Conclusion 92

5 3D Microscopic Stack of Cerebellar Cortex: Segmentation 96

5.1 Introduction 96

5.2 Context 97

5.3 Literature Review on Microvascular Segmentation in Neural Imaging 99

5.3.1 Image Preprocessing 99

5.3.2 Soma Segmentation 100

5.3.3 Neurite Segmentation 100

5.4 Proposed Pixel-level Patch-based Segmentation with MLP Training 101

5.4.1 3D to 5D Orientation Mapping to Preserve Angular Information and Line Curvature 101

5.4.2 Filter Realization of 3D to 5D Orientation Mapping 106

5.4.3 Supervised Machine Learning on Constructed High-dimensional Feature Patch 110

5.5 Discussion 120

6 Evaluation and Validation 121

6.1 Introduction 121

6.2 2D Retinal Funduscopy: Receiver Operating Characteristic (ROC) 122

6.2.1 Definition: Receiver Operating Characteristic (ROC) 122

6.2.2 DRIVE Database: Results 123

6.2.3 STARE Database: Results 126

6.3 2D+t Sequential Fluorescein Angiogram: Centreline Error 131

6.3.1 Context & Literature Review : Registration Accuracy for 2D+t Sequential Fluorescein Angiography 131

6.3.2 Definition: Centreline Error 132

6.3.3 Results 133

6.4 3D Microscopic Stack of Cerebellar Cortex 138

6.4.1 Surrogate Stack: Implementation and Results 138

6.4.2 Experimental Data : Results 148

6.5 Discussion and Conclusion 158

7 Discussion and Conclusion **159**

Bibliography **165**

List of Tables

2.1 *Comparison between MLP and SVM* 43

3.1 Transformation Models 71

6.1 *AUC* and *Acc* comparison for 10 different segmentation methods and human observer on 20 test images from DRIVE database. 127

6.2 *AUC and Acc comparison for our PP-MLP method evaluated on the STARE database with different training patch size, incrementing from less than 0.5% to 2% of the total image pixels.* 130

6.3 *AUC* and *Acc* comparison for 8 different segmentation methods and human observer on 20 images from STARE database. 130

6.4 *Statistical comparison of Centreline Error (CE) between our joint Global-RANSAC registration scheme with a controlled Global Affine transformation model on 384 angiograms.* 134

List of Figures

1.1 Two contrasting approaches. Top: Segmentation is followed by explicit edge linking; Bottom: The decision process at individual pixels utilises explicit information on the local phase-invariant orientation **field**. 2

1.2 A Topcon TRC NW6S digital non-mydriatic retinal fundus camera [1]. 6

1.3 An example retinal fundus image from the DRIVE database in original RGB channels [2]. 8

1.4 An example retinal fundus image from the DRIVE database, converted to gray-scale for processing. 8

1.5 An example retinal fundus image from an individual exhibiting normal vessels from the STARE database [3]. 10

1.6 An example retinal fundus image from a patient exhibiting abnormal vessels from the STARE database [3]. 10

1.7 An example hand-labelled vessels segmentation from two users. Left label was considered as more conservative as there were fewer pixels representing the vessels in the binary image than the right label [3]. 11

1.8 A single fluorescein angiogram from a pre-operative intervention sequence in our clinical data set. 15

1.9 Four types of light microscopy analyzing the same cell. (A) Brightfield microscopy; (B) Phase-contrast microscopy; (C) DIC or Nomarski microscopy; (D) Darkfield microscopy [4]. 16

1.10 Comparison of different set-ups of fluorescence microscopy: (A) *Epifluorescence* microscopy; (B) *Confocal* microscopy; (C) *Two-photon* microscopy [5]. 18

1.11 The schematic diagram of the layout of a transmission electron microscopy (TEM) [5]. 20

1.12 The schematic diagram of the layout of a scanning electron microscopy (SEM) [5]. 20

1.13 (A) *in vivo* two-photon microscopic imaging of rat cerebellum after calcium labeling. Three slices taken at increasing tissue depth to illustrate different structural layers: (A-1) the **molecular layer**, (A-2) the **Purkinje cell layer**, (A-3) the **granule cell layer**. (B) Maximum-intensity projection of 100 slices, each at 2 μm increments, to reconstruct a sagittal view of the labeled rat cerebellar cortex. The reconstruction begins at the brain surface and extends to the granule cell layer at a depth of 200 μm. (C) Illustrative diagram of cellular composition in (B), identifying the main cell types present: **Purkinje cells** in blue, **interneurons** in red, and **Bergmann glia** in green. (D) A zoomed-in image of the molecular layer, similar to (A-1). Top: The slice was taken at 96 μm below the surface; Bottom: The schematic diagram illustrates the main cellular components using the same color coding as in (C). Blood vessels are unstained and appear as dark structures. Note, all scale bars above represent 20 μm. Adapted from [6]. 23

1.14 A 3D schematic diagram of the interaction among components in the cerebellar cortex with Purkinje cells in blue and granular cells in green. The red-yellow circle marks the synapses formed by the ascending axon on Purkinje cell dendrites, and the red circle marks the synapses formed by the parallel fibers on Purkinje cells along their path. Adapted from [7]. 24

1.15 A 3D view of a sample rat cerebellar cortex microscopic tissue stack ($512 \times 512 \times 251$ at 8-bit). 25

1.16 A montage view of a sample rat cerebellar cortex microscopic tissue stack, only displaying every other slice along the Z axis. 26

2.1 Three-dimensional image of an intensity profile from a retinal vessel obtained by fluorescein imaging, cross-section resembles "double-Gaussian" construct. . . . 31

2.2 Estimated phase-invariant orientation field (in reds) of a 128 × 128 patch from a retinal angiogram with 3D view. 35

2.3 Estimated phase-invariant orientation field (in reds) of a 128 × 128 patch from a retinal angiogram with a zoomed-in detailed 2D view. 35

2.4 Estimated phase-invariant orientation field (in red) for a high-quality sub-region from a microscopy image. Regions are typically lower contrast and with greater degrees of noise. 36

2.5 Oriented band pass frequency response at 0 radians (a), $\pi/4$ radians (b), $\pi/2$ radians, and $3\pi/4$ radians (d). 37

2.6 Real and imaginary parts of oriented complex bandpass filter kernels used to produce the filter outputs in the spatial domain. 37

2.7 A simple Multilayer Perceptron with one input layer, one hidden layer and one output layer. 40

2.8 Sigmoid, Gaussian, Hyperbolic tangent and secant transfer functions normalize the output signal generated by each node. 41

2.9 The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 1 (left) in RGB scale. . . 45

2.10 The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 2 (left) in RGB scale. . . . 47

2.11 The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 3 (left) in RGB scale. . . . 47

3.1 Hierarchical stages of processing for sequential registration. 55

3.2 (a) Original circle with radius r; (b) For each pixel located on the contour (in dash line), a potential circle-centre locus is defined by a circle (in solid line) with radius r and centre at that particular pixel; (c) The frequency of which image pixels occur in the circle-centre loci is determined, the highest frequency pixel (black dot) represents the centre of the circle; (d) Circle detection in the presence of incomplete circle information (occlusion) and overlapping [8]. 57

3.3 Retinal "field-stop" extraction using the Hough Transform. The detected retina boundary is accentuated in red, and the centre of the retina detected is marked in blue. 59

3.4 The parameter space of the input retinal image. 59

3.5 Histogram equalization [8]. 60

3.6 The partial derivatives, I_x and I_y, of three image patches containing different structural characteristics, adapted from [9]. 64

3.7 Classification of image points using eigenvalues of Harris matrix, adapted from [9]. 64

3.8 The distribution of the partial derivatives are very different for the three different types of image patches/neighboughoods. This allows us to characterize each type of structure using the size and shape of the principal component ellipse. The R value for each structure type is also significantly different. 65

3.9 Left: A data set with many outliers for which a line has to be fitted: Right: Fitted line (in blue) using RANSAC. Outliers (in red) have no influence on the result. 68

3.10 Different transformations. (a) the fixed image; (b) the moving image with a grid overlayed; (c) the moving image aligned using a translation transformation; (d) aligned using a rigid transformation; (e) aligned using an affine transformation; f) aligned using a B-spline transformation. The overlay grids indicate the deformations imposed on the moving image. Adapted from [10]. 72

3.11 The intensity difference between two frames half a sequence apart **without** pairwise registration (SD = 32.48). 76

3.12 The intensity difference between the two frames **after** pairwise registration (SD = 19.54). 76

3.13 Scaled average fluorescein concentration within field-stop region per frame in the pre-operative intervention sequence against the frame number. 78

3.14 Scaled average fluorescein concentration within field-stop region per frame in the **post**-operative intervention sequence against the frame number. 78

3.15 The global reference in pre-op sequences is framed by a blue box, and the global reference in post-op sequences is framed by a red box. They are both automatically determined from the prior intensity-time course plots. 79

3.16 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 1. 81

3.17 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 2. 81

3.18 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 3. 82

3.19 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 4. 82

3.20 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 5. 83

3.21 Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 6. 83

4.1 Three-dimensional intensity profile (in 256-level gray scale) of a branching retinal microcapillary near macula region, obtained by fluorescein imaging. Here troughs refer to vessel pixels, while peaks refer to background pixels. 86

4.2 Left: Input image of rice grains at courser scale ($l = 2$); Right: Measure of edge likelihood computed using $cos(\Psi^{(l=2)})$. 90

4.3 Tracked retinal vascular centreline (in red) with the boundary pixels in blue, from a sub-section of a retinal angiogram. 94

4.4 Three-dimensional intensity profile (in 256-level gray scale) of a branching retinal microcapillary near macula region with extracted centreline (in blue and green) superimposed on the profile. Here troughs refer to vessel pixels. 94

4.5 Macular vasculature (centreline in magenta) imposed on pre-operative fluorescein angiogram. 95

4.6 Macular vasculature (centreline in magenta) imposed on post-operative fluorescein angiogram; with visible disappearance of microvascualtures (in red) compared to the pre-op centrelines above, possibly occluded by microemboli. 95

5.1 A partial pyramidal cell represented by maximum intensity projection of different portions of the original 3D fluorescence microscopy image stack. At neuron level (A), at the level of dendritic branches and bifurcations (B), at the level of individual spines (C), adapted from [11]. 98

5.2 The normal vectors (in green) in the two red-dashed windows point in opposite directions even though the orientation of content in both windows is the same. . 102

5.3 (a) Synthetic image with concentric circles; (b) the angular representation with interval $[-\pi/2, \pi/2]$ (light, dark) with a discontinuity mod π; (c) a CT cross-section of a tree trunk; (d) the angular representation of the CT slice, adapted from [12]. 103

5.4 Definition of X in polar coordinates. 105

5.5 A 2D orientation half plane in 3D is mapped onto a 2D cone in 5D with opening angle of 120 degrees, adapted from [13]. 105

5.6 An icosahedron (one of the five Platonic polyhedra), adapted from [13]. 108

5.7 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_1}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_1}$. Note, the slice is taken at the Z-X plane at the centre of the volume (Y=33). 111

5.8 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_2}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_2}$. Note, the slice is taken at the Z-X plane at the centre of the volume (Y=33). 111

5.9 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_3}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_3}$. Note, the slice is taken at the Y-X plane at the centre of the volume (Z=33). 112

5.10 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_4}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_4}$. Note, the slice is taken at the Y-X plane at the centre of the volume (Z=33). 112

5.11 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_5}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_5}$. Note, the slice is taken at the Z-Y plane at the centre of the volume (X=33). 113

5.12 A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_6}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_6}$. Note, the slice is taken at the Z-Y plane at the centre of the volume (X=33). 113

5.13 Directional filter set response on test spheres. Left: The surface visualisation of three test spheres with varying radius, hollow interior and Gaussian blurred surface in a $256 \times 256 \times 256$ test volume. Right: Directional filter set response (surface representation with maximum filter response) on the down-sampled (by a factor of 2) test volume for faster computation. The 6 different colours represent 6 different directional filters. 114

5.14 Segmented blood vessels from rat cerebellar cortex microscopic tissue stack using intensity-only volume patch for training. 115

5.15 Iso-surface plots of spatial auto-correlation functions and corresponding energy distributions in the Fourier domain. Adapted from [14]. 117

5.16 Segmented blood vessels from rat cerebellar cortex microscopic tissue stack using patch consist of 3 eigenvalues from GST for training. 118

5.17 A dissection (z=96) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity). 118

5.18 A dissection (z=135) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity). 119

5.19 A dissection (z=75) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity). 119

6.1 The segmentation results (right) of our **PP-MLP** algorithm on Test Image 1 (left). 124

6.2 The segmentation results (right) of our **PP-MLP** algorithm on Test Image 20 (left). 124

6.3 20 Test images from the DRIVE Database. 125

6.4 The segmentation results (posterior probability map, or "soft decision") of our **PP-MLP** algorithm on all the test images. 126

6.5 The ROC curves of our proposed "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method evaluated on the DRIVE database with different training patch size, incrementing from less than 1% to 3.6% of the total image pixels. 127

6.6 The ROC curves for 20 test images using one MLP model. 128

6.7 The ROC curves for lower and upper bound, and for the entire test data. . . . 128

6.8 The ROC curves of our proposed PP-MLP method evaluated on the STARE database with different training patch size, incrementing from less than 0.5% to around 2% of the total image pixels. The performance of Hoover threshold probing approach is superimposed in dotted black line. 129

6.9 Temporal average $f_{average}$ of registered sequences (pre- and post- coronary artery bypass graft surgery) for each patient (a total of 6 patients). 132

6.10 Same sub-region in all 66 frames for the inter-sequentially registered pre- **and** post-operative sequences of one patient, all with extracted frame-specific putative microvascular centreline estimates. 135

6.11 The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=12}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 12_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=12}(x, y)$ are highlighted by red diamond shapes. 135

6.12 The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=31}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 31_{st} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=31}(x, y)$ are highlighted by red diamond shapes. 136

6.13 The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=49}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 49_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=49}(x, y)$ are highlighted by red diamond shapes. 136

6.14 The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=60}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 60_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=60}(x, y)$ are highlighted by red diamond shapes. 137

6.15 Sample neurons from the NeuroMorpho database [7] based on principal cell types: (A) ganglion cell (mouse retina), (B) granule cell (rat hippocampus), (C) motoneuron (cat spinal chord), (D) Purkinje cell (mouse cerebellum), (E) pyramidal cell (rat cerebral cortex), (F) stellate cell (rat cerebral cortex), (G) tangential cell (blowfly visual lobe), (H) relay cell or interneuron (mouse dorsal thalamus). 137

6.16 Cylindrical mesh-based surface fitting between nodes (represented by spherical structures) along dendrites. 139

6.17 A reconstructed rat cerebellar Purkinje neuron morphology. 140

6.18 A montage view (along z-axis) of a sample surrogate stack ($213 \times 149 \times 35$). Reconstructed Purkinje neuron (stack) with Gaussian white noise ($Var = 0.1$) added. 142

6.19 A montage view (along z-axis) of a sample surrogate stack ($213 \times 149 \times 35$). Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron (stack). . 143

6.20 A 180° rotated view of the maximum intensity projection of the sample surrogate stack, reconstructed Purkinje neuron (stack) added with Gaussian white noise ($Var = 0.1$). 144

6.21 A 180° rotated view of the maximum intensity projection of the sample surrogate stack, Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron (stack). 144

6.22 Left: A 3D view of the test stack where Gaussian white noise $Var = 0.1$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 145

6.23 Left: A 3D view of the test stack where Gaussian white noise $Var = 0.2$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 145

6.24 Left: A 3D view of the test stack where Gaussian white noise $Var = 0.3$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 145

6.25 Left: A 3D view of the test stack where Gaussian white noise $Var = 0.4$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 146

6.26 Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.1$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 146

6.27 Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.2$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 146

6.28 Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.3$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 147

6.29 Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.4$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch. 147

6.30 Four rotated 3D view of segmented neural cellular structures (in green) from the experimental rat cerebellar cortex microscopic tissue stack using our "PP-MLP" method. 149

6.31 A montage view of the segmented neural cellular structures from the experimental rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method, only displaying every other slice along the Z axis. 150

6.32 Four rotated 3D view of segmented neural cellular structures (in blue) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch. 151

6.33 A montage view of the segmented neural cellular structures from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch, only displaying every other slice along the Z axis. 152

6.34 A 3D view of a sample rat cerebellar cortex microscopic tissue stack ($512 \times 512 \times 223$ at 8-bit). 153

6.35 Four rotated 3D view of segmented neural cellular structures (in green) from the experimental rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method. 154

6.36 Four rotated 3D view of the segmented Pukinje cells (two in total) from rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method. 155

6.37 Four rotated 3D view of segmented neural cellular structures (in blue) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch. 156

6.38 Four rotated 3D view of segmented neural cellular structures (in aqua) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on GST-based 3D neighbourhood feature patch. 157

Chapter 1

Introduction

1.1 Theme

Any attempt to investigate and measure the *in vivo* biomedical structures faces the following technical challenges: image registration, structural segmentation and extensive validation. Over the years, the rapid improvement in imaging techniques and equipment has in turn driven the demand for more automation in handling large biomedical data sets. A wealth of mathematical models has been introduced as optimal solutions for their designated data sources of varying resolutions, modalities and contrast (staining) mechanisms. Yet there is a shortage of unified frameworks that are somewhat transferable among multi-disciplinary data sources.

The first theme for this research is to carry out **adaptive** investigation with **resolution specific** to each image modality, spanning across standard 2D retinal funduscopy, non-standard 2D+t sequential retinal fluorescein angiography and 3D two-photon microscopic cerebellar cortex neural image stacks. The impact of increasing open source sharing of original biomedical data inevitably steer the demands for image processing away from conventional refinement and optimization on existing imaging models and techniques on limited amount of data sources. Rather, the principles and techniques that are more "transferable" and "dimensional-scalable" towards a variety of multi-modal data are gaining

more momentum. We aim to provide, in this thesis, one of such approach that can be easily scaled up to spatial 3-dimensional data volume from the concept proved viable when applying to 2-dimensional data of a different nature.

The second theme for this research is to offer a novel process that can **differentiate and link elongated, fine, vessel-like structures that contain intermittent gaps**. Our pixel-level segmentation scheme involves: i) constructing a phase-invariant orientation field of the local spatial neighbourhood; ii) combining local feature maps with intensity-based measures in a structural patch context; iii) using a complex decision process to interpret the combination of all the elements in the patch in order to reach a classification decision. This local-to-global structural "feature" interpretation has the advantage of pixel-level decision making for both segmentation and linkage, without processing information at the global scale, thus leads to less computation on the redundant macro structures of little research interest. Figure 1.1 illustrates the principles behind conventional edge-linking post segmentation and our approach.

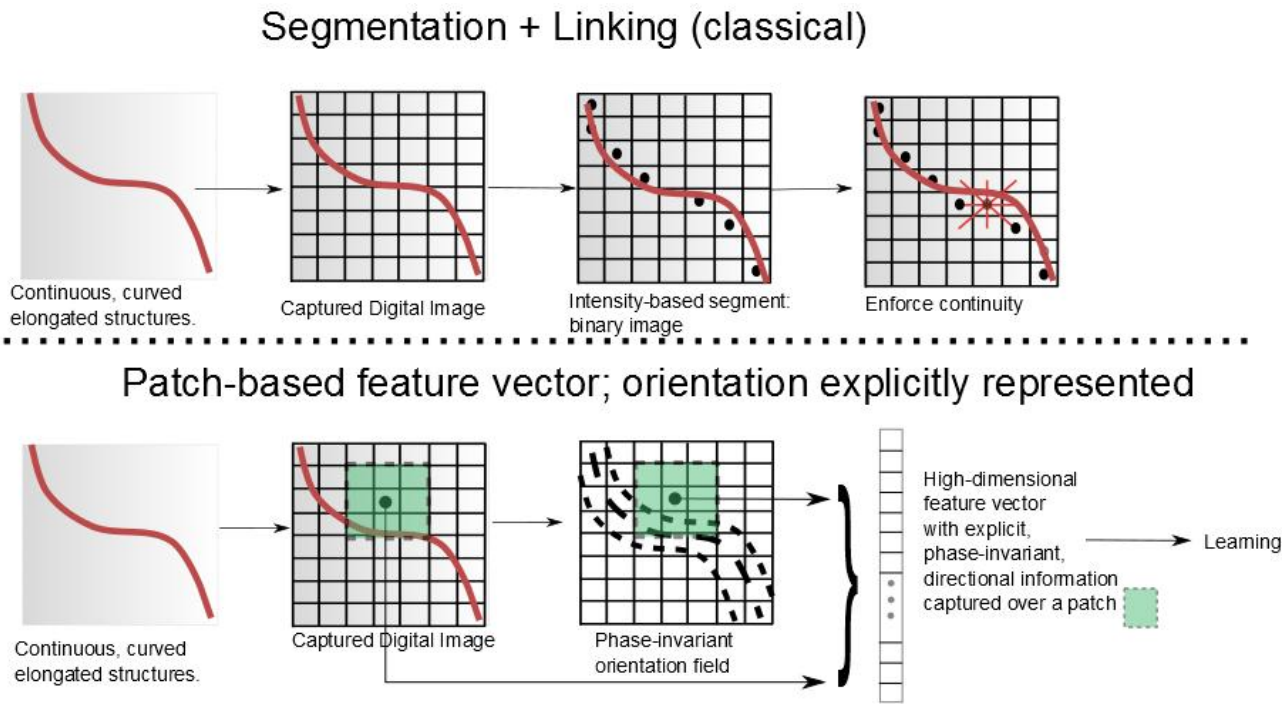


Figure 1.1: Two contrasting approaches. Top: Segmentation is followed by explicit edge linking; Bottom: The decision process at individual pixels utilises explicit information on the local phase-invariant orientation **field**.

1.2 Background of Image Modalities

1.2.1 Retinal Imaging

Ocular Haemodynamics with Associated Diseases

Attempts to study the retinal vascular system date back to the origins of ophthalmoscopy in 19th century [15] [16] [17]. As the retina is highly metabolically active with a double blood supply, the retina allows direct non-invasive investigation of disease-related circulatory complicated by the presence of diabetes, hypertension and other cardiovascular diseases. The most recent overviews [18] [19] on the retinal manifestations of eye and system-related disease listed 5 major categories:

1. Diabetes: *Diabetes mellitus* causess hyperglycemia (presence of elevated blood glucose), which in turn damages blood vessels, small or larger, and nerve cells, and thereby damages the kidneys, heart, brain and eyes. Treatment is primarily through diet changes, administration of insulin with/or anti-hyperglycemic drugs.

2. Diabetic Retinopathy: Diabetic retinopathy (DR) is a complication of *Diabetes mellitus* and one of the prevalent causes of blindness and visual loss in the world [16]. In the eye, hyperglycemia damages the retinal vessel walls, which can lead to ischemia, resulting in the growth of new blood vessels which may bleed and cause retinal detachment ("proliferative diabetic retinopathy"). Severe consequence is the breakdown of the blood-retinal barrier leading to fluid leakage, Diabetic Macular Edema (DME) and damage to photoreceptors, which often result in vision loss. The breakdown of the blood-retinal barrier causes leakage of dilated hyperpermeable capillaries and microaneurysms into intra- and extra-cellular retinal tissue with fluid accumulation. The leaking exudate in DME initially enters the cytoplasm of Müllers cells (radial glial cells of the retina), preferentially in the outer retina, though fluid accumulation has been found to extend through most macular layers in more advanced stages of DME. Retinal extracellular fluid ("cysts") occur predominantly in the outer retina. Over time,

cysts tend to fuse and extend from the outer into the inner retina. In these cases, atrophy or apoptosis of the remaining retinal tissue occurs. Hard exudates can occur and tend to be located at the level of the outer plexiform layer. Patients with longstanding DME with impaired visual acuity show decreased directional sensitivity of photoreceptors and decreased visual pigment density. The management of DR includes laser photocoagulation and administration of anti-vascular growth factors.

3. Age-Related Macular Degeneration: Age-related macular degeneration (AMD) has two major forms, dry and wet. Dry AMD typically leads to gradual loss of visual acuity. Wet AMD, also known as "choroidal neovascularization" (CNV), is more visually threatening. It is characterized by ingrowth of a choroidal vascular structure into the macula accompanied by increased vascular permeability. The increase in vascular permeability leads to abnormal fluid collection within or below the retina that causes visual dysfunction when it involves the center of the macula. The natural course of CNV is rapidly deteriorating acuity, scarring of the pigment epithelium, and permanent visual loss or blindness. Progression of dry AMD can be slowed in many patients through dietary supplements, while visual loss from wet AMD is treated with intravitreal administration of anti-vascular growth factor.

4. Glaucoma: Glaucoma is characterized by gradual damage to the optic nerve and resultant visual field loss. Early diagnosis and optimal treatment have been shown to minimize the risk of visual loss due to glaucoma. Glaucoma is primarily a neuropathy, not a retinopathy, and acts on the retina by damaging ganglion cells and their axons. The hallmark of glaucoma is cupping of the optic disc, which is the visible manifestation of the optic nerve head (ONH) 3-D structure. The optic disc can be imaged two-dimensionally either through indirect stereo biomicroscopy or with stereo color fundus photography. The ratio of the optic disc cup and neuroretinal rim surface areas in these images, called cup-to-disc ratio, is an important structural indicator for assessing the presence and progression of glaucoma. Glaucoma is typically treated with ocular pressure lowering drops, and in refractory cases through surgery.

5. Cardiovascular Disease: Hypertension and atherosclerosis can change the wall thickness of retinal arteries and veins. A thinning of the arteries and widening of the veins, is often associated with an increased risk of stroke and myocardial infarction. Hypertension can also invoke direct retinal ischemia, which causes retinal infarcts visible as cotton wool spots and choroidal infarcts visible as deep retinal white spots. In addition, systemic vascular disease can cause arterial and venous occlusions, known as central and branch arterial occlusions and central and branch venous occlusions.

Over the years, clinical and research assessment techniques have evolved from the physical description of the retinal arteries and veins to the quantitative measurements of the associated ocular haemodynamics parameters [20] [21] [22]. These quantitative parameters, describing the characteristics of retinal vasculature or ocular blood perfusion, are now the key indicators for clinical disease diagnosis. [23] suggested that hypertension reduces the average vessel diameter; [24] suggested that hypertension increases the arteriolar length to diameter ratio; [25] found that diabetes increases venous diameters and reduces blood velocity; [26] reported diabetes increases clinically perceived leukocyte velocity; [27] suggested that diabetes prolongs arteriovenous passage time and reduces capillary blood flow velocities; [28] showed that glaucoma lowers pulsatile ocular blood flow.

Standard 2D Retinal Funduscopy

Imaging Principle Funduscopy, or fundus photography is the photography of the interior surface of the eye, including the retina, optic disc, macula, and posterior pole [16]. It is often used by optometrists, ophthalmologists, and trained medical professionals for monitoring progression of a disease, diagnosis of a disease (combined with retinal angiography), or in screening programs, where the photos can be analysed later. It has the advantage of availability for examination at different geographical and temporal location as well as documentation of the medical information for future reference. Thus, fundus photography is widely used for population-based, large scale detection of diabetic retinopathy, glaucoma, and age-related macular degeneration.

The fundus camera consists of a specialized low power microscope with an attached digital camera, see Figure 1.2.



Figure 1.2: A Topcon TRC NW6S digital non-mydriatic retinal fundus camera [1].

The optical design of fundus cameras is based on the principle of monocular indirect ophthalmoscopy [16]. A fundus camera provides an upright magnified view of the fundus with the typical camera views ranging between 30 to 50 degrees of retinal area. The optics of a fundus camera are similar to those of an indirect ophthalmoscope as the observation and illumination systems follow dissimilar paths. The observation light is focused via a series of lenses through a doughnut shaped aperture, which then passes through a central aperture to form an annulus, before passing through the camera objective lens and through the cornea onto the retina. The light reflected from the retina passes through the un-illuminated hole in the doughnut formed by the illumination system. As the light paths of the two systems are independent, there are minimal reflections of the light source captured in the formed image. The image forming rays continue towards the low powered telescopic eyepiece. When the button is pressed to take a picture, a mirror interrupts the path of the illumination system allowing the light from the flash bulb to pass into the eye. Simultaneously, a mirror falls in

front of the observation telescope, which redirects the light onto the capturing medium, whether it is film or a digital CCD. Because of the eyes tendency to accommodate while looking though a telescope, it is important that the exiting vergence is parallel in order for an in focus image to be formed on the capturing medium.

DRIVE database The fundus images from the DRIVE (Digital Retinal Images for Vessel Extraction) database [2] [11] were obtained from a diabetic retinopathy screening program in The Netherlands. The screening population consisted of 453 subjects between 31 to 86 years of age. Each image was JPEG encoded, the common practice in screening programs. Of the total 40 images in the database, 7 contain pathology, such as exudates, hemorrhages and pigment epithelium changes. The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45 degree field of view (FOV). Each image was captured using 8 bits per color plane at 768×584 pixels. The FOV of each image was circular with a diameter of approximately 540 pixels. The images were cropped around the FOV, and a mask image was provided that delineates the FOV.

The set of 40 images was divided into a test and a training set, both containing 20 images. Three observers, instructed and trained by an experienced ophthalmologist, manually segmented all the pixels on the basis that they were at least 70% certain that these pixels belonged to the retinal vasculatures. For the training set, a single manual segmentation was carried out between two observers with one segmenting 14 images and the other segmenting the remaining 6 images. For the test set, two manual segmentations were carried out. First segmentation was shared between the prior two observers, and was used as gold standard. They marked 577,649 pixels as vessel and 3,960,494 as background (12.7% vessel). The second segmentation was completed by a third observer alone, and was used to compare computer generated segmentations with those of an independent manual segmentation. This segmentation resulted in 556,532 pixels marked as vessel and 3,981,611 as background (12.3% vessel).



Figure 1.3: An example retinal fundus image from the DRIVE database in original RGB channels [2].



Figure 1.4: An example retinal fundus image from the DRIVE database, converted to gray-scale for processing.

STARE database The fundus images (see Figure 1.5 and 1.6) from the "STructured Analysis of the Retina" database [3] were selected from method described as follows:

Twenty retinal fundus images were captured by a TopCon TRV-50 fundus camera at 35 field of view. Each slide was digitized to produce a 605×700 pixel image, 24 bits per pixel (standard RGB). Ten of the images were of healthy individuals with no known pathology. Ten of the images contained pathologies thus exhibited non-standard blood vessel appearance in varying portions of the image. This selection was made to develop successful segmentation methods with certain degrees of confidence even in the presence of abnormal vessel appearances for potential clinical applications.

Each of these 20 images was carefully labelled by hand, to produce a ground truth vessels segmentation. The tool used for hand labelling was adapted from the tool described in prior literature [29], which allowed the user to magnify the image to a level for appropriate labelling of individual pixels one at a time. The process of labelling an image took several hours, depending on the user and image. The classification of a majority of the pixels was often clear to a human observer. However, some of the pixels, such as those on the boundary of a vessel or on small vessels or near sites of pathology were less easily labelled. To estimate this variance in observation, a second person produced an additional set of hand labelling for the 20 test images. This second labelling was used to establish a reference for performance comparison. On average, the first person labelled 32200 pixels in each image as vessel, while the second person labelled 46100 pixels in each image as vessel. Subsequent review indicated that the first person took a more conservative view of the boundaries of vessels and in the identification of small vessels than the second person. Both labelling for one of the normal images are shown in Figure 1.7.

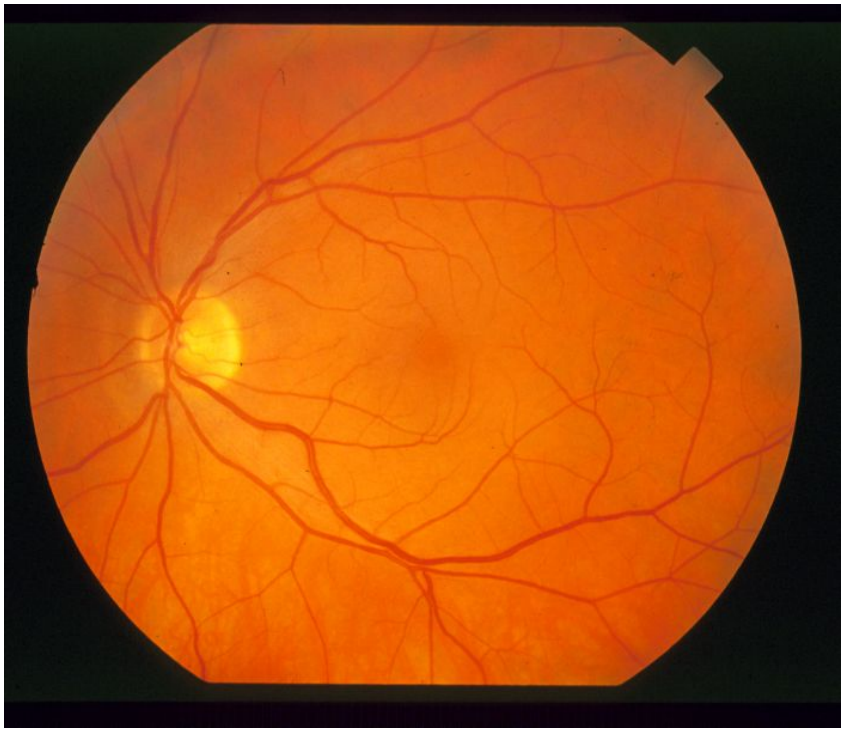


Figure 1.5: An example retinal fundus image from an individual exhibiting normal vessels from the STARE database [3].



Figure 1.6: An example retinal fundus image from a patient exhibiting abnormal vessels from the STARE database [3].

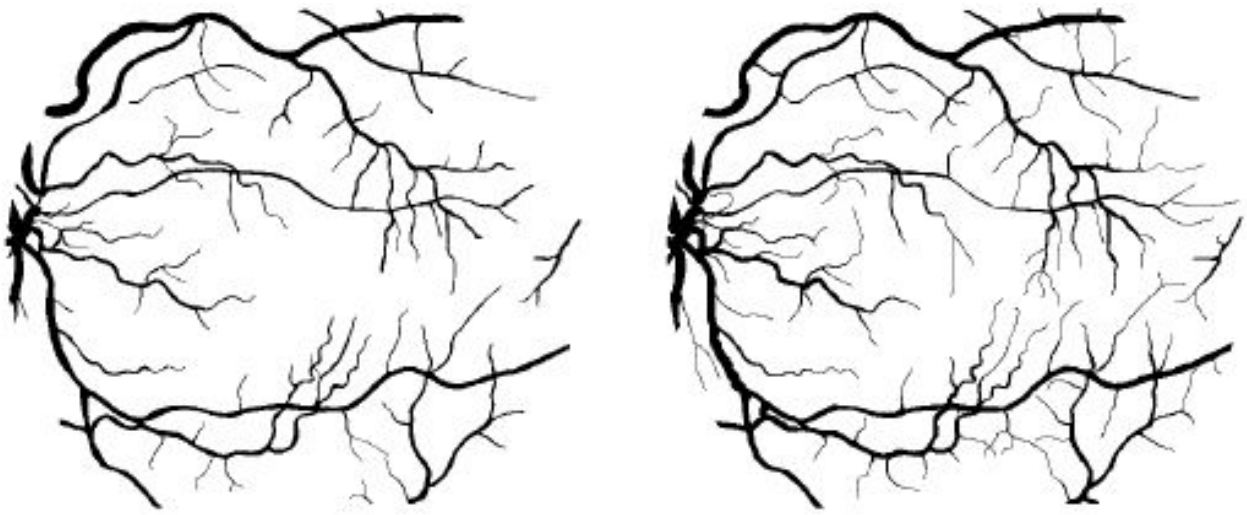


Figure 1.7: An example hand-labelled vessels segmentation from two users. Left label was considered as more conservative as there were fewer pixels representing the vessels in the binary image than the right label [3].

Non-standard 2D+t Sequential Fluorescein Angiograms

Imaging Principle Since its inception [30], fluorescein angiography has become a well-established technique for clinical assessment of the retina. It is widely used in the diagnosis and management of patients with diabetic retinopathy, macular degeneration, and inflammatory retinal diseases. The fluorescein dye concentration in the retinal vessels is quantified by the distribution and the variation of the fluorescence intensity. The passage of fluorescein dye through the retina thus reflects both the vessel structure and the retinal blood flow velocity. A fundus camera continuously photographs the retina from the onset of dye injection over a period of 3 to 5 minutes. This captures filling and the subsequent elimination of dye in the retinal vessels. Determined by capture time from point injection and filling of the vessels, the angiogram sequences are roughly divided into arterial phase, arteriovenous phase, venous phase and recirculation phase [31]:

- Arterial phase: The retinal arteries are filled (typically seconds after the choroidal flush).
- Arteriovenous phase: The retinal capillaries achieve complete filling and the retinal veins begin to exhibit laminar flow.

- Venous phase: Large retinal veins achieve complete filling and maximum vessel fluorescence occurs.
- Late phase (Recirculation phase): The veins and the arteries retain approximate equal brightness. The intensity of the fluorescence diminishes slowly, demonstrating gradual elimination of dye from the retinal and the choroidal vasculature.

Conventionally, 20 minutes before photography, the patient's pupil is dilated to about 7 mm with a couple drops of 1% tropicamide. This is then followed by injection into (commonly) the antecubital vein with 5 ml of 10% sodium fluorescein dye. A fundus camera built-in with appropriate filters then sequentially photograph the retina thereby capturing the passage of the dye through retinal and choroidal vessels. Frames are captured continuously during the initial phase of the angiogram at around one frame per second for the first 20 seconds. This rate is reduced to around a frame every 5 or 10 seconds up to a minute after injection and finally a few late frames are obtained at around 3 and 5 min. An excitation filter ensures illumination by light within a certain band of wavelengths 465-490 nm to excite the fluorescein dye, and a barrier filter restricts emission of fluorescein to wavelengths in the range of 520-530 nm.

A novel hemodynamic measure of ocular microcirculation For early stage detection, however, subtle changes in the retinal microvasculature requires sufficiently fine-resolution imaging and dependable image analysis to diagnose vessels at the micrometer scale. Standard parameters such as "AVP time", "mean dye velocity", and "time to maximum image" [20] [31] [27] were developed to investigate the vascular network in localized regions near the optical disk, but none of these analyses capture individual microvascular vessels.

Yet, it is known that blood clots occur during surgery; such clots can impact on the outcome of surgery and are thought to contribute to the development of post operative blood clots such as deep vein thrombosis and pulmonary emboli. Orthopaedic operations, particularly total knee replacement surgery, is associated with thrombolytic events and it has been reported that the risk of deep vein thrombosis ranges from 31%–82%, and the risk of fatal

pulmonary emboli (PE) from 0–0.4% [32]. This has led to a number of pre-operative and post-operative strategies including early mobilisation, compressive stockings, external pneumatic compression and chemical prophylaxis [33] [34]. However, despite this, a number of patients still go on to suffer from deep vein blood clots. This is thought to be related to the surgery and the unavoidable release of fat and bone marrow (emboli) into the vascular system which can initiate the clotting cascade [35].

If clinicians or surgeons can identify people during surgery that have a more aggressive reaction in the form of a large volume of microemboli (very small blood clots), more aggressive post operative management to avoid these small blood clots from aggregating to become large blood clots could be taken. This treatment has to be managed carefully because the administration of anti-clotting agents post-surgery carries serious risks.

Moreover, there has been speculation that changes in cognitive function observed in cardiothoracic patients post by-pass surgery [36] could be due to the migration of emboli towards the brain. Again, monitoring of microvasculature, particularly, in the retina may be a good predictor of potential cognitive impairment or even morbidity. Preventative measures may therefore well improve the patient's quality of life.

[37] first suggested that comparison of pre- and post- operative retinal fluorescein angiograms might indicate the existence of microemboli that could be associated with cognitive impairment or even morbidity. As the presence of these microemboli may be manifested as slower blood perfusion or even disappearance of vessels and/or vessel branches (micro vascular occlusions). This opens up the possibility of detecting with greater sensitivity the existence of microemboli that could be associated with cognitive impairment. However, only the macula region of one pre-op and one post-op image were used. The analysis of sequential retinal angiograms has not been widely exploited; the dynamics of blood flow, the inter-individual variability of retinal vascular patterns, and the temporal diffusion of the dye are either difficult to reliably compensate for or computationally too demanding. However, with retinal arterioles typically ranging from approximately 20 to 200μm in diameter [20], fluorescein angiography allows visualization of the microvasculature down to 10μm in diameter, which is

not yet achievable by either funduscopy or transcranial Doppler ultrasound (TCD) [38]. By recruiting all frames in both sequences, it is possible to study retinal microvasculature dynamics and even identify small, but potentially significant embolic events.

Fluorescein Angiograms Sequences by Arnold (FASA) database The retinal angiogram sequences we employed in our analysis were collected in a standardized clinical protocol (both anaesthetic and surgical) before and after coronary artery bypass graft surgery. A modification of the procedure allowed a larger number of image frames to be obtained within a shorter time frame, thereby minimizing the dye usage and time taken to scan the patient.

The right pupil of the patient was first dilated with 1% tropicamide eye drops, and then exposed by an eyelid retractor. The cornea was kept moist with balanced saline solution applied every 10 seconds. A modified and vertically mounted Zeiss (Oberkochen) 30 Fundus camera was moved to the head of the table. When the camera was positioned and focused on the macula, 5ml of 20% sodium fluorescein dye was injected as a rapid bolus into an intravenous cannula sited prior to the surgical procedure. As the dye took 10-15 seconds to circulate to allow the optimum images to be captured, the sequential exposures were first made sparsely over the initial period of 15 seconds post dye injection, and then with increasing yet uneven intervals (maximum rate at 0.7 second per frame) over a period of approximately 50 seconds. The eyelid retractor was then removed and the camera moved away from the operating table. The whole procedure took 1-2 minutes.

Procedures were carried out under the ethics guidelines of Hammersmith Hospital. Patient recruitment followed certain exclusion criteria for control purposes: the presence of diabetes mellitus, neurological, psychiatric, ocular, or carotid disease, or cardiac failure. Furthermore, pre- and post- operative ophthalmic studies (8 days before operation and 8 weeks following surgery) were identical and consist of visual acuity (Snellen chart), visual fields (Friedmann analyser) and bilateral colour fundus photography.

The pre- and post-operative angiograms correspond to an area of a 30 degrees field of the

retina ($31mm^2$) centred on the fovea. Each individual frame (see Figure 1.8) was scanned by a high resolution JVC TV camera and digitized to a 4288-by-2848 array of 8 bit pixels using the GOP-302 image analyzer (Teragon-Context AB). We used in total 12 retinal angiogram sequences from 6 patients with approximately 32 frames per angiogram sequence for image analysis.

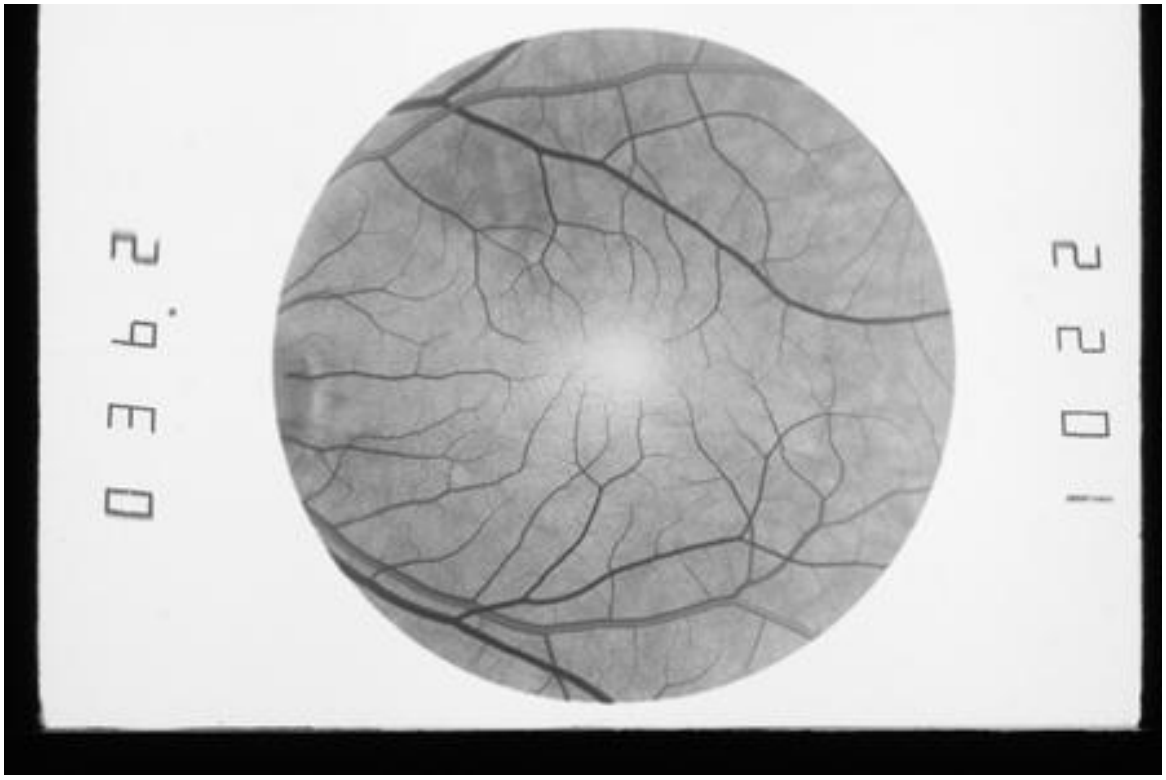


Figure 1.8: A single fluorescein angiogram from a pre-operative intervention sequence in our clinical data set.

1.2.2 Neural Imaging

Imaging and Staining Techniques in Neuroscience

Imaging is highly integrated in neuroscience research as the visualization of neural structures and neuronal functionalities at the cellular level is highly dependent on the tools and techniques chosen. Various molecular and cellular imaging techniques aims to unravel different biochemical activities of various neural cells with their associated molecules at a much higher resolution in space and time when compared with biomedical imaging techniques

such as Magnetic Resonance Imaging (MRI), Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET).

The microscope is now an indispensable tool in neuroscience. Because organelles, glial cells, neurons, and even populations of neurons cannot be seen by the naked eye, a microscope is essential to examine the nervous system at the cellular level. A **Light** microscope uses visible light to illuminate and image a specimen, thus can magnify the specimen up by a factor of 1000. *Brightfield* microscopy is a common form where light directly passes through the specimen. providing access to the structure of a cell and its local environment. *Phase-contrast* microscopy, on the other hand, amplifies variations in cell density by interpreting difference in refraction indices from structural change. *Darkfield* microscopy uses oblique illuminating lights to enhance the contrast in organelles that are highly refractive. *Differential interface contrast (DIC) or Nomarski* microscopy provides optical modifications on light-scattering properties of cellular structures. Figure 1.9 illustrates the different appearances of the same cell under the four types of light microscopes described above.

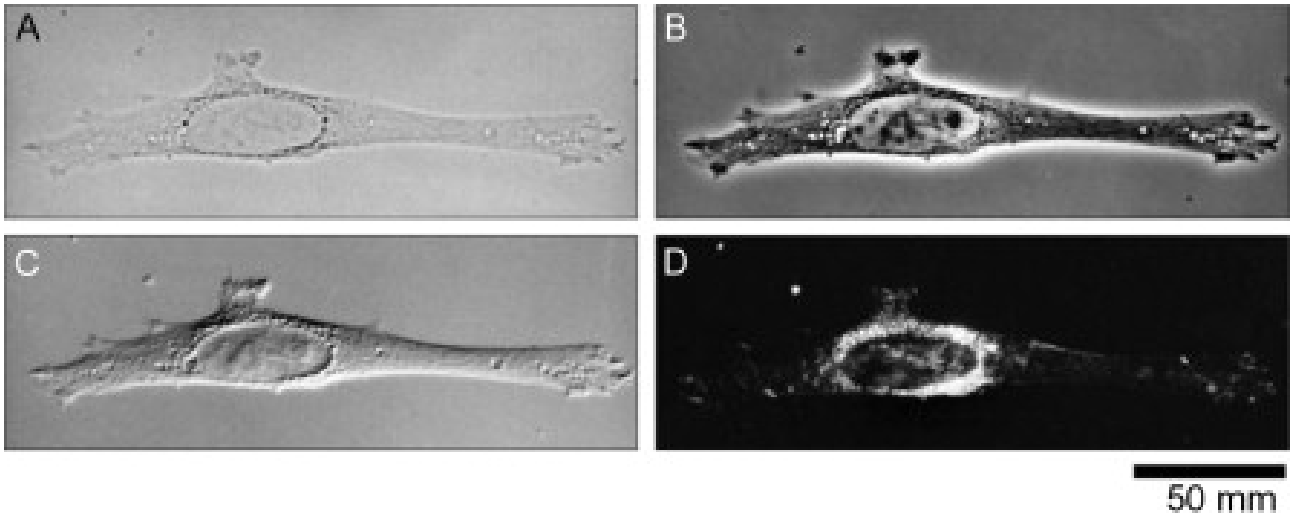


Figure 1.9: Four types of light microscopy analyzing the same cell. (A) Brightfield microscopy; (B) Phase-contrast microscopy; (C) DIC or Nomarski microscopy; (D) Darkfield microscopy [4].

Fluorescence microscope examines specimens when they are labelled with fluorescent dye or expressing genetically modified fluorescent proteins. For *Epifluorescence* microscopy (excitation and observation of the fluorescence are from above the specimen), the illuminating

light is first passed through an "excitation filter" that only allows the "excitation" wavelength through to the specimen. The wavelength-limited light excites the fluorescent agents ("fluorophore") in the specimen and the light emanates at a longer "emission" wavelength. This re-emitted light is then collected by the objective lens and passed through an "emission filter" prior to being picked up by the detector. This process allows a wide range of fluorophores to be simultaneously present at the same specimen as long as each fluorophore possess unique excitation and emission spectra. In *Confocal* microscopy, light is focused from a pinhole aperture onto a single point on the specimen. Fluorescent light emanating from the specimen is transmitted through a second pinhole aperture before hitting a detector. The two apertures are confocal with each other, so out-of-focus fluorescence is avoided by the detector pinhole. As only light produced by fluorescence very close to the focal plane can be detected, the image's optical resolution, particularly in the sample depth direction, is much better than epifluorescent microscopes. However, as much of the light from sample fluorescence is blocked at the pinhole, this increased resolution is at the cost of decreased signal intensity, hence long exposures are often required. *Two-photon* microscopy, or *Two-photon laser scanning* microscopy (TPLSM) is a further refinement of precision as a fluorophore can be excited by effectively absorbing two photons at the same time. Pulsed laser light illuminates the sample with light twice the excitation wavelength. When focused on a specific plane of section, two photons can arrive near-simultaneously at one fluorophore in that plane and deliver enough energy to excite the fluorophore. Above and below the plane in focus, light is too low-energy to excite fluorescence, eliminating out-of-focus fluorescence. TPLSM can penetrate deeper into tissue than epifluorescence and confocal microscopy, in the range of $500\mu m$–1mm [39]. It is thus particularly useful in investigating brain tissues from living specimen or fine-scale neural structures. Figure 1.10 compares the principles of epifluorescence, confocal and two-photon microscopy.

Electron microscope (EM) can increase the resolving power to the scale of 0.2nm [5], about the radius of a glutamate molecule. It is essential for examining neural structures at a molecular level, such as synapses, synaptic vesicles and ion channels. The major limitation of EM is that specimens must be fixed, dehydrated and chemically treated, thus EM cannot be

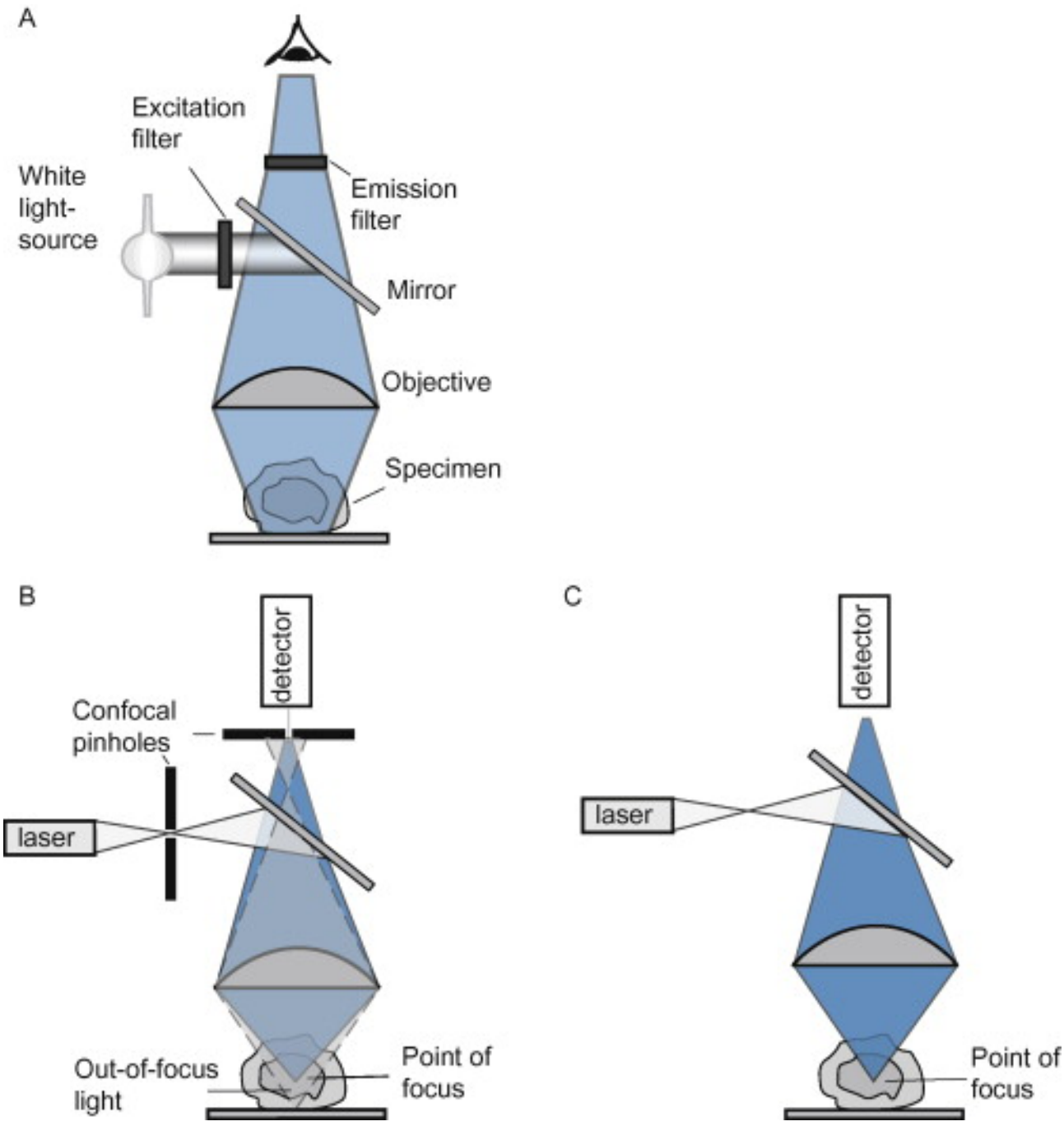


Figure 1.10: Comparison of different set-ups of fluorescence microscopy: (A) *Epifluorescence* microscopy; (B) *Confocal* microscopy; (C) *Two-photon* microscopy [5].

used to examine living samples. Also the specimen often have to be viewed in vacuum, as the molecules that make up air would scatter the electrons [5]. In *transmission electron* microscopy (TEM), a beam of electrons is aimed through a very thin ($< 100nm$)) section of the specimen underwent chemical treatment for contrast enhancement. The preserved tissues are often stained with heavy metals attracted to cellular components under investigation.

These heavy metal stains are electron-dense, so when the electron beam hits these metal atoms, the electrons are either absorbed or scattered. Electromagnets focus and magnify structures of interest by bending charged electron trajectories, which are picked up by special detector. TEM produces 2D images of thin tissue sections extremely useful to image the internal cell structures, so the technique is frequently used to define synapses as the postsynaptic density appear to be electron-dense when in close contact with presynaptic membranes containing neurotransmitter vesicles [39]. *Scanning electron* microscopy, on the other hand, is often employed to capture the dynamics of a specimen's surface. A high-energy electron beam scans across the surface under investigation, often coated with a thin film of gold or platinum to enhance contrast and improve signal-to-noise ratio [5]. As the beam scans across the specimen, interactions between the surface and the electron beam induce varying electron signals emitted. Low-energy secondary electrons excited in the specimen surface provide information at the surface level, while high-energy backscattered electrons and X-rays are emitted from below the specimen surface provide information on the composition of the interior. Figure 1.11 and 1.12 compare the schematic layout of transmission electron and scanning electron microscopy.

Staining is critical in automated analysis of neural tissues as it provides a "simplification" towards complex network of structural components in neural tissues. By accentuating only specific structures sensitive to dye, it results in simpler, higher contrast images with reduced "noise" or clutter. The choice of a staining method depends on several factors, such as the goal of the experiment, the source of the tissue (such as vertebrate or invertebrate), and the region of the nervous system to be examined [40]. Labels that can be injected into the cell are often preferred for modeling studies because the electrode that is used for dye injection can also be used to obtain electrophysiological data on the same neuron.

Compounds that produce dense products are most often chosen for the analysis of dendritic trees and axons at the light microscope level and for studies that require combined light and electron microscope imaging. Currently, biocytin is the most common **intracellular** label for both vertebrate and invertebrate neurons when a dense product is required [41]. Another useful label is biotinylated dextran amine (BDA) that doest not require intracellular dye

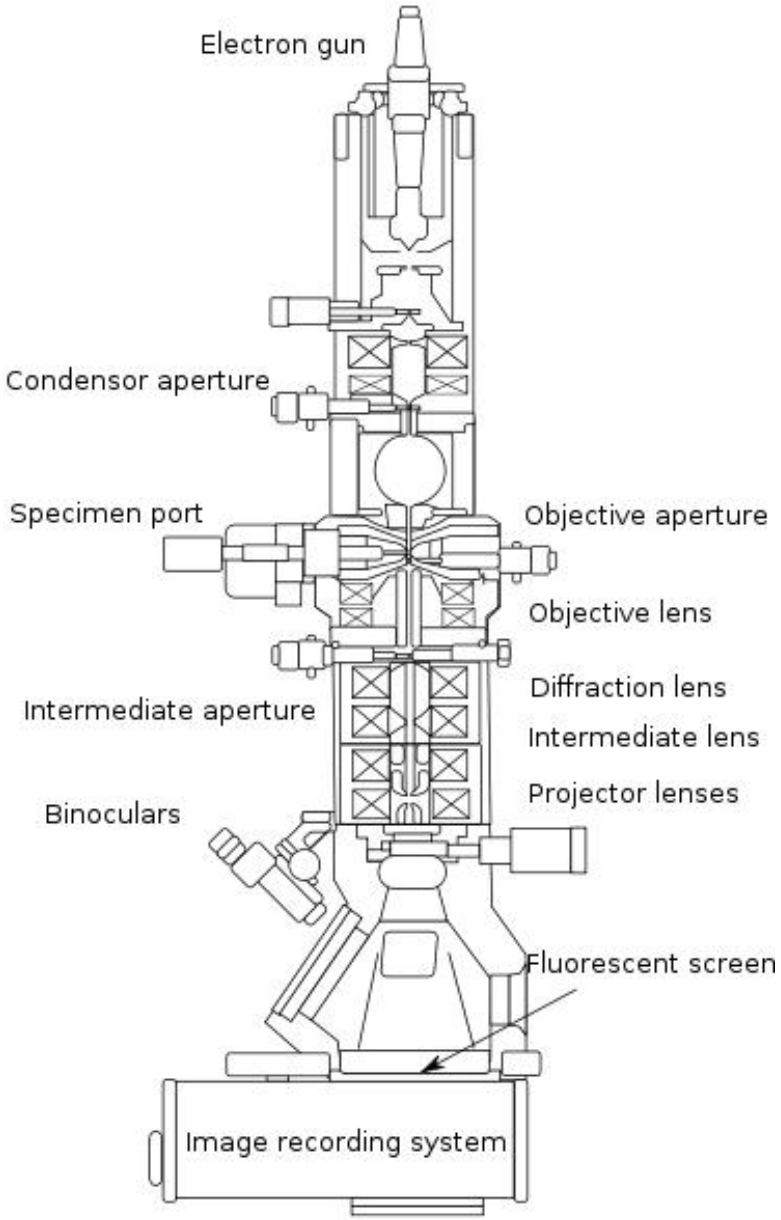


Figure 1.11: The schematic diagram of the layout of a transmission electron microscopy (TEM) [5].

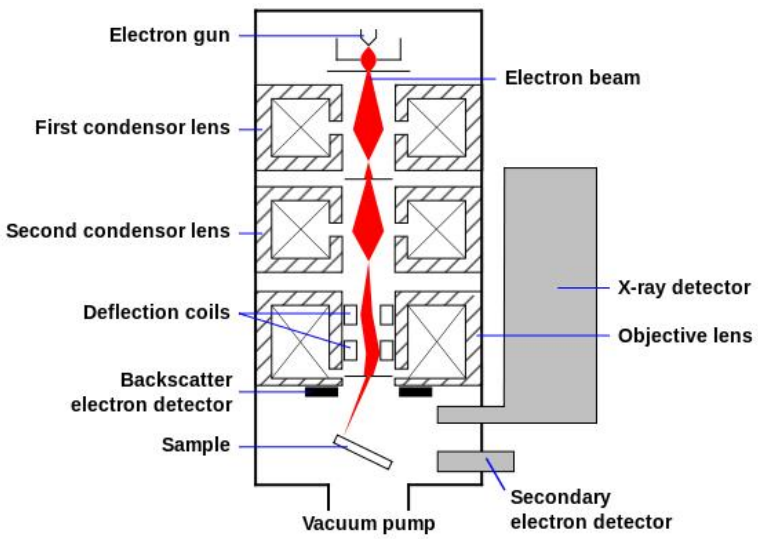


Figure 1.12: The schematic diagram of the layout of a scanning electron microscopy (SEM) [5].

injection. BDA is advantageous for both light and electron microscopy [42].

Fluorescent dyes are preferred for studies that require visualization of the neuron during the labeling procedure or that require the use of a laser-scanning microscope to analyze both neuronal branching patterns and the details of other morphological features such as dendritic spines or axonal boutons. More recently, fluorescent hydrazide and hydroxylamine derivatives have been developed for intracellular staining (Alexa Fluor Dyes); they are photostable, they can be fixed with aldehyde-based fixatives, and they are water soluble.

Fluorometric Ca^{2+} imaging is another sensitive method for monitoring neuronal activity in real-time [43]. In living cells, most depolarizing electrical signals are often associated with Ca^{2+} influx from the activation of voltage-gated Ca^{2+} channels, thus, elementary forms of neuronal communication, such as chemical synaptic transmission, can result in transient intracellular elevations in Ca^{2+} concentration which in turn can be captured by two-photon fluorescence microscopy imaging. One advantage is that it allows simultaneous recordings from many individual cells and has been used successfully in numerous in vitro cellular studies.

In recent years, transgenic mice have been developed that express endogenous green fluorescent protein (GFP) in specific populations of neurons [44]. Varying numbers of neurons express GFP in the different lines. Two lines are particularly useful for quantifying the morphology of hippocampal neurons: the “M” line and the “O” line [44]. In these lines, GFP is expressed under the control of neuronspecific elements from the Thy1 gene, an immunoglobulin gene that is expressed by projection neurons in many parts of the nervous system, as well as in some nonneuronal cells. Deletion of a specific intron abolishes expression in non-neuronal cells. GFP can also be used to label the entire structure of neurons homozygous for specific genes in the intact brain using the mosaic analysis with repressible cell marker (MARCM) technique [45]. Neurons labeled in this way can be reconstructed from stacks of confocal images for morphometric analysis.

3D Two-photon Microscopic Cerebellar Cortex Image Stack

The molecular layer of the cerebellum is a site of great anatomical and synaptic divergence [6]. The **mossy fiber-granule cell pathway** consists of around 100 million granule cells in rat cerebellum [46]. Each granule cell extends into the molecular layer with parallel fiber axons that make hundreds of synapses onto Purkinje cells and inhibitory interneurons. These connections re-converge as around 200,000 parallel fiber synapse with each Purkinje cell dendrite and the estimated 300,000 Purkinje cells pass on neural signals to axons that project out of the cerebellar cortex. Thus, one can conclude that the molecular layer is where the majority of the cerebellar processing take place [6]. Figure 1.13 illustrates the structural diversity in rat cerebellar cortex when labelled with Calcium Green-1 acetoxymethyl (AM) ester [43] [6].

In order to monitor the collective activities of these structures simultaneously in living animals, two-photon microscopy of fluorescent calcium indicators has been demonstrated to be an excellent tool for studying *in vivo* neuronal physiology [43]. Living cells can be loaded in bulk with membrane permeant form of Calcium indicator dyes where the charged carboxylate groups are covered by acetoxymethyl (AM) ester groups, which are then removed by intracellular hydrolases and esterases [43]. This allows simultaneous monitoring of the activity of thousands of individual neurons located up to $200\mu m$ below the surface of a slice [43].

The 3D microscopic neural image stacks we employed in our analysis were were collected by experimental procedures as follows [47]:

The care and experimental manipulation of animals was performed in accordance with the regulations of the United Kingdom Home Office. Sprague Dawley rats (postnatal days 18–29) were anesthetized with urethane (1.2 g/kg) or with a ketamine (50 mg/kg)– xylazine (5 mg/kg) mixture. The level of anesthesia was routinely monitored by observing whisker movements and/or monitoring the hindlimb leg-withdrawal reflex and supplementary doses of anesthetic administered if required. A craniotomy was made to expose folium Crus IIa, the dura was carefully removed, and a stainless steel frame was glued to the skull with dental

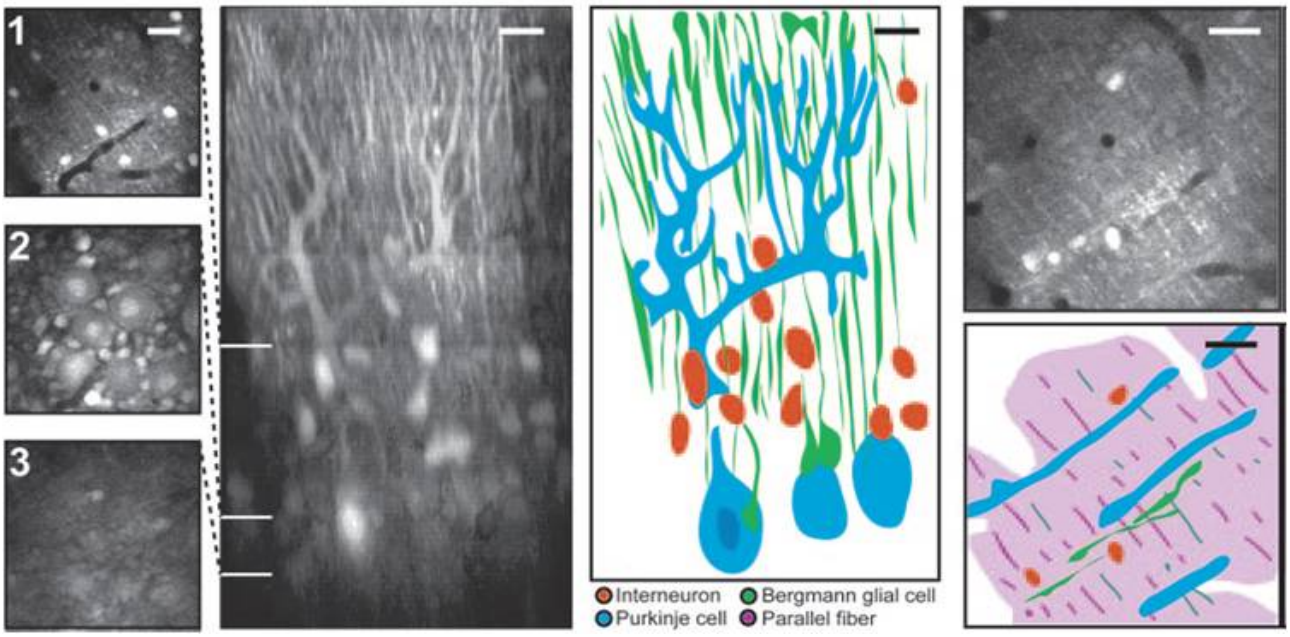


Figure 1.13: (A) *in vivo* two-photon microscopic imaging of rat cerebellum after calcium labeling. Three slices taken at increasing tissue depth to illustrate different structural layers: (A-1) the **molecular layer**, (A-2) the **Purkinje cell layer**, (A-3) the **granule cell layer**. (B) Maximum-intensity projection of 100 slices, each at 2 μm increments, to reconstruct a sagittal view of the labeled rat cerebellar cortex. The reconstruction begins at the brain surface and extends to the granule cell layer at a depth of 200 μm. (C) Illustrative diagram of cellular composition in (B), identifying the main cell types present: **Purkinje cells** in blue, **interneurons** in red, and **Bergmann glia** in green. (D) A zoomed-in image of the molecular layer, similar to (A-1). Top: The slice was taken at 96 μm below the surface; Bottom: The schematic diagram illustrates the main cellular components using the same color coding as in (C). Blood vessels are unstained and appear as dark structures. Note, all scale bars above represent 20 μm. Adapted from [6].

cement. The craniotomy was filled with 1.5–2% agarose in Ringer's solution, and a coverslip was clamped above the agarose to suppress brain movement, leaving an opening for microelectrode access. Body temperature was maintained at 37-38°C using a heating blanket.

Bulk loading of AM ester calcium indicator dye was performed using a protocol similar prior literatures [43] [6]. Briefly, for each experiment, 1 aliquot (50 μg) of Oregon Green BAPTA-1 (OGB-1) AM calcium indicator dye (Invitrogen) was dissolved in DMSO containing 20% w/v Pluronic F-127 (Invitrogen) and diluted 3:100 in a buffer solution containing 150 mm NaCl, 2.5 mm KCl, and 10 mm HEPES, pH 7.4, to give a final dye concentration of 300 μm OGB-1 AM. Alexa 594 was also added to the solution to aid visualization of dye ejection (final concentration, 10 μm). A micropipette was filled with this solution and inserted into the

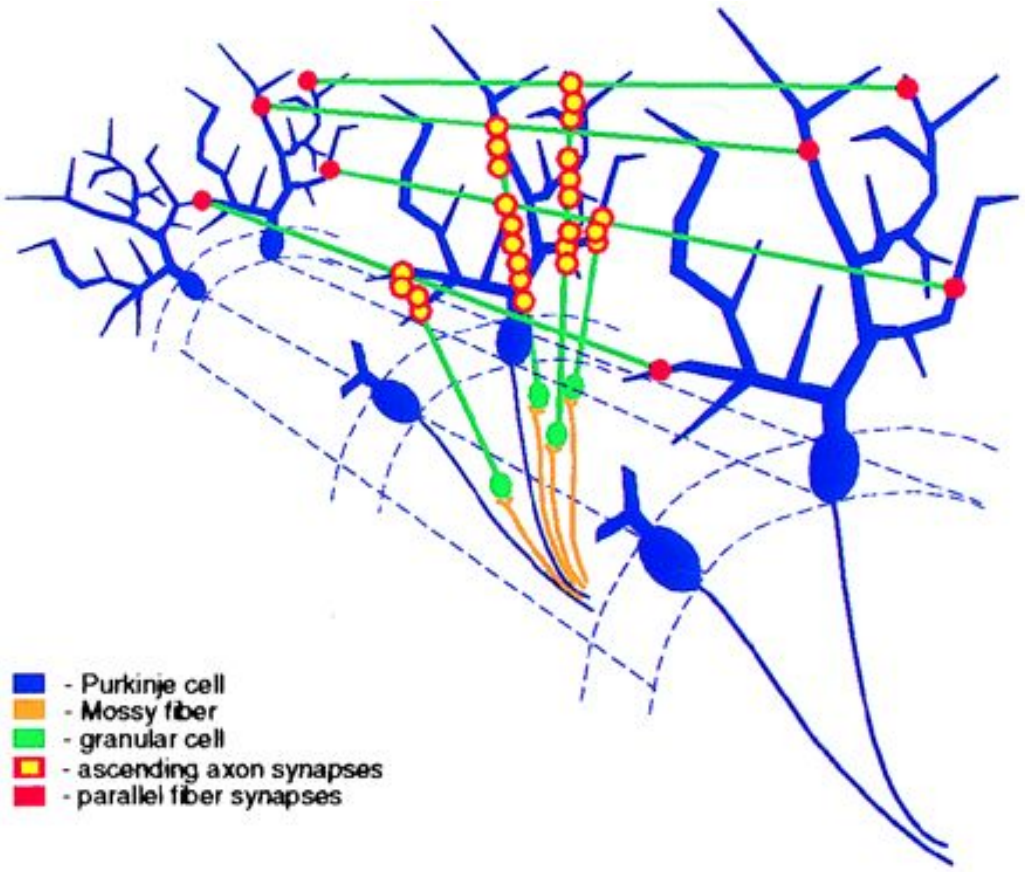


Figure 1.14: A 3D schematic diagram of the interaction among components in the cerebellar cortex with Purkinje cells in blue and granular cells in green. The red-yellow circle marks the synapses formed by the ascending axon on Purkinje cell dendrites, and the red circle marks the synapses formed by the parallel fibers on Purkinje cells along their path. Adapted from [7].

cerebellar molecular layer (depth typically 100–150 μm below pia). Dye was ejected by applying pressure at 40 kPa for $\sim$ 1 min using a Picospritzer (General Valve) or by repeated brief puffs with a 5 ml syringe.

Imaging was performed $>$ 30 min after dye ejection, using a two-photon laser scanning microscope (Prairie Technologies). Excitation was provided by a pulsed titanium/sapphire laser system operating at 810 nm (Mai Tai; SpectraPhysics) with $<$ 100 fs pulse width and 80 MHz repetition rate, focused using a 40$\times$, 0.8 numerical aperture objective lens (Olympus). Images were acquired using ScanImage for MATLAB. For every region imaged, a high-resolution reference image was first acquired (512 $\times$ 512, average of five frames), followed by a spontaneous activity run (typically 256 $\times$ 64 or 256 $\times$ 32 resolution, 1000–2000 frames, 128s of data) and, if appropriate, a series of one to five sensory-evoked runs (typically

500-1000 frames, 64s each). Raster lines making up each frame were of 2 or 2.3 ms duration, resulting in frame rates of 7-16 Hz. In some experiments, harmaline was administered by intraperitoneal injection (10 mg/kg). Imaging resumed 7-10 min after harmaline injection.

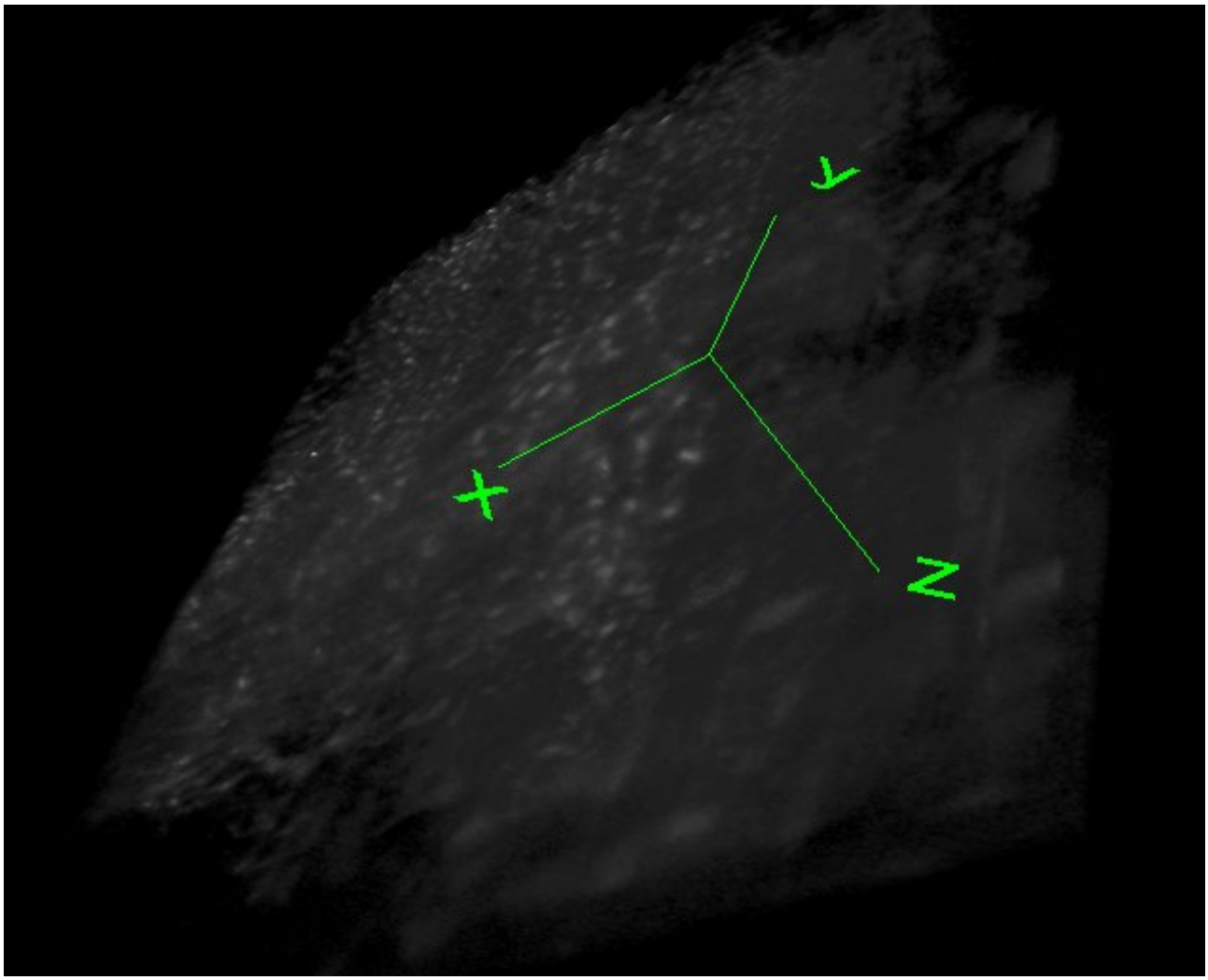


Figure 1.15: A 3D view of a sample rat cerebellar cortex microscopic tissue stack (512×512×251 at 8-bit).

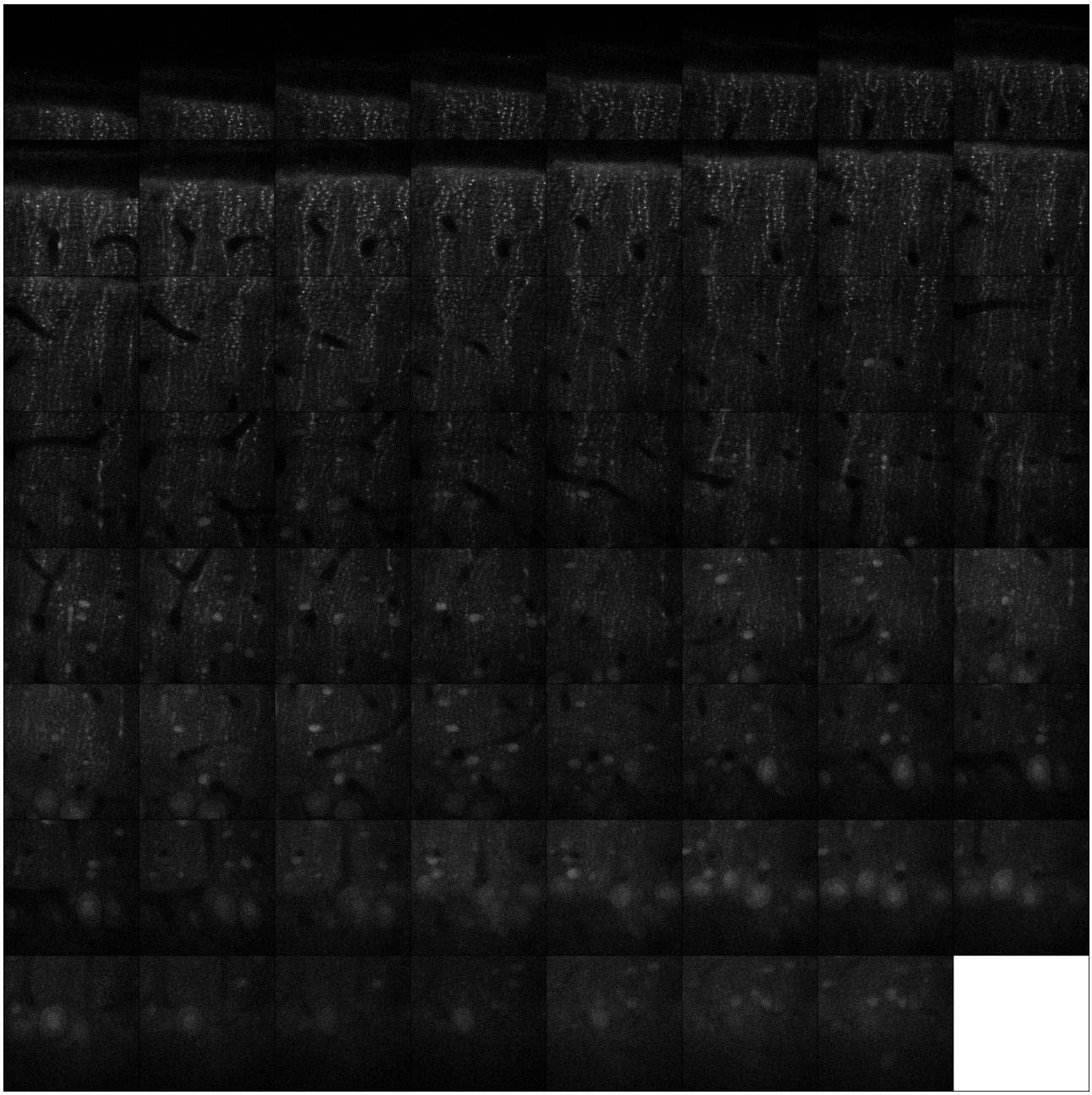


Figure 1.16: A montage view of a sample rat cerebellar cortex microscopic tissue stack, only displaying every other slice along the Z axis.

1.3 Thesis Structure

The thesis is structured as follows:

In Chapter 2, we present the problem of 2D retinal funduscopy segmentation on two public available databases, DRIVE and STARE. We first introduce our proposed framework for 2D elongated linear structure segmentation. The principles behind construction of

phase-invariant orientation field of the local spatial neighbourhood, and the combination of intensity-based measures and local orientation feature maps to form local "feature patches" would unify throughout the thesis. However, the availability of known "ground-truth", as well as the image quality and resolution often constrain the selection for classification techniques. Here, we apply supervised machine learning (MLP or SVM) as an option for "soft" classification to segment single retinal fundus images.

In Chapter 3, we first address a novel co-registration procedure to align and utilize the temporal factor in 2D+t sequential fluorescein retinal angiograms. At the pairwise level, we combine projective RANSAC with a quadratic homography transformation to map the coordinate systems between any two frames. At the joint level, we construct a hierarchical approach in order for each individual frame to be registered to the global reference intra- and inter- sequence(s).

In Chapter 4, we then take a non-training approach that searches in both the spatial neighbourhood of each pixel and the filter output across varying scales to locate and link microvascular centrelines to (sub-) pixel accuracy. In essence, this "link while extract"" piece-wise segmentation approach combines the local phase-invariant orientation field information with additional local phase estimates to obtain a soft classification of the centreline (sub-) pixel locations. This technique does not require training on large data sets, and is particular useful for high-resolution retinal angiograms without "ground-truth" labelling. However, image registration, introduced in Chapter 3, is a necessary pre-processing step to allow vessel segmentation and tracking on angiograms sequences with varying spatio-temporal components.

In Chapter 5, we first introduce a 5D orientation mapping to capture 3D geometric orientation properties. This is essential to differentiate neural cellular structures, such as Purkinje cells, neural dendrites and inter-neurons, as they all display certain elongation along their medial axes, yet each class has a characteristic shape captured by an orientation field that distinguishes it from other structures. This 5D-to-3D mapping is incorporated into the local feature maps description prior to a learning machine.

In Chapter 6, we summarise performance evaluations and/or validation of each of the techniques presented in this thesis. For retinal fundus images, we compute Receiver Operating Characteristic (ROC) curves on existing public databases (DRIVE & STARE) to assess and compare our algorithms with other benchmark methods. For 2D+t retinal angiography sequences, we compute the error metrics ("Centreline Error") of our scheme with other benchmark methods. For microscopic cortical data stacks, we present evaluation results on surrogate data with known ground-truth.

In Chapter 7, we summarize our contribution of each chapter and discuss the pros and cons of our suggested methods in each image modality when compared with other approaches as well as approaches we have taken that are less successful than our proposed methods. And finally, we conclude the thesis with suggestions to areas/topics of future work.

Chapter 2

2D Retinal Funduscopy: Segmentation

2.1 Introduction

Segmentation is an important first step in the automated analysis of several types of biomedical image data. Unlike object recognition, the techniques that have been developed for structure segmentation are optimised for particular resolutions, modalities and contrast (staining) mechanisms. The diversity of these techniques (literature review see Section 2.3) is a reflection of the fact that there are few methods that generalise well across the many contrasting mechanisms increasingly found in biomedical imaging.

In this chapter, we describe (in Section 2.4) the use of phase-invariant complex wavelet filters, coupled to a training process involving a small, high-quality training dataset, to build an image segmentation system capable of performing in low signal-to-noise, in the presence of pathological structural changes, and under conditions of strong object-background contrast change. Lastly, we discuss the benefits of our approach and conclude this chapter in Section 2.5.

2.2 Context

A common problem of medical or biomedical image segmentation is that the connectivity of boundaries should be enforced. This is often achieved through a post-processing method known as *edge-linking*. Incorrect edge linking can be a source of poor segmentation results. Capturing the properties of pixels across larger scales of space provides a method of making edge-linking more reliable. But it is inelegant to extend basic edge-linking algorithms beyond immediate neighbourhood pixels, because the number of decisions and the complexity of the decision process become high. Our aim is to use a sufficiently rich feature vector that incorporates neighbourhood information appropriate to describing the weak, thin nature of the ocular vascular boundaries, and implicitly encodes the continuity information in the feature vector itself. Rather than using explicitly coded rules to enforce continuity, the feature vector and a training process are used instead, making this a more adaptable approach, should imaging statistics change.

2.3 Literature Review on Retinal Vascular Segmentation

Retinal vascular segmentation techniques utilise the fact that there is, to certain extent, sufficient contrast and characteristic difference between the retinal vasculature and its surrounding background to make a viable decision on the classification of image pixels (whether they belong to the blood vessel or not). The cross-sectional grey-level profile of a typical vessel conforms to a Gaussian shape (see Figure 2.1), while the vasculature is piecewise linear and may be represented by a series of connected line segments and that the vasculature originates from the same point (the optic disc) and all vessels are connected [48] [49] [3].

The main techniques used to segment the vasculature form retinal fundus images can be categorized as:

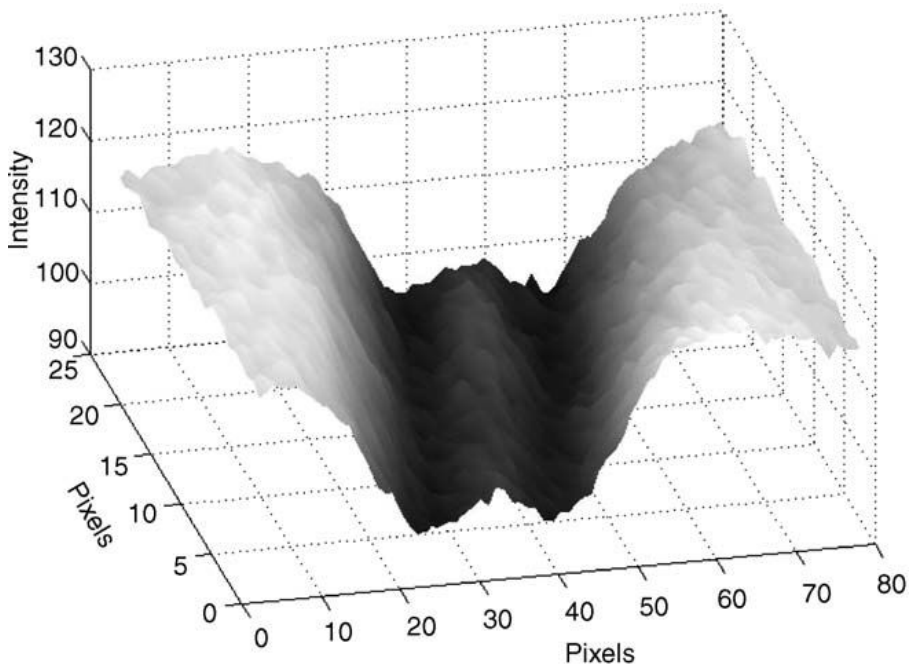


Figure 2.1: Three-dimensional image of an intensity profile from a retinal vessel obtained by fluorescein imaging, cross-section resembles "double-Gaussian" construct.

- **Matching filter** (MF): This approach exploits the known "piecewise-linear approximation" of the vasculature, the decrease in vessel diameter along vascular length and the Gaussian-like intensity profile at vessel cross-section to segment vessels. It involves using a rotated 2D linear "structural element" (kernel) that has a Gaussian cross-profile section to identify the cross-profile of the blood vessel, which typically has a Gaussian or a Gaussian derivative profile [48] [49] [29] [50]. The kernel is rotated into many different orientations in order to fit into vessels of different directions. The image is then thresholded (an arbitrary chosen grey level divides all features into a binary classification, depending on whether they have a greater or lesser intensity level than the brightness threshold) to extract the vessel silhouette from the background. This approach works reasonably well on images of healthy retina but may experience difficulties in the presence of diabetic retinopathy. Also, using too long a structuring element may have difficulty in fitting into highly tortuous vessels.

- **Multiscale**: This approach is based on scale-space analysis. As the width of a vessel gradually decreases across its length after each bifurcation when it travels outward from the optic disc (OD), scale-space representation seems appropriate to address the change

in scale of vessel-width. Multiscale eigenvalue decomposition of the second order local structure of an image ("Hessian" matrix) can be used to extract vessels [51]. Hessian matrix maxima, when combined with edge information and the maximum principal curvature can further quantify the width, size and orientation of retinal blood vessels [52] [53]. By interpreting multi-scale filter outputs using statistical maximum likelihood estimates [54] [55], an estimate of the width, contrast, and direction of a putative blood vessel at each point in the image can be determined in the presence of additive noise or a featureless background.

- **Morphological processing**: This approach exploits the known "piecewise-linear approximation" of the vasculature to segment vessels. Structuring elements of a certain intensity can be added ("dilation") or subtracted ("erosion") to the underlying image [56] [57]. Opening (erosion followed by dilatation) with a structuring element of a certain shape can separate objects in an image by preserving structures that match the scale and the shape of the structural element and removing those that don't. Closing (dilatation followed by erosion) can be used to "fill-in" small holes within an image. Combined with Laplacian-of-Gaussian (LoG) filtering and cross-curvature evaluation, binary image morphology has been demonstrated to be reasonably fast and noise resistant [57].

- **Vessel tracking**: This approach locates vessel centrelines by searching over each vascular cross-section along the longitudinal axis of the vessel, between a given start and end point. It tends to work on single retinal vessels rather than the entire vasculature and requires manual seeding/identification of start and end points [58]. It may be confused by vessel crossings and bifurcations but can provide accurate measurements of vessel widths and tortuosity.

- **Pattern-recognition**: Published methods include ridge-based primitives classification [2], k-Nearest Neighbor (kNN) classifiers [11], Principal Component Analysis (PCA) [59], Gabor wavelet with supervised classification using a mixture of Gaussian [60], and line operators and Support Vector classification [61]. The unsupervised approach, on the

other hand, do not rely on any prior labeling knowledge to differentiate between vessel and the background pixels. Published methods include the maximum likelihood estimation of vessel parameters [54], the fuzzy C-Means clustering algorithm [62] and local entropy thresholding using gray-level co-occurrence matrix [63].

Yet, most algorithms are developed on either red-free fundus images ([48], [11] and [54]) or fluorescein angiograms ([57], [64] and [52]), and are typically tested and refined on few databases.

Recent reviews [18] [19] on retinal imaging and image processing have spotted the industry migration towards digitization and integration of patients' health records. However, longitudinal integration of medical records taken over the span of decades is by no means an easy task. The fast advancement in biomedical imaging devices and technologies results in supplier-specific non-standardized medical data management and handling, thus techniques that can easily adapt to scale and format differences among various image modalities yet employ certain parameters that are unique and distinctive enough to classify structures of interest are highly desirable and practical.

2.4 Proposed Pixel-level Patch-based Segmentation with Training

2.4.1 Construction of Phase-invariant Orientation Field of the Local Spatial Neighbourhood

The feature descriptor that we use here is first defined in [65], [66] and [67]. To take into account variations in retinal images, this approach compactly encodes the neighbourhood phase-invariant structural orientation and spatial symmetry estimate. Moreover, it shares some benefits of the scale space approach as the filter scale selection is automated to be "scalable" to the local vessel diameter. This tuning takes the form of a weighting applied to

each scale of the filter output, which is based on the likelihood of that filter scale being well matched to the local vessel size. In-depth discussions and applications of "steering in scale" and "detecting intrinsic scale" in a more generalized setting can be found in [67].

The phase-invariant orientation field is a representation of local image structure that is relatively discriminating yet stable in the presence of illumination change. It may be thought of as a quasi-illumination invariant estimation of local structure orientation, independent of local spatial symmetry about a selected dominant axis. A sample of the estimated phase-invariant orientation field imposed over a patch from a retinal fundus image is shown in Figure 2.2 and 2.3. A sample of the estimated phase-invariant orientation field imposed over a sub-region in thoracic aortic microscopy image is shown in Figure 2.4. Note, in Figure 2.4, the contrast in the sub-region has been manually enhanced for the purpose of visibility in print. Contrast enhancement is not used prior to calculating the orientation field itself. The inclusion of non-retinal image here is a generalization of the 2D composition of our phase-invariant orientation field on other image modalities. More detailed discussion can be found in our recent paper [68].

It is constructed as:

$$\mathrm{O}^{(l)}(x,y) = \frac{\sum_{k=0}^{K/2-1} |g_k^{(l)}(x,y)|\overrightarrow{u_k}}{p + (\sum_{k=0}^{K/2-1} |g_k^{(l)}(x,y)|^2)^{\frac{1}{2}}} \tag{2.1}$$

$g_k^{(l)}(x,y)$ where $k = 0, 1, 2, ..., \frac{K}{2}$-1 denotes the output of the k^{th} oriented bandpass complex analysis filter from image f at level l. Level l can be interpreted as equivalent to "scale" in scale-space filtering approach. p is a conditioning constant set at 1.25% of the maximum value of the image [66].

The normalized magnitude of the scaled vector response ranges from 0 to $1 - \epsilon(p)$, and it can be used as an indication of **anisotropy**: strongly isotropic neighbourhoods will produce values close to 0 whilst strongly anisotropic neighbourhoods will produce values close to 1.

For $K = 8$, we use the directional vectors $\overrightarrow{u_k} = [1, 0], [0, 1], [\text{-}1, 0], [0, \text{-}1]$. For each of the base filter, in the Fourier domain, a sampling on the 2D interval $[-\pi, \pi] \times [-\pi, \pi]$ is first performed with a grid spacing of $\pi/64$ in each cartesian direction. The choice of an odd matrix size for

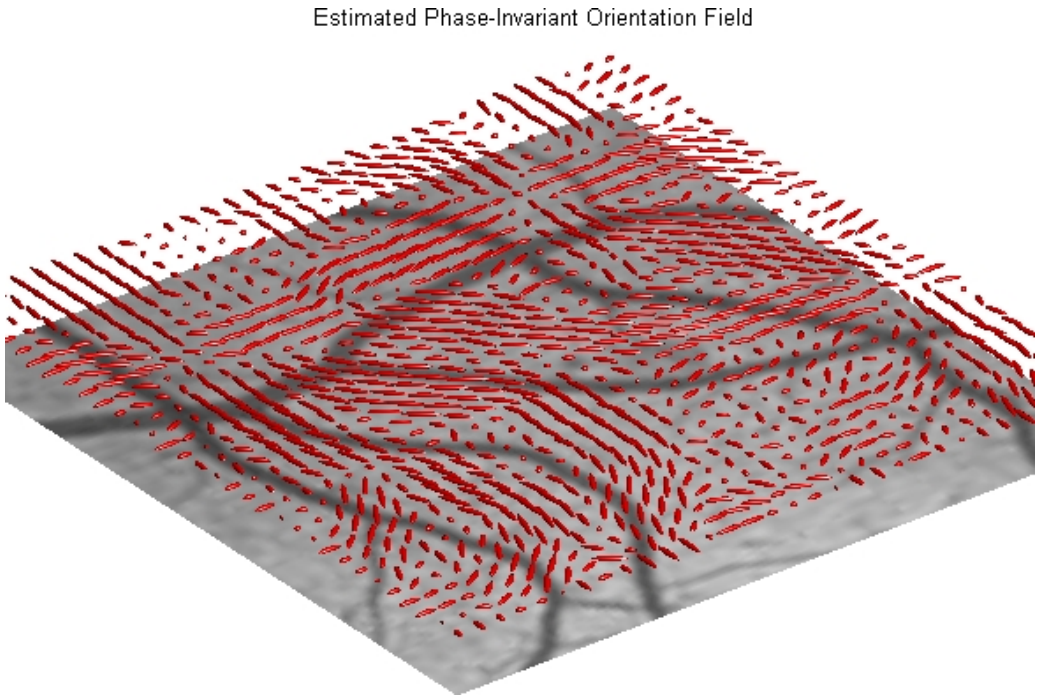


Figure 2.2: Estimated phase-invariant orientation field (in reds) of a 128×128 patch from a retinal angiogram with 3D view.

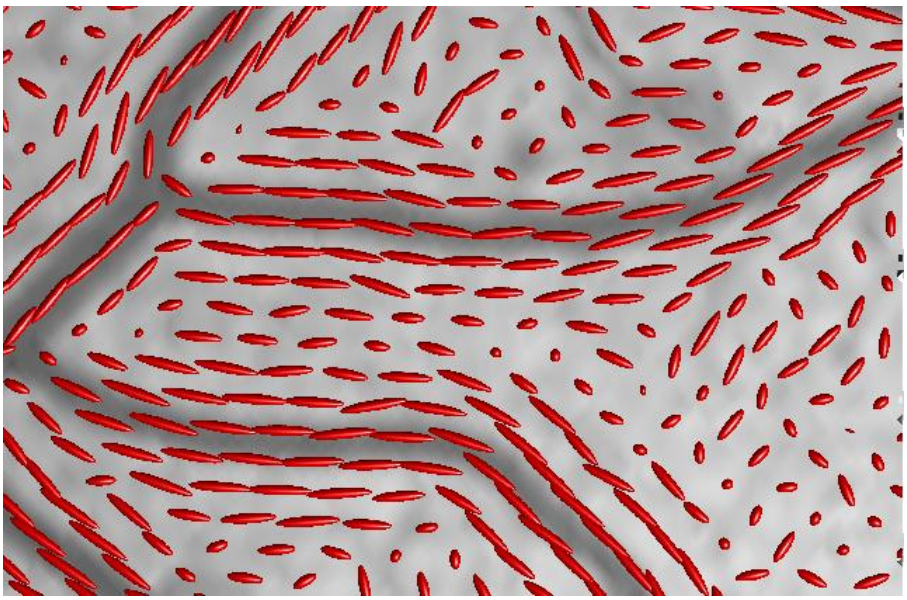


Figure 2.3: Estimated phase-invariant orientation field (in reds) of a 128×128 patch from a retinal angiogram with a zoomed-in detailed 2D view.

constructing the Fourier domain representation is tied to the symmetry of the filter kernels, which we have observed to be better on odd-sized grids [66]. The sampled Fourier spaces for each of four oriented kernels can be visualized in Figure 2.5.

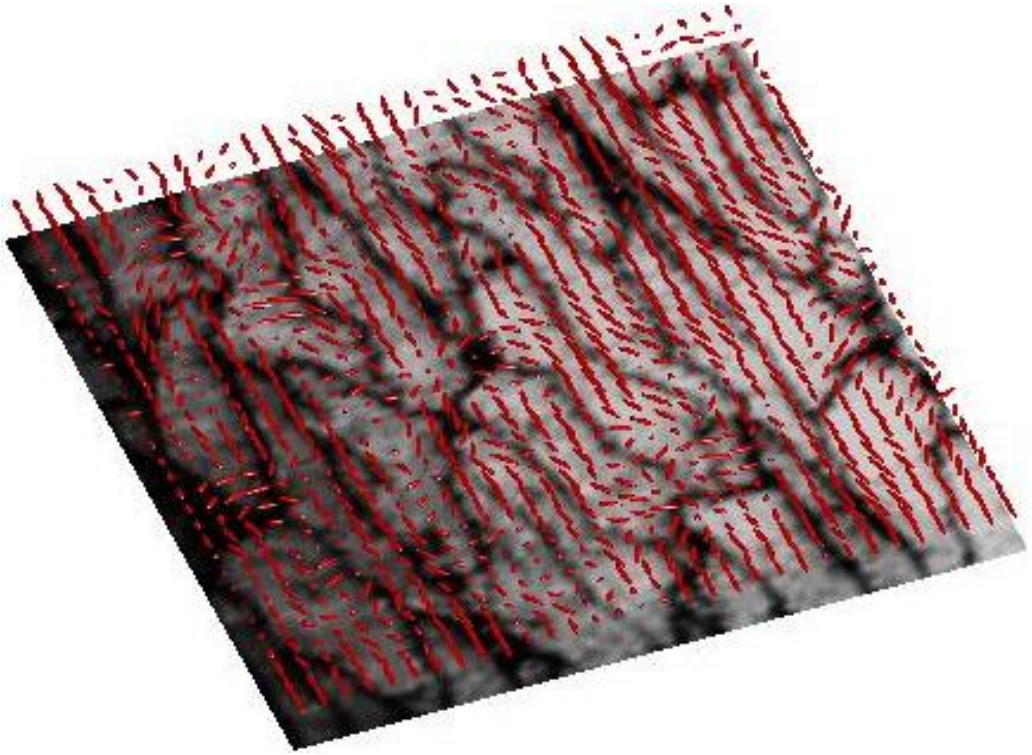


Figure 2.4: Estimated phase-invariant orientation field (in red) for a high-quality sub-region from a microscopy image. Regions are typically lower contrast and with greater degrees of noise.

The inverse discrete 2D Fourier Transform is then computed to extract 65×65 spatial impulse responses. These responses are each truncated to fit into a set of four 15×15 complex arrays. The complex "Steerable Wavelet Filter" (SWF) arrays extracted can be visualized in Figure 2.6.

2.4.2 Construction of Feature Patches from Intensity-based Measures and Local Orientation Feature Maps

The orientation field $\mathbf{O}^{(\ell)}$ of the image, $I(x, y)$, can be described at decomposition level, ℓ, for a multi-rate, quadrature-pair steerable pyramid by

$$\mathbf{O}^{(\ell)} = \frac{u^k \psi_k^{(\ell)}}{\varepsilon + \sqrt{\sum_{k \in \{0,1,2,3\}} (\psi_k^{(\ell)})^2}} \tag{2.2}$$

where a term ab signifies the geometric product between multivectors in a $\mathcal{C}\ell_2$ algebra;

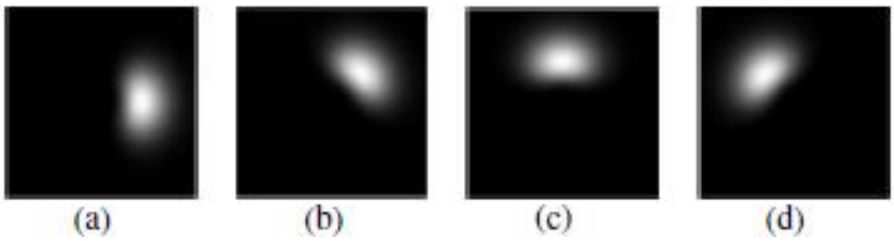


Figure 2.5: Oriented band pass frequency response at 0 radians (a), $\pi/4$ radians (b), $\pi/2$ radians, and $3\pi/4$ radians (d).

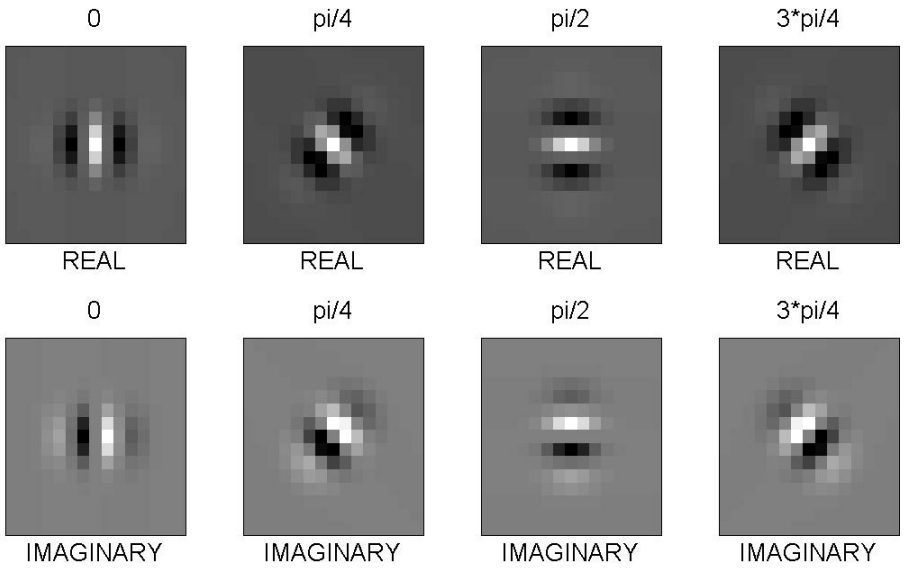


Figure 2.6: Real and imaginary parts of oriented complex bandpass filter kernels used to produce the filter outputs in the spatial domain.

Einstein's summation convention is used in the numerator of Equation (2.2). ε is a constant, described in [66], and the variables $\psi_k^{(\ell)},\ k = 0, 1, 2, 3$ refer to the *magnitude* outputs of four, directed, quadrature filter pairs which are applied by spatial convolution to $I(x, y)$ at multiple scales. The orientation field $\mathbf{O}^{(\ell)}$ is represented as a field in $\mathcal{C}\ell_2$ by weighting unit-length grade-1 vectors by the magnitudes of phase-quadrature filter outputs. The unit-length vectors, $u^k, k = 0, 1, 2, 3$, are formed by $u^k = v^k v^0 v^k$ (in the non-commutative $\mathcal{C}\ell_2$ algebra), where $v^k, k = 0, 1, 2, 3$ represent the unit-length grade-1 vectors describing the k^{th} quadrature filter axis of spatial phase symmetry/antisymmetry. The RHS of Equation (2.2) takes into account the double-coverage of the directed quadrature filter, and introduces a phase-invariant nature to the spatial orientation field $\mathbf{O}^{(\ell)}$. The magnitude of this field ranges from 0 to 1 and can be used as an indication of **local anisotropy**, in which strongly isotropic neighbourhoods

produce values near to 0 and strongly anisotropic neighbourhoods produce values near to 1.

For some segmentation problems, it is also advantageous to use 6 statistical features from 3×3 intensity neighborhoods of the candidate pixel. These feature vectors are z-scale transformed, i.e. normalised in standard deviation and zero-mean transformed using: $\nu(i,m) = \frac{\nu(i,m)-\mu_m}{\sigma_m}, \forall i, m,$

where $\{\mu_m\}_{m\in\{1,2,...,M\}}$ and $\{\sigma_m\}_{m\in\{1,2,...,M\}}$ are the mean and standard deviation in the training population of the m^{th} feature. M is the total number of features.

$\{\nu(i,m)\}_{(i,m)\in\{1,2,...,N\}\times\{1,2,...M\}}$ denote the feature vector components before normalisation. One has to exercise caution *not* to normalise the orientation field information from $\mathbf{O}^{(\ell)}$ – which is already well-behaved and bounded – in this way, as the information encoded by the relation between vector components of individual observation vectors will be destroyed.

Our 27-dimensional **input** feature vector (prior to classification) is defined as:

$$\overrightarrow{FV_n} = [I_{3\times3}, \overrightarrow{O}_{(3\times3\times2)}] \tag{2.3}$$

Our unique descriptors for segmentation are assembled as follows: For each pixel, we first assemble its 3×3 neighbourhood intensity. We then separate the horizontal and vertical components of the phase-invariant orientation field, and assemble the $2 \times 3 \times 3$ neighbourhood orientation field. The horizontal and vertical components of the orientation vector are normalized against vector magnitudes to conserve the angular representation. For training, we use randomly selected patch (similar to a discretized Wiener process) on each training image and validate the accuracy of the classification output against "ground-truth".

2.4.3 Supervised Machine Learning (MLP or SVM) for "Soft" Classification

For the classification of pixels into object or background classes, the high-dimensional nature of the problem might suggest the use of either a Multilayer Perceptron (MLP) or a Support Vector Machine (SVM). The former is known to perform fast classification decisions and achieve good recognition rate of curvilinear shapes [69]. The latter is known to perform well on high-dimensional input vectors [70], partly due to the use of a margin-based training approach.

Meanwhile, prior literature [11] [19] indicates that a supervised pixel classification method was proposed by one of the originators of the DRIVE database where a 31-dimensional feature vector was constructed for each pixel in the image, and a kNN-classifier was trained with these feature vectors. Their study also showed that the kNN-classifier performed better than a linear or a quadratic classifier. This architecture first appeared as *convolutional neural nets* [71] used for recognizing handwritten characters.

This combination of feature space and classifier is intended to create higher complexity in decision surface as neither the feature vectors or the classifiers are linear.

Multilayer Perceptron A multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLPs utilize a supervised learning technique ("backpropagation") for training the network [72]. MLP is a modification of the standard linear perceptron and can distinguish data that is not linearly separable. The unique characteristics of MLP, when compared with a perceptron, is that each neuron uses a nonlinear activation/transfer function which was developed to model the frequency of action potentials, or firing, of biological neurons in the brain. This function can be modeled in a variety of ways, but must always be normalizable and differentiable [72]. Furthermore, MLP

consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes (See Figure 2.7). Each node in one layer connects with a certain weight to every node in the following layer.

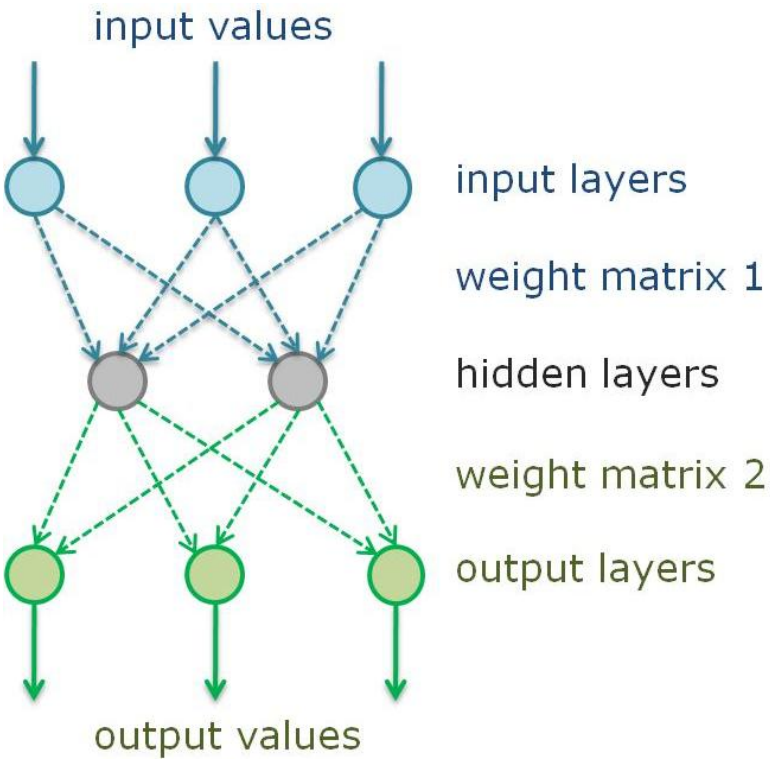


Figure 2.7: A simple Multilayer Perceptron with one input layer, one hidden layer and one output layer.

Action/transfer Function in Artificial Neural Network (ANN) A node's transfer functions serves the purpose of controlling the output signal strength for the node (except for the input layer which uses the inputs themselves). These functions set the output signal strength between 0 and 1. The input to the transfer function is the dot product of all the node's input signals and the node's weight vector. Commonly, four transfer functions: *Sigmoid, Gaussian, Hyperbolic tangent and Hyperbolic secant* are used for training. The functions are selectable on a layer-by-layer basis and networks can be created that incorporate multiple types. The figure below illustrates the behavior of each function.

The *Sigmoid* function is the most widely used function for backpropagation neural networks. It acts as an output gate that can be opened (1) or closed (0). Since the function is continuous, it also possible for the gate to be partially opened (i.e. somewhere between 0 and

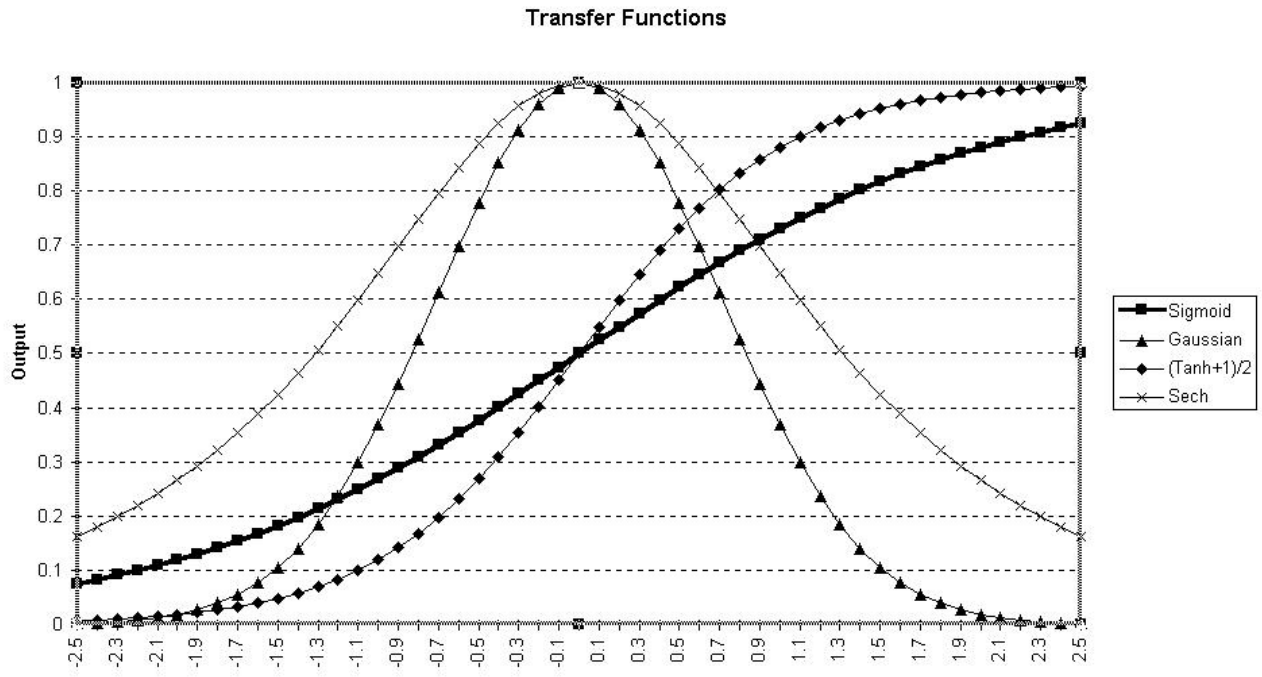


Figure 2.8: Sigmoid, Gaussian, Hyperbolic tangent and secant transfer functions normalize the output signal generated by each node.

1). Models incorporating sigmoid transfer functions often have generalized learning characteristics and yield models with improved accuracy. Use of sigmoid transfer functions can also lead to longer training times.

Gaussian transfer functions significantly alter the learning dynamics of a neural network model. Where the sigmoid function acts as a gate (opened, closed or somewhere in-between) for a node's output response, the Gaussian functions act like a probabilistic output controller. Like the sigmoid function, the output response is normalized between 0 and 1, but the Gaussian transfer functions are more likely to produce the "in-between state". It would be far less likely, for example, for the node's output gate to be fully opened (i.e. an output of 1). Given a set of inputs to a node, the output will normally be some type of partial response, due to the partial opening of the output gate. Gaussian based networks tend to learn quicker than sigmoid counterparts, but can be prone to memorization.

Hyperbolic tangent and *Hyperbolic secant* functions are counterpart to sigmoid and Gaussian functions. The hyperbolic tangent is similar to the sigmoid but can exhibit different learning

dynamics during training. It can accelerate learning for some models and also have an impact on predictive accuracy. Experimenting with transfer functions for each individual model is the only conclusive method to determine if any of the non-sigmoid transfer functions will offer both good learning and accuracy characteristics.

For our applications, we apply MLP with the Levenberg-Marquardt (LMA) training algorithm [73] over a small sample of hand-labelled pixels. The network uses a mean squared error performance function, hyperbolic tangent sigmoid transfer functions for both hidden and output layers, and scaled conjugate gradient descent for iteration.

Support Vector Machine A Support Vector Machine (SVM) [74] can be used to construct a separating hyper-plane which maximizes the distance to the nearest training data inputs. If the data is linearly separable in the feature space implied by the kernel function then the decision rule (in standard algebra) will be:

$$f(\boldsymbol{\nu}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{\nu}, \boldsymbol{\nu}_i) + \beta\right) \tag{2.4}$$

in which $\boldsymbol{\nu}$ is the normalised feature vector. N is the number of training examples. $y_i \in \{-1, 1\}$ are the scalar labels. K is the training kernel. $\{\alpha_i\}_{i=1,2,\ldots,N}$, and β are scalar parameters selected to maximize the margin that partitions the positively and negatively labelled data. $\{\boldsymbol{\nu}_i\}_{i=1,2,\ldots,N}$ are M-dimensional training vectors. The training examples ($\boldsymbol{\nu}_i$) are called *support vectors*, such that $\alpha_i \neq 0$.

In reality, perfect separation of the feature vectors into non-overlapping categories may not be possible, or it may result in a model with so many feature vector dimensions that the model does not generalize well to other data. To allow some flexibility in separating these categories, SVM models have an "error" cost parameter C that regularizes the trade-off between allowing training errors and forcing "rigid" margins. Increasing the value of C increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well. Thus tuning C correctly is a vital step in best practice in the use of SVMs.

For this work, we used `libsvm` [75] via its *Matlab* interface, and restricted our tests to the in-built kernels.

The Radial Basis Function (RBF) kernel $K(\boldsymbol{\nu}, \boldsymbol{\nu}_i) = \exp(-\gamma||\boldsymbol{\nu} - \boldsymbol{\nu}_i||^2)$ is used, where γ determines the area of influence $\boldsymbol{\nu}_i$ has over the data space. We find the RBF kernel effective for pixel-level segmentation of 2D biomedical data [68]. The SVM is first trained using the input feature space with its *a priori* determined ground-truth on a small training image of size 128×128. The accuracy of the SVM model is strongly dependent on the selection of the model parameters, and our search over these parameters is computationally expensive, trying values of each parameter across a wide range. We search over the two "standard" parameters (the aforementioned parameters C and γ), using a range of -4 to +4 with an interval of 0.1 for C and a γ range from 0.1 to 2 with an interval of 0.001. 5-fold cross-validation is done on the training image patch. This validation step is important to ensure the learned model is stable and generalises well to new data. Once the model with the smallest error is generated, segmentation of new data is achieved by casting a new image into 27-dimensional feature space, normalising the vectors and applying the SVM on a pixel-by-pixel basis.

Comparison between MLP and SVM Table 2.1 summarizes the different characteristics between MLP and SVM training approach. Both MLP and SVM have been applied to many classification problems, and generally yielding good performance [69]. The main difference between MLP and SVM training is the principle of risk minimization, also known as the optimization technique. In MLP, traditional empirical risk minimization is used minimizing the error on the training data. In SVM, structural (the support vectors) risk minimization principle is applied by minimizing an upper bound on the expected risk.

As in the past, we have successfully applied SVM for cell membrane segmentation [68]; we initially considered applying SVM in favour for its reported generalizability in higher dimensional data sets [74]. However, we were disappointed by the 70%-80% accuracy from SVM training, and switched to MLP instead, which results in >90% accuracy (results see Table 6.1-6.3 in Chapter 7).

Table 2.1: *Comparison between MLP and SVM*

Multi-class Classification	MLP can solve multi-class problem by simply assigning multi-class labels prior to learning, therefore is more flexible in application.	SVM (a binary classifier) requires using successive classification ("one-against-many") to solve for multi-class problem.
Optimization Technique	In MLP, network weights are adjusted during training such that the sum-of-square error between the network output and the actual value (target) is minimized. This is susceptible to result in local minima if the search starts near a local minimum, hence the global minimum might not be guaranteed by using MLP.	SVM requires an explicit determination of the decision boundaries directly from the training data. This forms the basis of the optimization step that minimizes the aggregate distance between the maximum-margin hyper plane and the support vectors. Thus, SVM guarantees to deliver unique solution (the global minima).
Training Time	MLP training can be parallelized using map reduction therefore is much faster and can withstand large computational load.	SVM training requires solving the associated Lagrangian dual problem which is a quadratic optimization problem with large number of variables, therefore is much slower.
Usability	MLP is often easy to train and fast to use.	Configuration of SVM is often hard as it requires the optimal choice of a variety of parameters: the kernel type, the kernel parameters, and the margin parameters.

We think the success of MLP-classifiers in segmenting retinal vasculature in funduscopy is due to two reasons. First, the specific layout of MLP-classier we applied was initially configured for handwriting recognition, as hand-written alphabets to some extent also exhibit elongation and intermittent gaps at times; the specific configuration may be more fine-tuned for our

biomedical applications. Second, the high SNR and varying illumination, often observed in retinal imaging, may complicate SVM kernel configuration thus leads to suboptimal results.

2.5 Discussion and Conclusion

Figure 2.9–2.11 demonstrates the segmentation results of the global vessel-bed on 3 test images from the DRIVE database test image sets. Receiver Operating Characteristic (ROC) curves on DRIVE and STARE when compared against other segmentation benchmarks are presented in details in Section 6.2.



Figure 2.9: The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 1 (left) in RGB scale.

In this chapter, we first present a literature review on existing methods for vascular segmentation in retinal funduscopy. We then present an explanation and justification of the feature sets that we use, and justify the use of a phase-invariant patch representation to capture spatial structure. When combined with the appropriate machine-learning classifier, the system can perform satisfactory vascular segmentation of the global vessel-bed yet using a

pixel-level local approach. At this stage, we would like to point out that the use of machine-learning classification here is due to the availability of training data with labelled "ground-truth" on the public databases. This general framework is extended from 2D to 3D application in a later Chapter (Chapter 5) where we are able to create 3D surrogate neuronal training data. However, the complexity of neural structures requires our feature representation to encode additional orientation information (our introduction of "3D to 5D orientation mapping") prior to the machine learning classification of structural of interest.

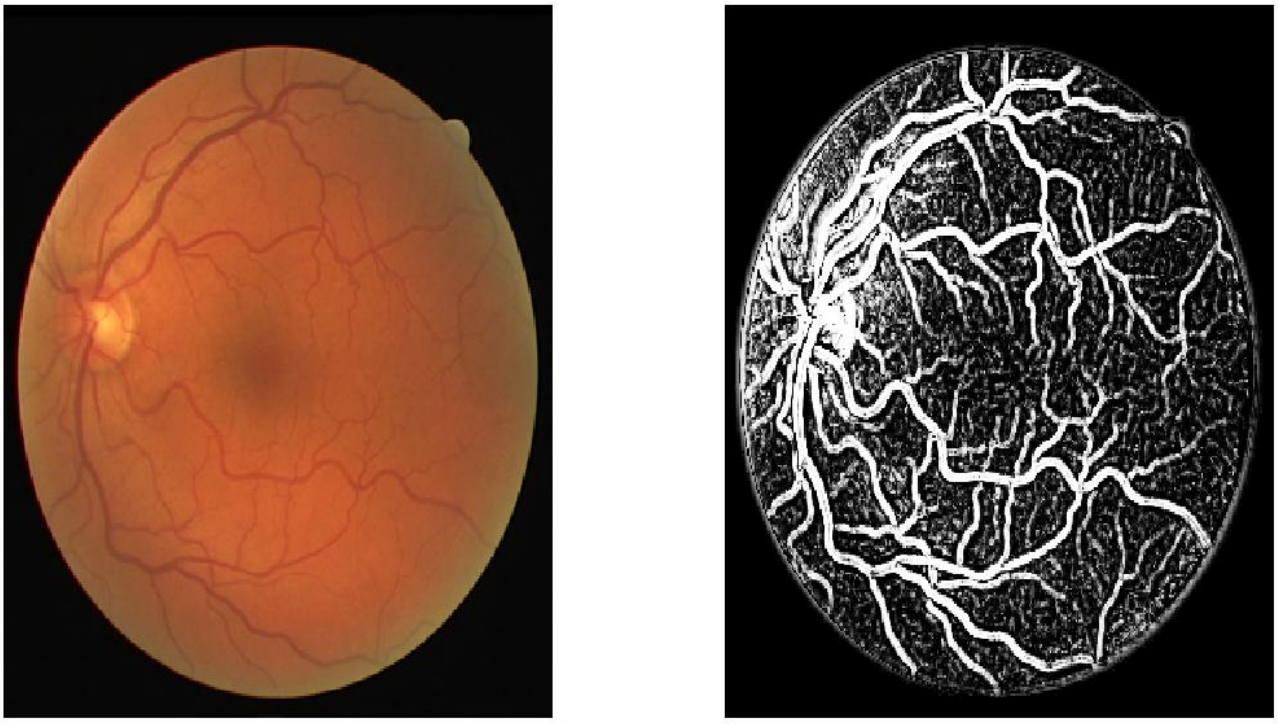


Figure 2.10: The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 2 (left) in RGB scale.



Figure 2.11: The segmentation results (right) (posterior probability map, or "soft decision") of our algorithm (using MLP training) on Test Image 3 (left) in RGB scale.

Chapter 3

2D+t Sequential Fluorescein Angiogram I: Co-Registration

3.1 Introduction

The geometric alignment or registration of images over time is a fundamental task when tracking temporal change is essential. It can be viewed as a preliminary step that allows segmentation to be carried out in both spatial and temporal context. In this chapter, we address our co-registration problem for 2D+t sequential retinal angiograms by pairing a 'pairwise' RANSAC-iterated-homography with a local-to-global hierarchical spatio-temporal 'joint' registration framework (Section 3.4). But first, we start by introducing the technical context for image registration in Section 3.2 and the relevant literature reviews both in generalized computer vision aspect and more specific medical imaging in Section 3.3. Lastly, we discuss the benefits of our approach and conclude this chapter in Section 3.5.

3.2 Context

To automatically assess or characterize the geometrical changes in retinal blood vessels, we have to first address the issue of **image registration**. Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns two images (the reference and sensed images) so that image analysis tasks to combine the information obtained from possibly various sources can be carried out.

Image registration is usually classified by taking into account the dimensionality of the data (i.e. 2D, 3D, 4D), the complexity of image deformations, the computational budget, domain of application (i.e. medical or remote sensing), and the mathematical principle behind the algorithm.

Various review papers [76] [77] [78] [79] [80] provide in depth coverage of the subject, yet each have a slight different view on the taxonomy of the classification. To take into account the fuzzy nature of such classifications, we outline the registration literature in a more generic manner and adopt an application-driven perspective, with a further split into feature- and area- based methods, culminating in a review of image registration techniques.

3.2.1 Application-Based Relevance

There are four subcategories: multi-view; multi-temporal; multi-modal and scene-to-model.

- *Multi-view registration* re-constructs a larger 2D or 3D representation of the scene from images acquired at different viewpoints, and is often referred to as mosaicing (2D) or shape-recovery (3D).

- *Multi-temporal registration* detects and evaluates changes from images of the same scene acquired at different times and possibly under different conditions. It plays a key role in pathological monitoring and diagnosis in medical imaging

- *Multi-modal registration* integrates information obtained from different source streams to gain more complex and detailed scene representation. In radiotherapy and nuclear medicine, it is more and more common to fuse MRI, ultrasound or CT images that record anatomical structure with PET, SPECT or MRS images that monitor functional and metabolic activities.

- *Scene-to-model registration*, sometimes referred to as atlas-patient registration in the medical imaging context, localizes the acquired image (scene) in the model (i.e. model of the scene, possibly a computer representation or "average" specimen). In the computer vision context, this can be applied in target template matching with real-time images or automatic quality inspection with respect to a reference or a prototype.

With regard to our application, the principal technique amongst those above that is of relevance to us would be a multi-temporal registration. However, we also wish to extract temporal information from what is often considered to be "noise" in the registration context. Our approach actually makes use of temporal information to extract local flow information in vessels, specifically by not "ignoring" the noise.

3.2.2 Feature-/Area- Based Relevance

Conventionally, image registration techniques can be classified as feature- or area- based.

Feature-based techniques typically consist of four steps: Feature detection, feature matching, transformation model estimation, image resampling and subsequent transformation.

- *Feature detection* extracts salient structures or distinctive objects in the images to be registered either manually or automatically. Features can be *regions* (projections of general high contrast closed-boundary regions of an appropriate size [81]; affinely invariant neighbourhoods [82]), *lines* (line segments [83]; object contours [84]) or *points* (region corners [85]; intersections of straight lines [86]; centres of gravity of closed-boundary regions [87]; high variance points [88]; points of inflection on curves of

high curvature [89]; the most distinctive points with respect to a specified measure of similarity [90]).

The common aspect amongst all feature-based techniques is that the spatial features to be used for registration should be distinct, spatially well distributed and detectable ("remarkableness") in both sensed and reference images.

Ideally, the algorithm should be able to detect the same features in all projections of the scene regardless of potential object occlusions or image degradation. This demands high physical interpretability and accuracy in the localization of the features.

- *Feature matching* establishes pairwise feature correspondence between the sensed image and the reference image. The development of discriminative and robust feature descriptors is key to being able to counter temporal instability in respective features across time. Typically, in cases when captured features exhibit ambiguity or their neighbourhoods are locally distorted, additional spatial relationships among these features (local structural information) could be taken into account for matching. This can be accomplished through the use of clustering in transform space by [86] or chamfer matching [91] to minimize distance/root mean square average between features.

  Alternatively, much attention has been focused on developing descriptors resilient to image deformation. Common criteria for these features would be a mix-and-match combination of the following:

  - *invariance* of the same feature in different frames (i.e. at different points in time).
  - *uniqueness* in order to be able to distinguish among different features.
  - *stability* in order to extract the same feature in assumed deformed scenarios.
  - *independence* so that single feature without corresponding counterparts in one frame should not affect the performance of detection in the other frame.

- *Transform model estimation* is a process that constructs one or more mapping function(s) to align a sensed image with a previously sensed or stored reference image. The type of the mapping function is generally chosen based on *a priori* knowledge on

plausible geometric deformation, expected image degradations and the image acquisition process. To determine the parameters of the model, one can choose to use either global maps or local maps.

The global mapping model, typically composed of bivariate polynomials of relatively low order, aims to estimate one set of the mapping function parameters that is valid for the entire image. In cases that there are more corresponding feature pairs than is minimally required to determine the mapping function, a least-squares fit is applied to compute the set of parameters that minimize the sum of squared errors at the control points. This is implemented by [31], adopting technique published by [92] and taking in account scaling, rotation and translation.

On the other hand, a local mapping model decomposes the image into patches, each defined by a different set of parameters. Many locally sensitive registration methods perform image tessellation, and then collect local mapping functions valid within each patch. Notable approaches include piecewise linear mapping [81]; piecewise cubic mapping [93]; and most recently, local B-splines based on free-form deformation (FFD) proposed by [94], which are considered state-of-the-art.

- *Image resampling and transformation* is a necessary process that acts upon the sensed image, using either candidate or final mapping functions constructed from the previous step. Assuming that the input into image resampling algorithm consists of image values sensed at integer coordinates. The image values at the non-integer coordinates are computed by appropriate interpolation techniques.

  The transformation applied on the sensed image can be realized in a forward or backward manner. Each pixel from the sensed image can be directly transformed using the estimated mapping functions, hence called a forward method. This, however, can result in holes and overlaps in the output image from discretization and coordinate rounding.

  The backward method uses the coordinates of the target pixel (adapted from the reference image) and the inverse of the estimated mapping function. On the regular grid

of the sensed image, interpolation is realized via convolution of the image with an interpolation kernel, commonly nearest-neighbour or bilinear [95], on the basis of the trade-off between accuracy and computational complexity.

When images lack certain distinctive saliency or detail, area-based techniques can be used to estimate the correct mapping more reliably. Such methods work directly from pixel intensities and put more emphasis on matching, by optimizing appropriate similarity measures within a certain neighbourhood.

The similarity measures can vary from least mean square error, normalized cross-correlation [96]; entropy of each image [97]; pattern intensity [97]; phase correlation [98]; gradient correlation [99]; gradient difference [100]; to mutual information [101]. Pre-defined windows or even the entire image may be the neighbourhood for establishing correspondence between regions.

However, the major drawback of area-based approaches is that the transformation model is highly dependent on the window/neighbourhood. A simple rectangle or circular window may not be able to accommodate the complex geometric deformation (similarity, perspective transform for instance) present between images. Furthermore, lack of "remarkableness" within such window/neighbourhood due to large smooth/textureless areas could easily result in mismatch. Lastly, both the reference and the sensed images would have to share "similar" intensity functions for correlation-type methods to be applicable or at least display statistical similarity.

3.3 Literature Review on Registration in Fluorescein Angiograms

The purpose of registration for our problem is to estimate and model the distortion between frames in order to map each angiogram onto one common coordinate system ("*reference*"). In

the medical imaging context, it allows clinicians to review and compare spatially and temporally aligned medical scans for diagnosis and treatment.

In retinal angiogram analysis, feature-based methods rely on the stability of the feature descriptors and the accuracy of their transformation model estimation to achieve successful registration. Often, vascular bifurcations and crossing points are extracted as features [57] [52] [102]. An improved method [103] extracts both "corner" points and "face" points to improve the point-feature distribution. Other algorithms such as "Scale Invariant Feature Transform" (SIFT) [104] and "Speeded Up Robust Features" (SURF) [105] have also been suggested. For transformation model estimation, schemes such as "Iterative Closest Point" (ICP) [64] [106] [103] [107] are often compared with "RANdom SAmple Consensus" (RANSAC) [108] [102]. Accurate initialization of matching point correspondence is essential for ICP, as it has a narrow domain of convergence. Therefore, [103] used SIFT descriptors to guarantee at least one "correct" key point match.

Geometric distortion, radiometric degradation, and additive noise corruption contribute to the difficulty in registration. It is necessary to be able to distinguish between image deformation (our geometric "noise") and the real change of the scene (our "signal"). Problems specific to clinical retinal angiography are: the photographer's bias in image capture, the patient's involuntary movement, and the 3D to 2D warping between retina and the camera. Existing techniques, for example, approximate the retinal surface by a sphere [64], assume a perspective distortion [109], or model the intensity variation to address uneven illumination [110].

3.4 Proposed Co-registration Procedure: "Joint Global RANSAC" method

The chart to illustrate the hierarchical stages of our co-registration algorithm ("Joint Global RANSAC" method) can be found in Figure 3.1.

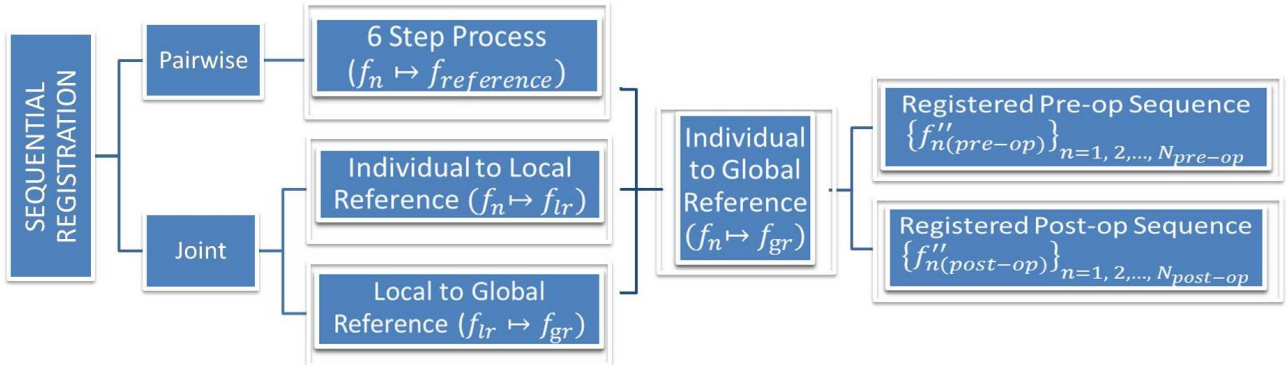


Figure 3.1: Hierarchical stages of processing for sequential registration.

3.4.1 Pairwise (6 step process): Projective RANSAC with Quadratic Homography Transformation

At the pairwise image level, we combine projective RANSAC with a quadratic homography transformation. This is described in 6 steps below:

1. Apply a circular Hough Transform to extract the "field-stop" mask.
2. Pre-process the image by contrast-enhancement (histogram equalization) within the field-stop.
3. Extract point features (vessel bifurcations and crossings) using the Harris corner measure [111] with self-threshold.
4. Putatively match pairs of feature points using normalized cross-correlation.
5. Optimize transformation mapping using iterative "RANdom SAmple Consensus" [108].
6. Apply a projective model by assessing registration parameters.

Step One

The first step, which extracts the binary "field-stop" mask, is an essential step in the "pre-processing" that improves the accuracy of detected features by taking into account the

spatial distribution of the extracted features. We observe non-negligible amount of "false" corners accumulating around the "field-stop". Their presence inevitably affects the accuracy of our global-mapping transform model. Therefore, we apply the **Hough Transform** [112] to contour-fit the "field-stop". This in turn becomes the location criteria to filter out false corners before estimating the model parameters: corners detected around and beyond the "field-stop" circle are eliminated in this process.

The Hough Transform is an effective method to segment overlapping or semi-occluded objects [8]. Often the edge map for the retinal boundaries are broken into segments, which can be rather difficult to link these segments into a coherent circle. Thus, Hough Transform's insensitivity to to missing parts of lines, to image noise, and to other non-line structures co-existing in the image [8] is a good fit for this task. The principle of using Hough Transform to search parametric space in order to locate circles is illustrated in Figure 3.2.

A circle with radius r and centre (a, b) can be described with the following analytical expression $\mathbf{f}(\mathbf{x}, \mathbf{a})$, where $\mathbf{a}$ is the vector of circle parameters and $\boldsymbol{x} = (x, y)$ represents a pixel on the circle:

$$(x - a)^2 + (y - b)^2 = r^2 \tag{3.1}$$

This results in a three dimensional parameter space: two parameters for the centre of the circle and one parameter for the radius of the circle. All pixel $\mathbf{x}$ whose edge magnitude exceeding some threshold can be considered possible circle pixels. All parameters a, b and r are then sampled into a limited number of values to enable the discretization of the parameter space into a 3D box structure of cells. This array of cells is called the **accumulator array A**, whose elements are **accumulator cells A(a)**. As illustrated by the psuedo-code below 1, the accumulation cell $\mathbf{A}(\mathbf{a})$ is incremented if the point (a, b) is at distance r from point $\boldsymbol{x}$, and this condition is valid for all triplets (a, b, r) satisfying the parameter space. If some potential centre (a, b) of a circle of radius r is frequently found in the parameter space, it is highly probable that a circle with radius r and centre (a, b) really exists in the processed image.

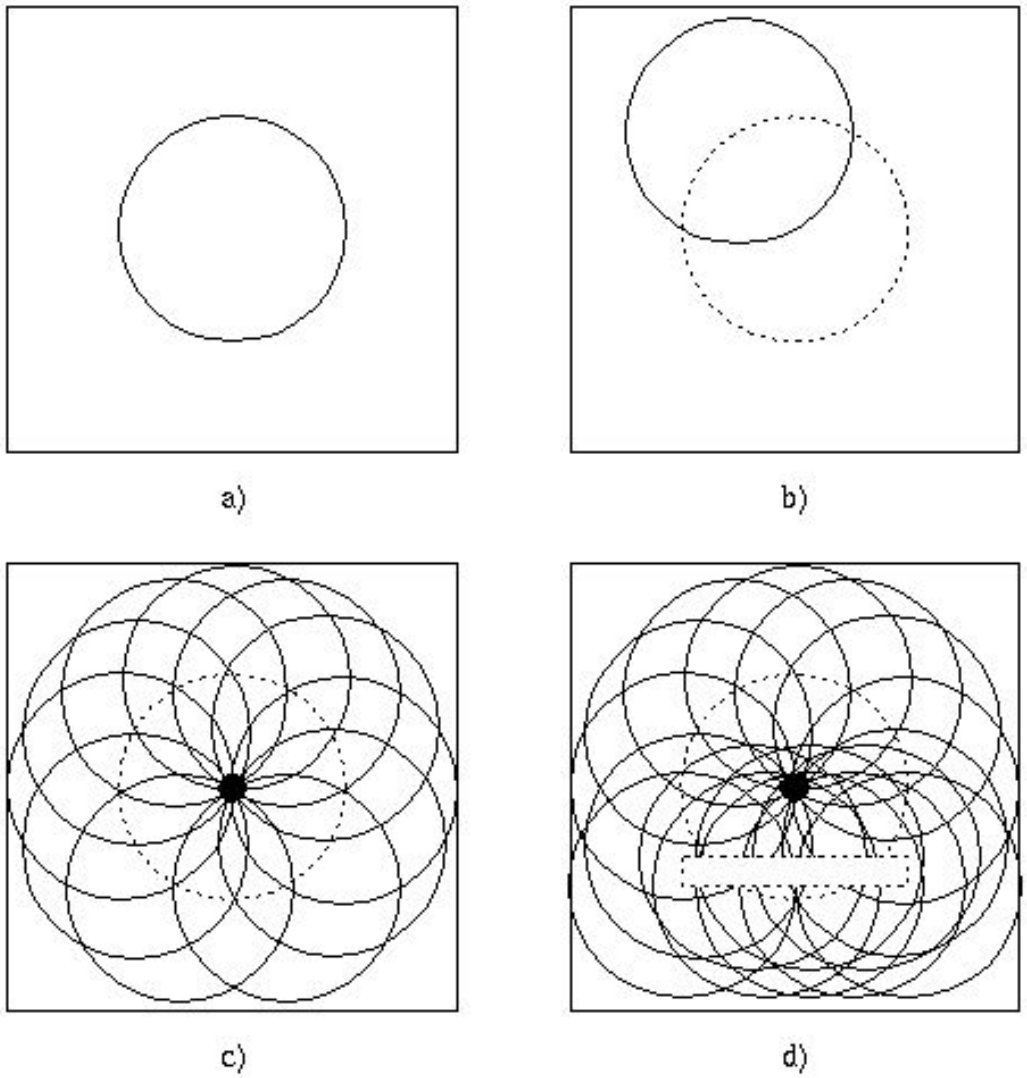


Figure 3.2: (a) Original circle with radius r; (b) For each pixel located on the contour (in dash line), a potential circle-centre locus is defined by a circle (in solid line) with radius r and centre at that particular pixel; (c) The frequency of which image pixels occur in the circle-centre loci is determined, the highest frequency pixel (black dot) represents the centre of the circle; (d) Circle detection in the presence of incomplete circle information (occlusion) and overlapping [8].

To avoid exponential growth of the accumulator data structure with the increase in dimensionality of the parameter space, one common approach to reduce the computation demand is to consider prior information about edge directions [8]. Without using edge direction information, all accumulator cells $\mathbf{A}(a, b)$ are incremented in the parameter space if the corresponding point (a, b) is on a circle with centre $\boldsymbol{x}$ and radius r. But if edge directions are quantized into 8 possible values, theoretically only 1/8 of the circle needs incrementation of accumulator cells. In practice, assume edge direction errors are in the scale of $\pi/4$, still

Algorithm 1 Hough Transform Circle-fit Pseudo-code

1. Quantize the parameter space within the limits of parameters **a**. The m dimensionality of the parameter space is given by the number of parameters of the vector a.
2. Form an m-dimensional accumulator array **A** with structure matching the quantization of parameter space; set all elements to zero.
3. For each image point $\boldsymbol{x}$ in the appropriately thresholds gradient image, increase all accumulator cells $\mathbf{A}(\mathbf{a})$ by an increment of $\triangle\mathbf{A}$ if $\mathbf{f}(\boldsymbol{x}, \mathbf{a}) = 0$

$$\mathbf{A}(\mathbf{a}) = \mathbf{A}(\mathbf{a}) + \triangle\mathbf{A}$$

   for all **a** inside the limits used in step 1.
4. Local maxima in the accumulator array **A** correspond to centres of circle in the original image.

only 3/8 of the circle requires incrementation of accumulator cells. This procedure is expressed in the equations below, where $\psi(\boldsymbol{x})$ refers to the edge direction in pixel $\boldsymbol{x}$ and $\triangle\psi$ is the maximum anticipated edge direction error, and accumulator cells in the parameter space are only incremented if (a, b) satisfy the equation 3.2.

$$a = x - rcos(\psi(\boldsymbol{x})) \tag{3.2}$$

$$b = y - rsin(\psi(\boldsymbol{x})) \tag{3.3}$$

$$\psi(\boldsymbol{x}) \in [\psi(\boldsymbol{x}) - \triangle\psi, \psi(\boldsymbol{x}) + \triangle\psi] \tag{3.4}$$

An example of the result of applying Hough Transform to extract retinal "field-stop" from our retinal angiogram is shown in Figure 3.3, with the parameter space displayed in Figure 3.4.

Step Two

It is commonly observed that the concentration of fluorescein dye in the retinal blood vessels is rather low in both the beginning (arterial phase) and the end (late venous and recirculation phase) of the sequence. This suggests that temporary adjustment of the dynamic range of

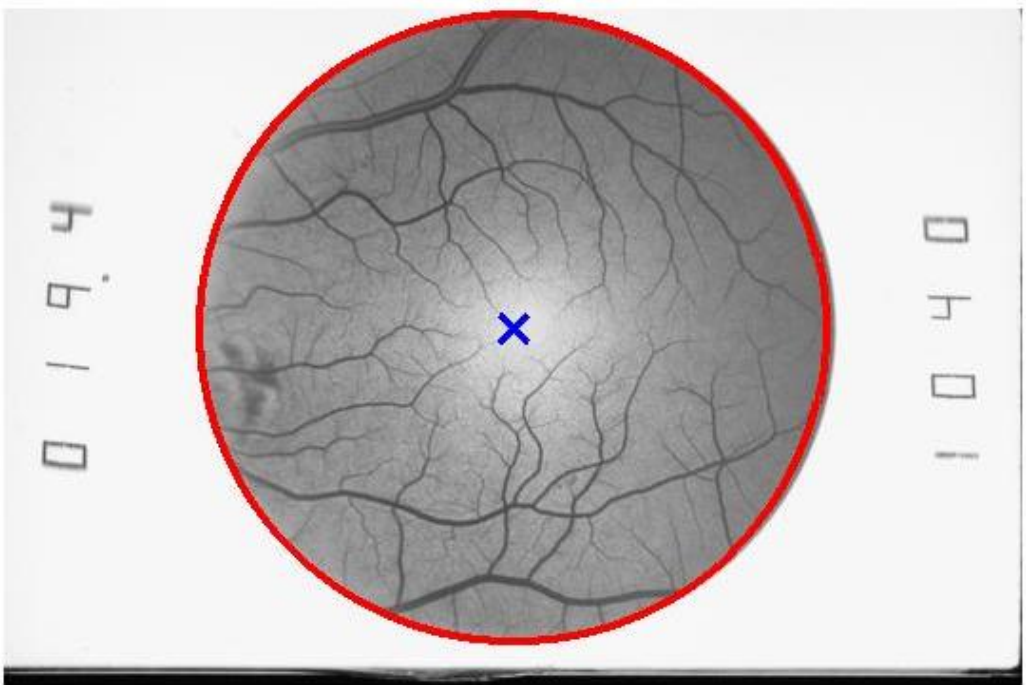


Figure 3.3: Retinal "field-stop" extraction using the Hough Transform. The detected retina boundary is accentuated in red, and the centre of the retina detected is marked in blue.

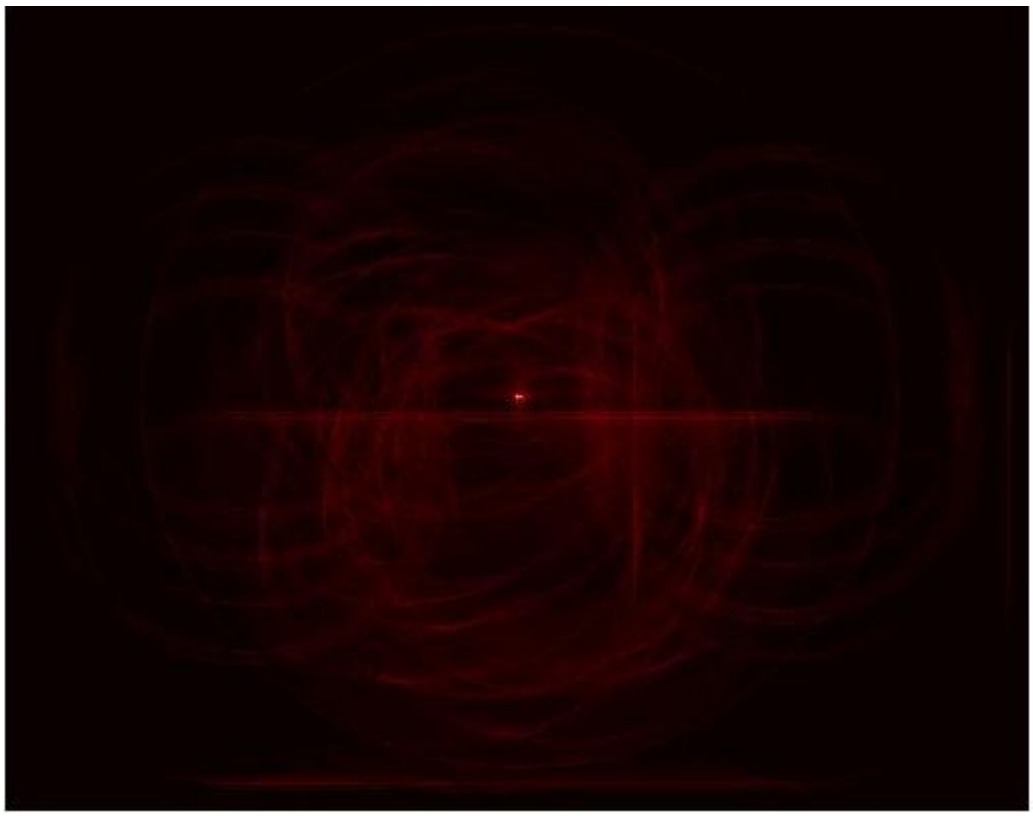


Figure 3.4: The parameter space of the input retinal image.

pixel intensities can boost local intensity contrast and improve the corner strength output of the Harris corner detector, described in subsequent Step 3.

We implement **histogram equalization** technique, a gray-scale transformation for contrast enhancement, on pixels located within the extracted "field-stop" from the previous step. This adjustment works on the principle of enhancing contrast for brightness values close to histogram maxima, and decreases contrast near maxima, thus spreading out the most frequent intensity values to better distribute the pixel intensity histogram (see Figure Figure 3.5).

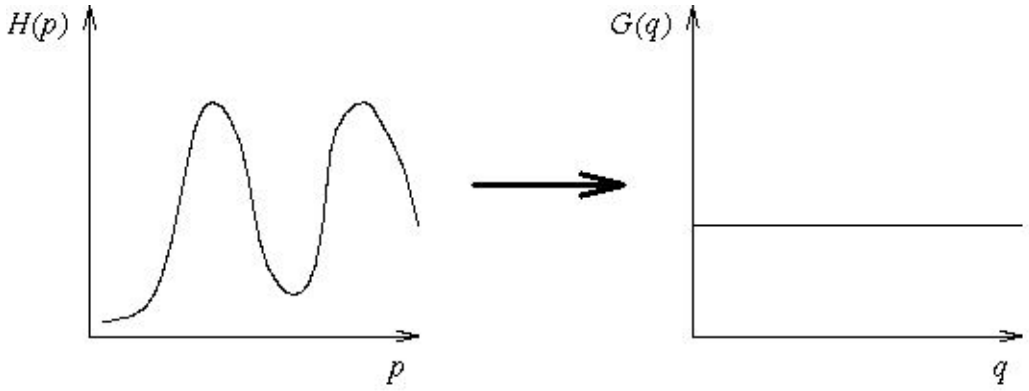


Figure 3.5: Histogram equalization [8].

A pixel brightness transformation $\mathcal{T}$ of the original image with brightness scale $[p_0, p_k]$ to the enhanced image with brightness scale $[q_0, q_k]$ can be defined as:

$$q = \mathcal{T}(p) \tag{3.5}$$

We denote the input histogram by $H(p)$ and the desired output histogram by $G(q)$, and treat the histograms as discrete distribution functions. We know from the monotonic property of the transform $\mathcal{T}$ that:

$$\sum_{i=0}^{k} G(q_i) = \sum_{i=0}^{k} H(p_i) \tag{3.6}$$

For image with N rows and N columns, the equalized histogram $G(q)$ of uniform probability density can be written as:

$$G(q) = \frac{N^2}{q_k - q_0} \tag{3.7}$$

Equation 3.6 now becomes:

$$\sum_{i=0}^{k} G(q_i) = \sum_{i=0}^{k} \frac{N^2}{q_k - q_0} = \frac{N^2(q - q_o)}{q_k - q_0} = \sum_{i=0}^{p} H(i) \tag{3.8}$$

The desired pixel brightness transformation $\mathcal{T}$ can now be derived as:

$$q = \mathcal{T}(p) = \frac{q_k - q_0}{N^2} \sum_{i=p_0}^{p} H(i) + q_0 \tag{3.9}$$

Algorithm 2 Histogram Equalization Pseudo-code

1. For an $N \times M$ images of G gray-scale levels, create an array H of length G initiated with 0 values.
2. Form the image histogram: Scan every pixel and increment the relevant member of H. For pixel p with intensity value g_p, perform:
$$H[g_p] = H[g_p] + 1$$
3. Form the cumulative image histogram H_c
$$H_c[0] = H[0] H_c[p] = H_c[p-1] + H[p] \quad p = 1, 2, ..., G-1$$
4. Set
$$T[p] = round\left(\frac{G-1}{NM} H_c[p]\right)$$
Note, this step offers a more efficient implementation by constructing a look-up table of the multiples of NM, and make comparison with the monotonically increasing values in H_c.
5. Rescan the image and write an output image with gray-scale levels g_q as:
$$g_q = T[g_p]$$

Step Three

We use a feature-based approach to ensure robustness and reliability in the presence of between-image geometric rotation and scaling, as described in prior sections. Generically, the bifurcation points or vessel crossings of a vascular tree (blood vessels) are often regarded as a

good source of landmark points (features) [64]. The choice of point features here is not only well-perceived by a human observer (hence easier for validation), but also allows higher invariance to imaging geometry relative to region or line features. We select the **Harris corner measures** [111] as indicators for vessel bifurcations and crossings distributed across the angiogram based on the conclusions from a comparative study [113] on 6 point detectors. Harris-based detectors demonstrated the best performance for image rotation, illumination variation and viewpoint change, against intensity- and contour-based detectors. We further modify the corner threshold to represent the spatial average of the Harris corner measure within the field-stop. This improves the algorithm's adaptability and automation.

The Harris corner detection can be explained as follows. Take an image patch from image f with image intensity I over the area (x, y) and shift it by $[u, v]$ so that it overlaps a second patch. The **weighted sum of squared differences** (SSD) between these two patches, denoted by E, is given by:

$$E(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2 \tag{3.10}$$

where $w(x, y)$ is a window function. In the case of Harris corner detector, Gaussian function with unit amplitude and zero mean is selected.

This can be expressed as:

$$w(x, y) = exp\left(- \frac{(x^2 + y^2)}{2\sigma^2} \right) \tag{3.11}$$

where σ denotes the standard deviation of the function.

Consider all small shifts, $E(u, v)$ can be approximated by Taylor's expansion. Let I_x and I_y be the partial derivatives of I, we can write:

$$E(u, v) = \sum_{x,y} w(x, y)[I_x u + I_y v + O(u^2, v^2)]^2 \tag{3.12}$$

This can be further written as:

$$\begin{aligned} E(u,v) &= Au^2 + 2Cuv + Bv^2 \\ A &= \sum_{x,y} w(x,y) I_x^2(x,y) \quad (3.13) \\ B &= \sum_{x,y} w(x,y) I_y^2(x,y) \quad (3.14) \\ C &= \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y) \quad (3.15) \end{aligned}$$

For small shift $[u,v]$, we can also express the equations above using bilinear approximation as:

$$E(u,v) \approx [u,v] M \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.16)$$

where the **Harris matrix** M is a 2x2 matrix computed from image derivatives where:

$$M = \sum_{x,y} W(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (3.17)$$

A corner is thus characterized by a large variation of E in all directions. An illustration of the partial derivatives I_x and I_y of neighbourhoods containing different structures (line edge, no interest points, corner) is shown in Figure 3.6.

By using eigenvalue analysis on the Harris matrix M, we can classify image points using criteria below based on the magnitude of the eigenvalues λ_1 and λ_2 (illustrated by Figure 3.7):

- If $\lambda_1 \approx 0$ and $\lambda_2 \approx 0$ then the pixel has no features of interest and is considered as "flat".
- If one eigenvalue has a large positive value while the other ≈ 0, then an "edge" is found.
- If both λ_1 and λ_2 have large positive values, then a "corner" is found.

Often, rather than exact calculation of the eigenvalue decomposition, a measure of corner

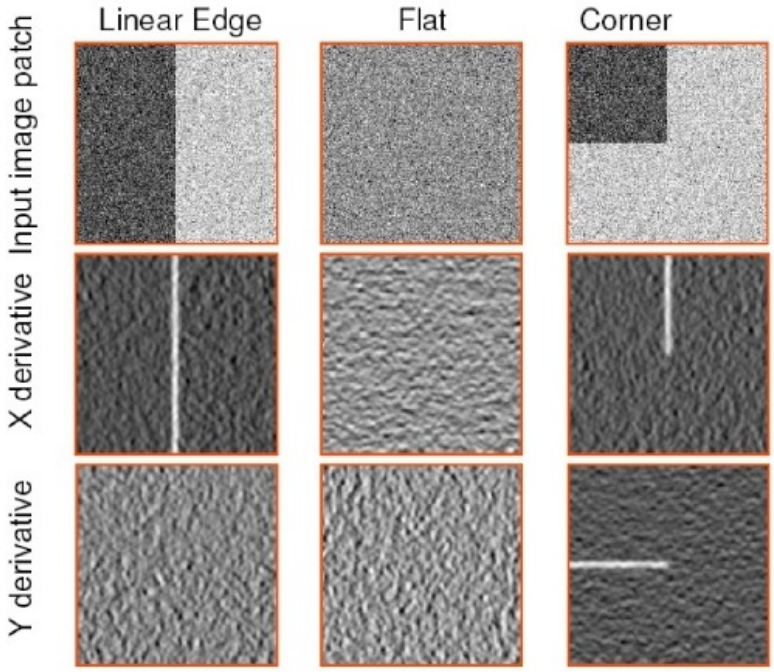


Figure 3.6: The partial derivatives, I_x and I_y, of three image patches containing different structural characteristics, adapted from [9].

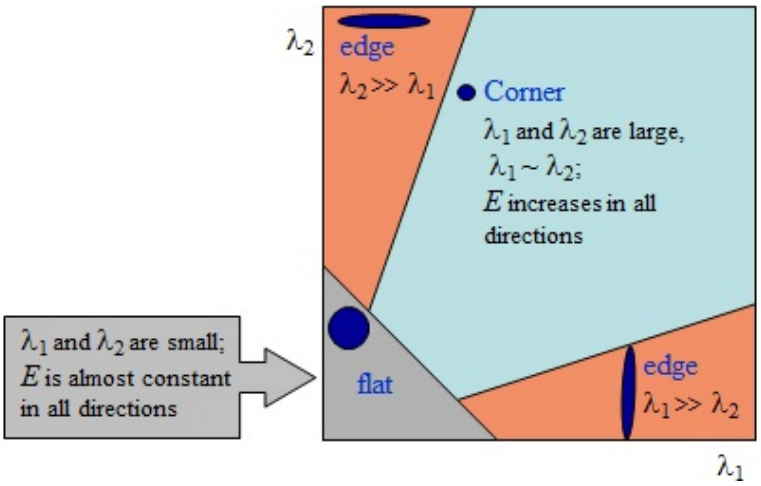


Figure 3.7: Classification of image points using eigenvalues of Harris matrix, adapted from [9].

response R is applied to evaluate the determinant and trace of Harris matrix to find corners, where κ is a tunable sensitivity parameter (feasible range from 0.04 to 0.15):

It is defined as:

$$R = \det M - \kappa(traceM)^2 \tag{3.18}$$

where

$$\begin{aligned} \det M &= \lambda_1 \lambda_2 \\ traceM &= \lambda 1 + \lambda_2 \end{aligned} \tag{3.19}$$

The R value for different structures is illustrated in Figure 3.8.

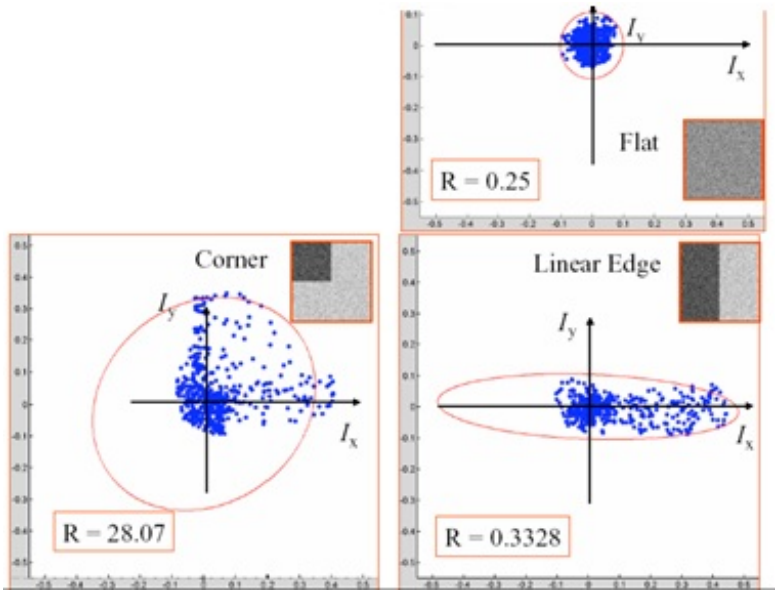


Figure 3.8: The distribution of the partial derivatives are very different for the three different types of image patches/neighboughoods. This allows us to characterize each type of structure using the size and shape of the principal component ellipse. The R value for each structure type is also significantly different.

We can summarize the Harris detection as follows (Algorithm 3):

Step Four

At step 4, these feature points are then putatively matched using **Normalized Cross-Correlation** [114] [115] of the neighbourhood pixel intensities and orientations at each feature points. This This N^2 comparison process filters out implausible correspondence feature pairs and reduces the computation load at the subsequent steps.

Cross-correlation is a measure of similarity of two signals. The normalization component is intended to partially remove the global illumination discrepancy between frames.

Algorithm 3 Harris Corner Detection Pseudo-code

1. Compute x and y derivatives of image by convolving the image with Gaussian kernels.

$$I_x = G_\sigma^x * I \quad I_y = G_\sigma^y * I$$

2. Compute products of derivatives at each pixel.

$$I_{x^2} = I_x \cdot I_x \quad I_{y^2} = I_y \cdot I_y \quad I_{xy} = I_x \cdot I_y$$

3. Compute the sums of the products of derivatives S at each pixel.

$$S_{x^2} = G_{\sigma'} \cdot I_{x^2} \quad S_{y^2} = G_{\sigma'} \cdot I_{y^2} \quad S_{xy} = G_{\sigma'} \cdot I_x y$$

4. Define at each pixel (x, y) the Harris matrix M.

$$M(x,y) = \begin{bmatrix} S_{x^2}(x,y) & S_{xy}(x,y) \\ S_{xy}(x,y) & S_{y^2}(x,y) \end{bmatrix}$$

5. Compute the response of the detector at each pixel.

$$R = \det M - \kappa(traceM)^2 \tag{3.20}$$

6. Threshold on the value of R. Compute non-maximum suppression.

For each extracted feature points, we first denote the rxr neighbourhood positioned at the feature point location (u, v) as t. For matching image f, the **correlation coefficient** γ can be defined as:

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y) - \overline{f}_{u,v}][t(x-u, y-v) - t]}{\sqrt{\sum_{x,y}[f(x,y) - \overline{f}_{u,v}]^2 \sum_{x,y}[t(x-u, y-v) - \overline{t}]^2}} \tag{3.21}$$

where $\overline{f}_{u,v}$ is the mean of $f(x, y)$ in the region under the neighbourhood, and $\overline{t}$ is the mean of the neighbourhood.

Equation 3.21 allows us to build a correlation matrix that holds the correlation strength between every feature point extracted in one frame and every feature points extracted in the other frame. Only the pairs that correlate strongly between the two frames are used as "control-point" to estimate a global-mapping transformation model.

Step Five

We then apply **"RANdom SAmple Consensus"** [108], abbreviated as **RANSAC**, at step 5, to iteratively estimate the "best-fit" global-mapping transformation model applicable to most putative matches. It is a non-deterministic scheme that produces a reasonable result only with a certain probability, with this probability increasing as more iterations are allowed. We set the "Number of Maximum Trials" at 1000 for such compromise between likelihood and speed. And we use the "symmetric transfer error" [115] of the homography applied to the matched "feature" points as the cost measure for optimization.

Here, RANSAC is applied as an internal mechanism for detecting and rejecting gross errors due to limitations from local feature detection. The assumption here is that the data consists of "inliers" which are data whose distribution can be explained by some set of model parameters, and "outliers" which are data that do not fit the model. Thus, rather than optimizing the fit of a functional description (the model) to all of the data present, it is more optimal to fit such functional description to only the "inliers". Another assumption here is that given a (usually small) set of inliers, there exists a procedure which can estimate the parameters of a model that optimally explains or fits this data. A simple example of model-fitting without and with RANSAC can be shown in Figure 3.9.

An addition note is that local feature detectors have two sources of errors: measurement or classification errors. Measurement errors occur when the feature detector correctly identifies the feature, but slightly miscalculates one of its parameters (e.g., its image location). Such errors generally follow a normal distribution, and therefore the conventional smoothing assumption (that the maximum expected deviation of any datum from the assumed model is a direct function of the size of the data set, and thus regardless of the size of the data set, there will always be enough good values to smooth out any gross deviations) is applicable to this case. Classification errors occur when a feature detector incorrectly identifies a portion of an image as an occurrence of a feature. These gross errors are not counteracted from averaging hence an iterative method such as RANSAC can function well to smooth the data.

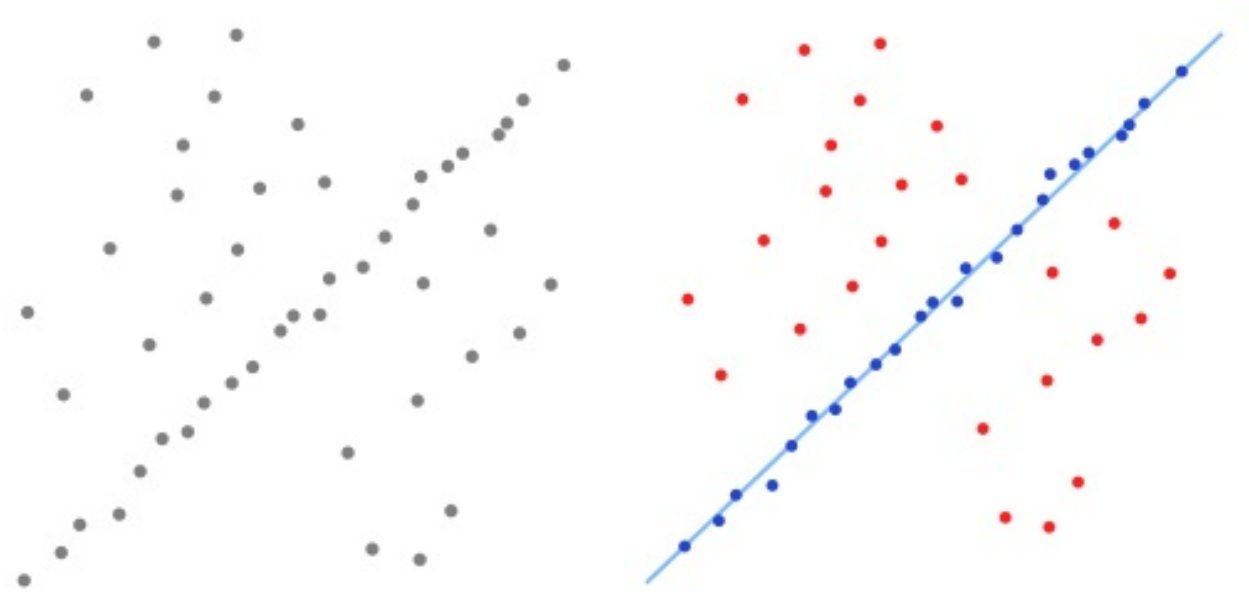


Figure 3.9: Left: A data set with many outliers for which a line has to be fitted: Right: Fitted line (in blue) using RANSAC. Outliers (in red) have no influence on the result.

"RANSAC" is formally stated as follows:

- Given a model that requires a minimum of n data points to instantiate its free parameters, and a set of data points P such that the number of points in P is greater than n [$\#(P) \geq n$], randomly select a subset S_1 of n data points from P and instantiate the model. Use the instantiated model $M1$ to determine the subset S_1^* of points in P that are within some error tolerance of M_1. The set S_1^* is called the consensus set of S_1.

- If $\#(S_1^*)$ is greater than some threshold T, which is a function of the estimate of the number of gross errors in P, use S_1^* to compute (possibly using least squares) a new model M_1^*.

- If $\#(S_1^*)$ is less than T, randomly select a new subset S_2 and repeat the above process. If, after some predetermined number of trials, no consensus set with T or more members has been found, either solve the model with the largest consensus set found, or terminate in failure.

- Single error tolerance (for model compatibility) for all data is often sufficient as variation within error tolerances is relatively small compared to the size of a gross error.

The idea behind RANSAC is to use as small an initial data set as feasible and enlarges this set with consistent data when possible. This results in robust estimation of the model parameters with a high degree of accuracy even when significant amount of "outliers" are present in the data set. Here, "outliers" could be due to presence of noise or erroneous measurements or incorrect hypotheses about the interpretation of data; while "inliers" are data whose distribution can be explained by some set of model parameters.

The pseduo-code for RANSAC is summarised in Algorithm 4. Importantly, the values of parameters t (the threshold value or error tolerance for establishing the fit of the data to the model) and d (the number of close data values required to assert that a model fits well to data) are determined from the application and the data-set specific requirements, often based on experimental evaluation.

Though the parameter k (the number of maximum iteration trials performed by the algorithm) can be determined theoretically. Let p_a be the probability that the RANSAC algorithm in some iteration selects only inliers from the input data set when it chooses the n points from which the model parameters are estimated. Let p_s be the probability of choosing an inlier each time a single point is selected. Assuming that the n points needed for estimating a model are selected independently, then p_s^n is the probability that all n points are inliers and $1 - p_s^n$ is the probability that at least one of the n points is an outlier, a case which implies that a bad model will be estimated from this point set. Thus, we can write:

$$1 - p_a = (1 - p_s^n)^k \tag{3.22}$$

The right hand side of the Equation 3.22, $(1 - p_s^n)^k$, denotes the probability that the algorithm never selects a set of n points which are all inliers, which can also be written as $1 - p_a$, the left hand side of the Equation 3.22.

Now this leads to:

$$k = \frac{log(1 - p_a)}{log(1 - p_s^n)} \tag{3.23}$$

Algorithm 4 RANSAC pseudo-code

input:
data - a set of observations
model - a model that can be fitted to data
n - the minimum data points required to fit the model
k - the number of maximum iteration trials performed by the algorithm
t - the threshold value or error tolerance for establishing the fit of the data to the model
d - the number of close data values required to assert that a model fits well to data

output:
best_model - model parameters which best fit the data (or nil if no good model is found)
best_consensus_set - data points from which this model has been estimated
best_error - the error of this model relative to the data

```
iterations := 0
best_model := nil
best_consensus_set := nil
best_error := infinity

while iterations < k
    maybe_inliers := n randomly selected points from data
    maybe_model := parameters fitted to maybe_inliers
    consensus_set := same as maybe_inliers

     for every point in data not in maybe_inliers
          if point fits maybe_model with an error smaller than t
              add point to consensus_set
     if the number of elements in consensus_set is > d
    (this implies that we may have found a good model, now test how good it is)
        this_model := model parameters fitted to all points in consensus_set
        this_error := a measure of how well this_model fits these points
         if this_error < best_error
        (This model is better than all previous ones, keep it until a better one is found)
            best_model := this_model
            best_consensus_set := consensus_set
            best_error := this_error
    increment iterations

return best_model, best_consensus_set, best_error
```

Equation 3.23 assumes that the n data points are selected independently, meaning once a point is selected once it must be excluded from subsequent iterations. When in practice it is unlikely that the points are replaced immediately between iterations, hence the derived k value is often taken as an upper limit.

Step Six

At step 6, we define the *inliers-to-outliers ratio* as the comparison between the number of pairs that can be described by the "best-fit" model within an offset threshold and those that cannot. This ratio helps to select the appropriate transformation model in a hierarchical manner, from lower-order transformation model [115] to higher-order transformation model [64].

First, we denote the misaligned **moving image** as $I_M(\boldsymbol{x})$; and the reference **fixed image** is denoted as $I_F(\boldsymbol{x})$. The moving image and the fixed image are both of dimension d and are each defined on their own spatial domain: $\Omega_M \subset \mathbb{R}^d$ and $\Omega_F \subset \mathbb{R}^d$. The transformation $\mathbf{T}$ is defined as a spatial (coordinate) mapping from the moving image domain to the fixed image domain:

$$\mathbf{T} : \Omega_M \subset \mathbb{R}^d \Rightarrow \Omega_F \subset \mathbb{R}^d \tag{3.24}$$

The parametrization of the transform, denoted by $\mathbf{T}_{\boldsymbol{\mu}}$, is indicated by the inclusion of the subscript $\boldsymbol{\mu}$ [116]. The vector $\boldsymbol{\mu}$ contains the "transformation parameters". Table 3.1 summarizes the transformation model $\mathbf{T}_{\boldsymbol{\mu}}$ (in 2D) in the increasing order of flexibility, while Figure 3.10 illustrates the different transforms.

2D Transformations	$\mathbf{T}_{\boldsymbol{\mu}}$	$\boldsymbol{\mu}$
Translation	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{t}$	$\boldsymbol{\mu} = \mathbf{t}$
Rigid	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = R(\boldsymbol{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c}$	$\boldsymbol{\mu} = (\theta_z, t_x, t_y)^T$
Similarity	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = sR(\boldsymbol{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c}$	$\boldsymbol{\mu} = (s, \theta_z, t_x, t_y)^T$
Affine	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \mathbf{A}(\boldsymbol{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c}$	$\boldsymbol{\mu} = (a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)^T$
Quadratic	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{\Theta}\mathbf{X} = \boldsymbol{\Theta}(\mathbf{x^2}, \mathbf{xy}, \mathbf{y^2}, \mathbf{x}, \mathbf{y}, \mathbf{1})^{\mathbf{T}}$	$\boldsymbol{\mu} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \end{pmatrix}$
B-spline	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \sum_{\boldsymbol{x}_k \in \mathcal{N}_{\boldsymbol{x}}} \mathbf{p}_k \boldsymbol{\beta}^3(\frac{\boldsymbol{x} - \boldsymbol{x}_k}{\boldsymbol{\sigma}})$	$\boldsymbol{\mu} = (p_{1x}, p_{2x}, ..., p_{P1}, p_{1y}, p_{2y}, ..., p_{P2})^T$
Thin-plate spline	$\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{A}(\boldsymbol{x}) + \mathbf{t} + \sum_{\boldsymbol{x}_k^{\text{fix}}} \mathbf{c}_k \mathbf{G}(\boldsymbol{x} - \boldsymbol{x}_k^{\text{fix}})$	$\boldsymbol{\mu} = (\boldsymbol{x}_{1x}^{\text{mov}}, \boldsymbol{x}_{1y}^{\text{mov}}, ..., \boldsymbol{x}_{Kx}^{\text{mov}}, \boldsymbol{x}_{Ky}^{\text{mov}})^T$

Table 3.1: Transformation Models

- **Translation**: The parameter vector $\boldsymbol{\mu}$ is simply defined by the translation vector $\mathbf{t}$, as a

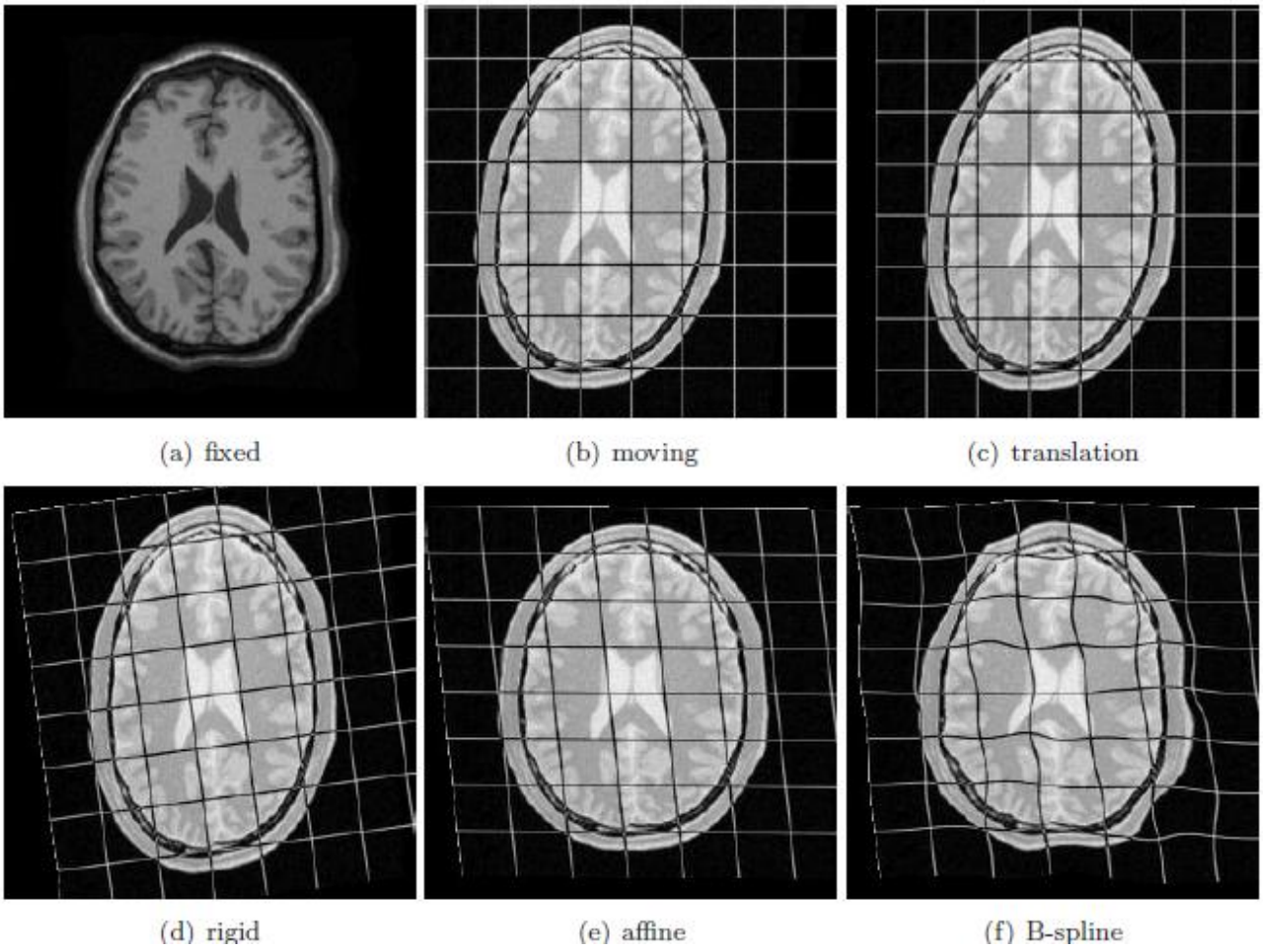


Figure 3.10: Different transformations. (a) the fixed image; (b) the moving image with a grid overlayed; (c) the moving image aligned using a translation transformation; (d) aligned using a rigid transformation; (e) aligned using an affine transformation; f) aligned using a B-spline transformation. The overlay grids indicate the deformations imposed on the moving image. Adapted from [10].

spatial shift between coordinates.

- **Rigid**: The image is treated as a rigid body, which can translate and rotate, thus preserves angles and curvature. However, it cannot be scaled or stretched. The transformation model $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{t}$ consists of the rotation matrix R, the centre of rotation $\mathbf{c}$ and the translation matrix $\mathbf{t}$. As the rotation matrix is parameterised by the Euler angle(s), the parameter vector $\boldsymbol{\mu}$ consists of the Euler angle(s) and the translation vector $\mathbf{t} = (t_x, t_y)$. R_z denotes the rotation around the axis normal to the image. Note that the centre of rotation, $\mathbf{c}$, (usually the centre of the image) is not part of $\boldsymbol{\mu}$ as it is fixed.

- **Similarity**: This transformation characterizes the scaled displacements or shift along both horizontal and vertical directions. This means the image is treated as an object, which can translate, rotate, and scale isotropically. The transformation model $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = sR(\boldsymbol{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c}$ consists of an isotropic scalar factor s, the rotation matrix R, the centre of rotation $\mathbf{c}$ and the translation matrix $\mathbf{t}$. The rotation matrix is parameterised by a Euler angle R_z in 2D, same as the case of rigid transformation. Thus, the parameter vector $\boldsymbol{\mu}$ consists of the isotropic scaling factor, the Euler angle and the translation vector.

- **Affine**: Affine models map a parallelogram onto a square while preserving straight lines. The image can be translated, rotated, scaled and sheared. The transformation model $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \mathbf{A}(\boldsymbol{x} - \mathbf{c}) + \mathbf{t} + \mathbf{c}$ has a more versatile matrix $\mathbf{A}$ with respect to the prior similarity transform. The parameter vector $\boldsymbol{\mu}$ is formed by the matrix elements a_{ij} and the translation vector.

- **Quadratic**: The quadratic surface model allows 12 degrees of freedom, the coefficients θ_{ij} sums up the contribution from weak-perspective camera projection between frames, the geometric estimate for the near spherical shape of the retina surface and a rigid transformation during sampling [64]. It generalizes the 2D affine transformation model induced by the rigid motion of a planar surface. For $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{\Theta}\mathbf{X}$, $\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_1^T \\ \boldsymbol{\Theta}_2^T \end{pmatrix}$ where $\boldsymbol{\Theta}_1^T = (\theta_{11}, ..., \theta_{16})$ and $\boldsymbol{\Theta}_2^T = (\theta_{21}, ..., \theta_{26})$. And $\mathbf{X} = (x^2, xy, y^2, x, y, 1)^T$.

- **B-spline**: B-spline is a common parametrization for the non-rigid transformations [94]. $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \sum_{\boldsymbol{x}_k \in \mathcal{N}_{\boldsymbol{x}}} \mathbf{p}_k \boldsymbol{\beta}^3(\frac{\boldsymbol{x} - \boldsymbol{x}_k}{\boldsymbol{\sigma}})$ where $\boldsymbol{x}_k$ are the control points, $\boldsymbol{\beta}^3(\boldsymbol{x})$ are the cubic multidimensional B-spline polynomial, $\mathbf{p}_k$ are the B-spline coefficient vectors, $\boldsymbol{\sigma}$ are the B-spline control point spacing, and $\mathcal{N}_{\boldsymbol{x}}$ is the set of all control points within the compact support of the B-spline at $\boldsymbol{x}$. The control points $\boldsymbol{x}_k$ are defined on a regular grid, overlayed on the fixed image. The control point grid is defined by the amount of space between the control points $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_d)$ (with d the image dimension), which can be different for each direction. B-splines have local support (as $|\mathcal{N}_{\boldsymbol{x}}|$ is small), which means that the transformation of a point can be computed from only a couple of

surrounding control points. This is beneficial both for modelling local transformations, and for fast computation. The parameters $\boldsymbol{\mu}$ are formed by the B-spline coefficients $\mathbf{p}_k$. The number of control points $\mathbf{P} = (P_1, ..., P_d)$ determines the number of parameters M, by $M = (P_1 \times ... \times P_d) \times d$. P_i in turn is determined by the image size s and the B-spline grid spacing: $P_i \approx s_i/\sigma_i$, as some additional control points are placed just outside the image. The parameter vector $\boldsymbol{\mu}$ in 2D is composed as: $\boldsymbol{\mu} = (p_{1x}, p_{2x}, ..., p_{P1}, p_{1y}, p_{2y}, ..., p_{P2})^T$.

- **Thin-plate spline**: Thin-plate splines are another well-known representation for nonrigid transformations. The transformation is based on a set of K corresponding landmarks in the fixed and moving image: $\boldsymbol{x}_k^{\text{fix}}$ and $\boldsymbol{x}_k^{\text{mov}}$ for $k = 1, ..., K$. The transformation is expressed as a sum of an affine component and a nonrigid component such that $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{x} + \mathbf{A}(\boldsymbol{x}) + \mathbf{t} + \sum_{\boldsymbol{x}_k^{\text{fix}}} \mathbf{c}_k \mathbf{G}(\boldsymbol{x} - \boldsymbol{x}_k^{\text{fix}})$. $\mathbf{G}(\mathbf{r})$ is a basis function and c_k are the coefficients corresponding to each landmark. The coefficients c_k and the elements of $\mathbf{A}$ and $\mathbf{t}$ are computed from the landmark displacements $\mathbf{d}_k = \boldsymbol{x}_k^{\text{mov}} - \boldsymbol{x}_k^{\text{fix}}$. The spline kernel transforms are often less efficient than the B-splines as they lack the compact support property of the B-splines, but they allow for more flexibility in placing the control points $\boldsymbol{x}_k^{\text{fix}}$ and $\boldsymbol{x}_k^{\text{mov}}$. The moving landmarks form the parameter vector $\boldsymbol{\mu}$. Both landmark sets are needed to define a transformation. In order to perform a registration, only the fixed landmark positions are given by the user; the moving landmarks are initialized to equal the fixed landmarks, corresponding to the identity transformation, and are subsequently optimized. The parameter vector in 2D is composed as $\boldsymbol{\mu} = (\boldsymbol{x}_{1x}^{\text{mov}}, \boldsymbol{x}_{1y}^{\text{mov}}, ..., \boldsymbol{x}_{Kx}^{\text{mov}}, \boldsymbol{x}_{Ky}^{\text{mov}})^T$.

As we need a global model sufficiently flexible to reflect the spherical distortion introduced by retinal imaging geometry, yet computational efficient. We have therefore decided to implement the quadratic transformation with 12 degrees of freedom to counteract warping. However, the mapping with a higher-order model (quadratic) is sensible if the lower-order model (affine) registration has been successful.

To generalize the above six-step pairwise registration approach, we suggest that the ROC

(sensitivity versus 1-specificity) characteristic be used to decide what performance trade-offs are acceptable for True Positive (TP) and False Positive (FP) rates.

Figure 3.11 and Figure 3.12 compare between-frame pixel intensity differences without and **after** our pairwise registration scheme.

3.4.2 Joint: Hierarchical Compound of Transformation Matrices

To analyse temporal information both within a single retinal angiogram sequence (intra-sequence) and across two different sequences taken before and after the operation (inter-sequence), we need a step-by-step process that first aligns each pixel intra-sequentially then cross-aligns the same pixel inter-sequentially. Temporal registration requires maximizing the point correspondence between similar structural features, while still allowing us to differentiate, detect or even monitor pathological changes. For each patient, the last post-op frame was acquired several hours after the first pre-op frame. This increases the chance that the vasculature alters either in width or in curvature, in addition to the natural variability of blood vessels.

For the i^{th} frame in a sequence S of length N, let (x_i, y_i) denote the location of each pixel in the frame coordinates at time t_i. The image is denoted by the function:

$$f_i := f_i(x_i, y_i; t_i), \;\; i \in \{1, 2, ..., N\} \tag{3.25}$$

Consider two unregistered angiogram sequences $S^{(A)}$ and $S^{(B)}$, acquired with frame-specific spatial coordinates relative to the camera lens, at unknown points in time relative to the cardiac cycle, and with non-uniform frame-to-frame intervals (from less than a second to tens of seconds) along each sequence:

$$S^{(A)} = \left\{ f_i^{(A)} \right\}_{i=1,2,...,N_A} \text{and} \quad S^{(B)} = \left\{ f_i^{(B)} \right\}_{i=1,2,...,N_B} \tag{3.26}$$

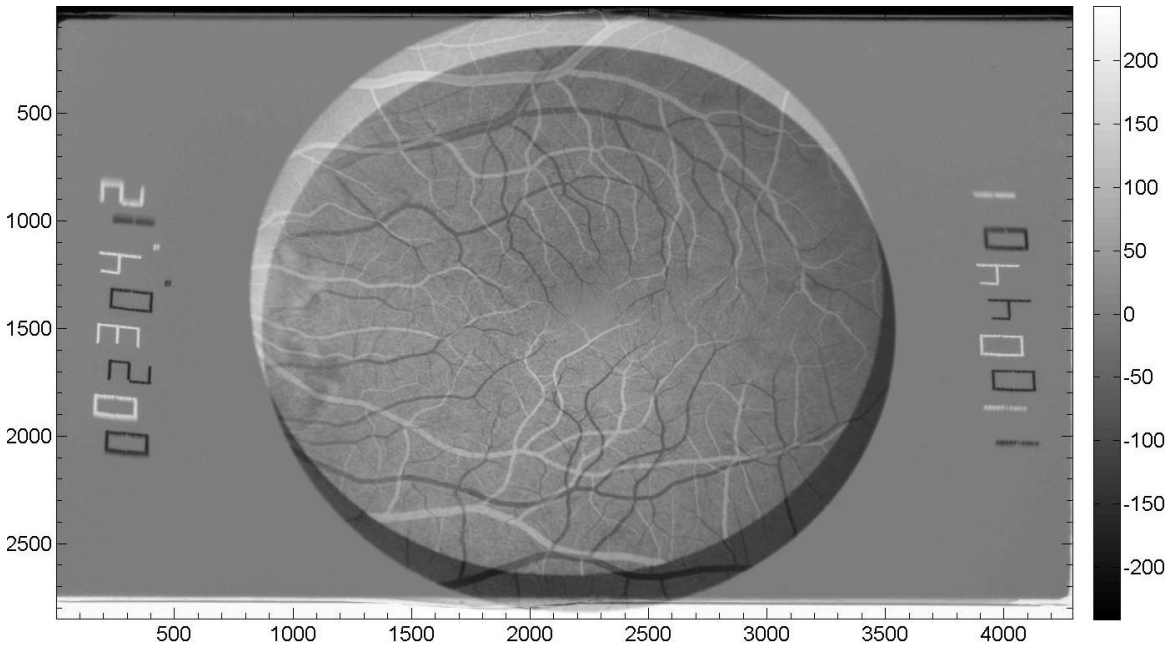


Figure 3.11: The intensity difference between two frames half a sequence apart **without** pairwise registration (SD = 32.48).

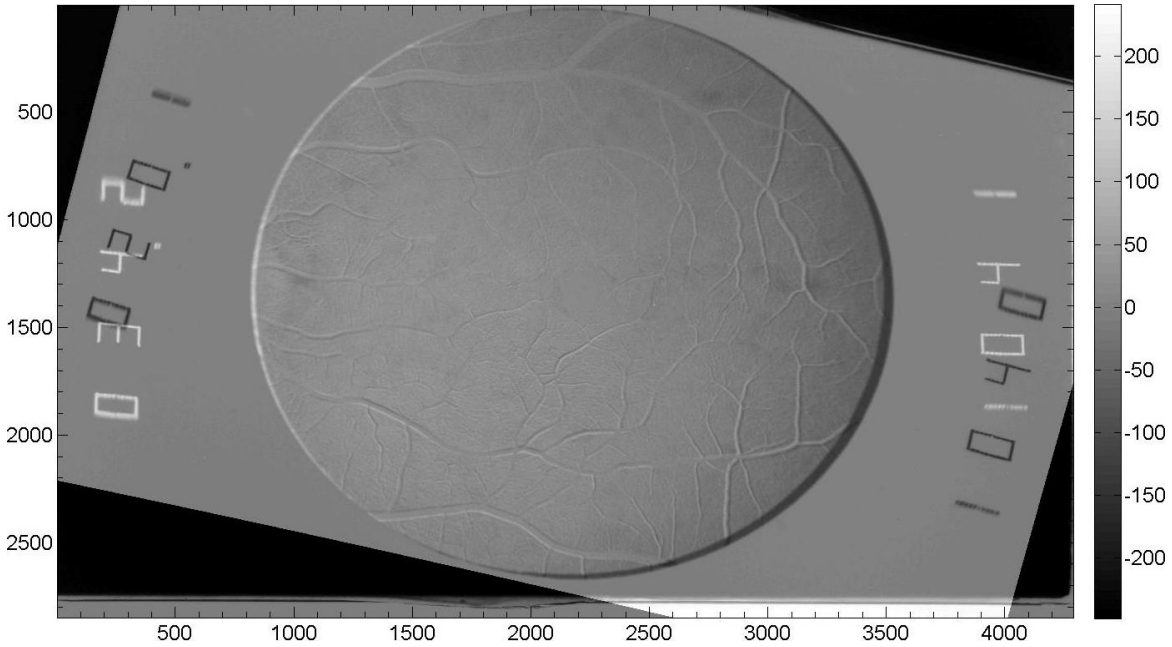


Figure 3.12: The intensity difference between the two frames **after** pairwise registration (SD = 19.54).

We first spatially register each individual image $f_i^{(A)}$ in $\mathrm{S}^{(A)}$ to a *local reference* spatial coordinate system defined by frame $f_{lr}^{(A)}$. This **individual-to-local reference** registration is also applied to sequence $\mathrm{S}^{(B)}$ with frame $f_{lr}^{(B)}$ as its local reference. We then register the two local references $f_{lr}^{(A)}$ and $f_{lr}^{(B)}$ separately to one *global reference* f_{gr}. This **local-to-global**

reference transformation is further combined with prior individual-to-local reference transformations to give the **individual-to-global reference** transformation that allows $S^{(A)}$ and $S^{(B)}$ to be co-registered to one global reference f_{gr}.

We adopt the clinical practice of selecting the darkest image as the *local reference*. Fluorescein angiograms are negative-contrasted where the higher the dye concentration the "darker" the frame appears. Thus, the darkest frame is essentially the frame with the highest vessel contrast and visibility. Our algorithm computes the sum of the pixel intensities within the field-stop and selects the frame at the peak of the dye-time course (when the image is the darkest) in each sequence:

$$f_{lr} = \{f_{i^*}\}, \quad i^* = \underset{i\in\{1,2,...,N\}}{\arg\min} \langle f_i, M_i \rangle \tag{3.27}$$

with $\langle \cdot, \cdot \rangle$ denoting a spatial inner product and M_i a binary spatial weighting function that is unity for points (x_i, y_i) within the field-stop region and 0 outside.

The ***individual-to-local reference*** registration is given as:

$$f_i' := f_i'\left(x'_{(lr)}, y'_{(lr)}; t_i\right) = R_{i(lr)}(f_i), \;\; i \in \{1, 2, ..., N\} \tag{3.28}$$

where individual frame f_i is mapped to the spatial coordinate system of local reference f_{lr} by function $R_{i(lr)}$. Note that (x', y') denotes the transformed pixel coordinates.

The ***local-to-global reference*** transformation is given as:

$$f_{lr}' := f_{lr}'\left(x'_{(gr)}, y'_{(gr)}; t_{lr}\right) = R_{lr(gr)}(f_{lr}) \tag{3.29}$$

where the local reference f_{lr} is mapped to the spatial coordinate system of global reference f_{gr} by function $R_{lr(gr)}$.

Lastly, the ***individual-to-global reference*** registration can be combined as:

$$f_i'' := f_i''\left(x''_{(gr)}, y''_{(gr)}; t_i\right) = R_{lr(gr)}\left(f_i'\right), \;\; i \in 1, 2, ..., N \tag{3.30}$$

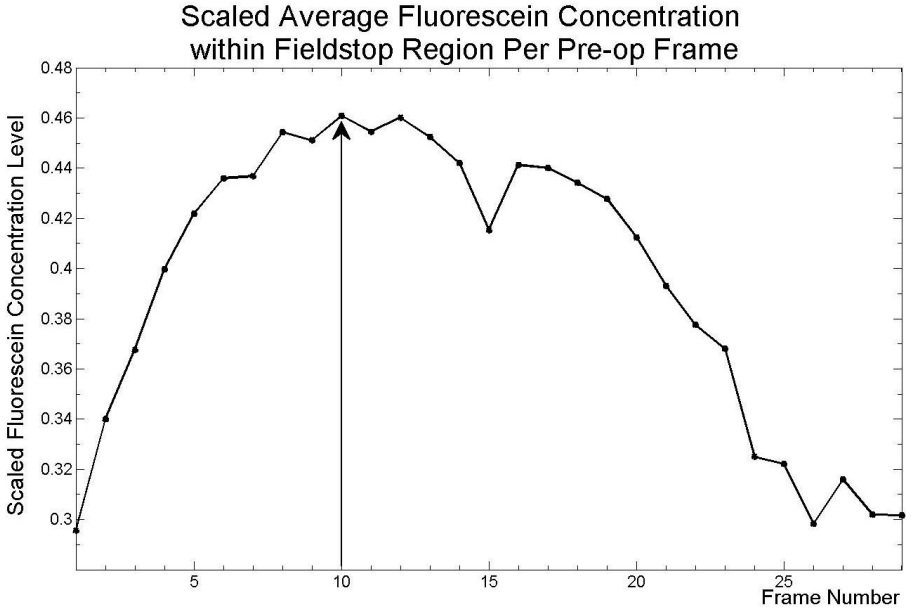


Figure 3.13: Scaled average fluorescein concentration within field-stop region per frame in the pre-operative intervention sequence against the frame number.

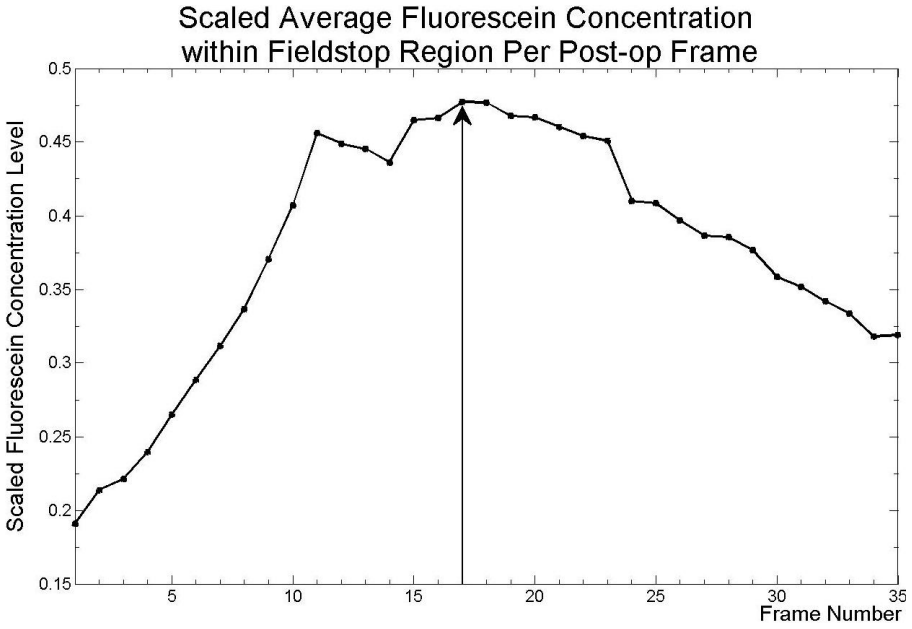


Figure 3.14: Scaled average fluorescein concentration within field-stop region per frame in the **post**-operative intervention sequence against the frame number.

where individual frame f_i is mapped to the spatial coordinate system of the global reference f_{gr}.

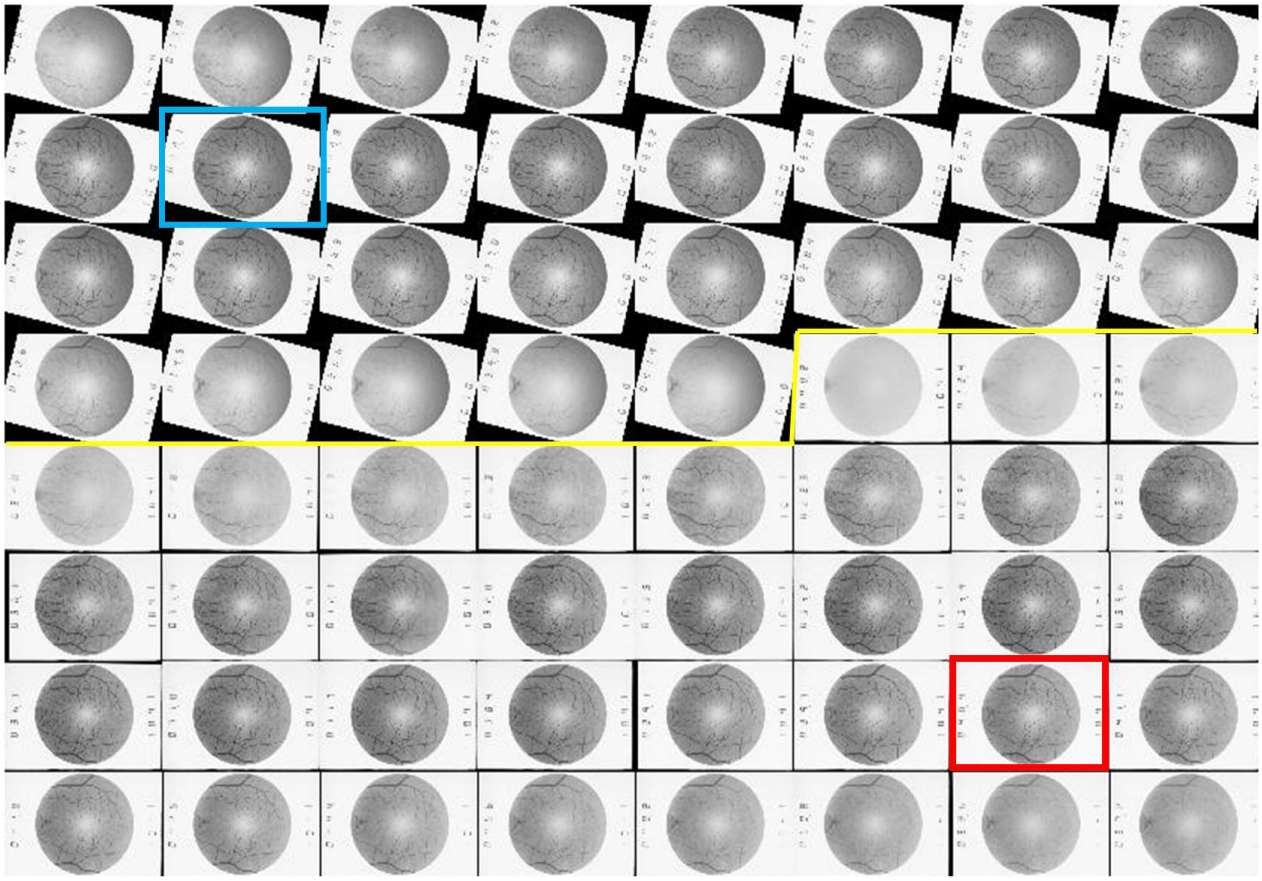


Figure 3.15: The global reference in pre-op sequences is framed by a blue box, and the global reference in post-op sequences is framed by a red box. They are both automatically determined from the prior intensity-time course plots.

3.5 Discussion and Conclusion

The purpose of 2D+t sequential retinal image registration is to spatio-temporally align set(s) of retinal images inter- and/or intra-sequentially to clinical assess disease-related ocular circulation. The aligned pre- and post-op sequences from 6 patients are presented in Figure 3.16–3.21 by the form of temporal pixel intensity average across spatially-aligned (with regards to one coordinate system shared in both pre- and post-op sequences for each patient). Notably, Figure 3.17–3.20 demonstrate that there exist a fair amount of coordinate and field-of-view shift between pre- and post- operative sequences. Yet our registration algorithm is able to perform a satisfactory results to map the coordinates without resorting to

state-of-the-art local-to-global bootstrapping methods that, though allow higher degree of transformation require higher computation power.

Another point to note is that, our "Joint Global RANSAC" method manages to remove manual bias in its automated reference frame selection by choosing the frame where fluorescein concentration peaks within field-stop which is the imperative approach adopted from clinical practices. (See Figure 3.13, 3.14 and 3.15).

In this chapter, we first present a literature review on existing methods for retinal registration. We then present an explanation and justification of image processing techniques we employed in our 6-step **pairwise** registration that incorporates projective RANSAC with quadratic homography transforms. This is then followed by our **joint** hierarchical compound of transformation matrices that eventually allow us to align fluorescein angiogram sequences taken before and after operation with spatio- and temporal variation on top of the exiting geometric distortion, radiometric degradation, and additive noise corruption between any two frames. Finally, we conclude the chapter by presenting the 12 aligned angiogram sequences taken from 6 patients, adopted from film-based images (4288×2848 resolution with approximately 32 frames per angiogram sequence) taken more than 20 years ago with varying temporal intervals.

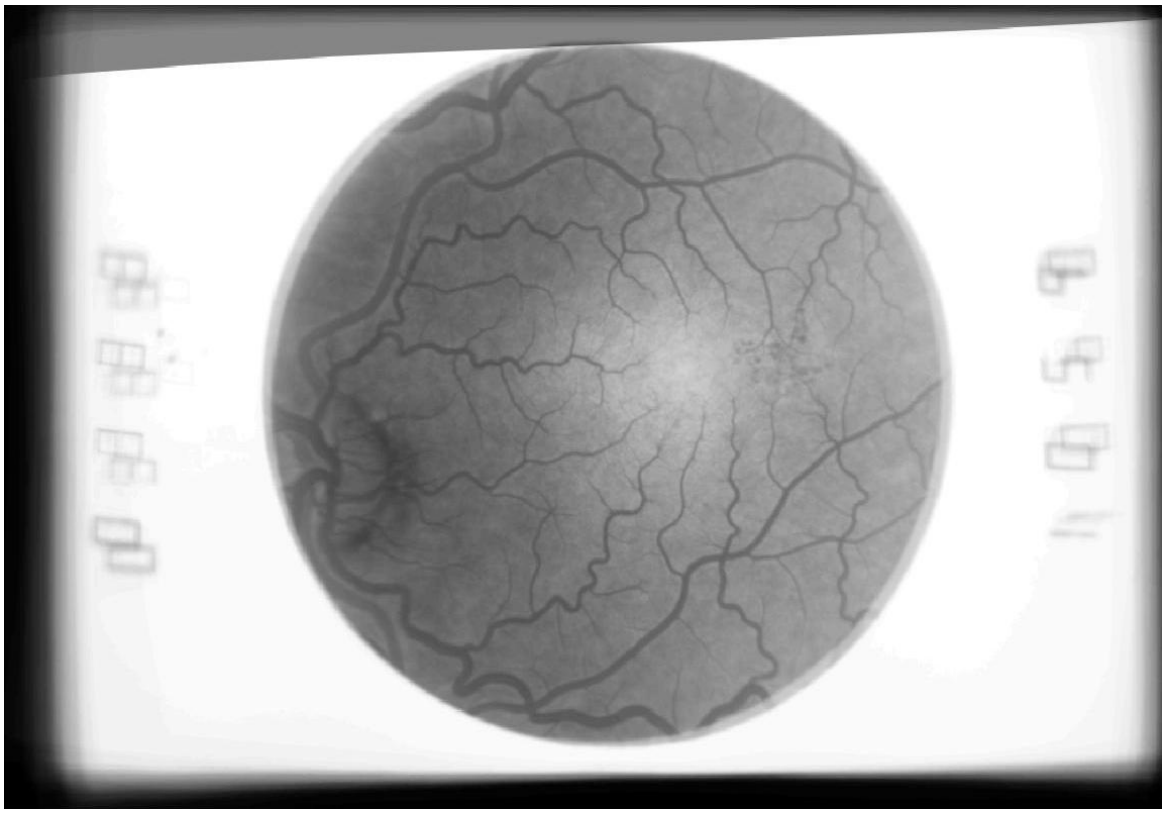


Figure 3.16: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 1.



Figure 3.17: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 2.

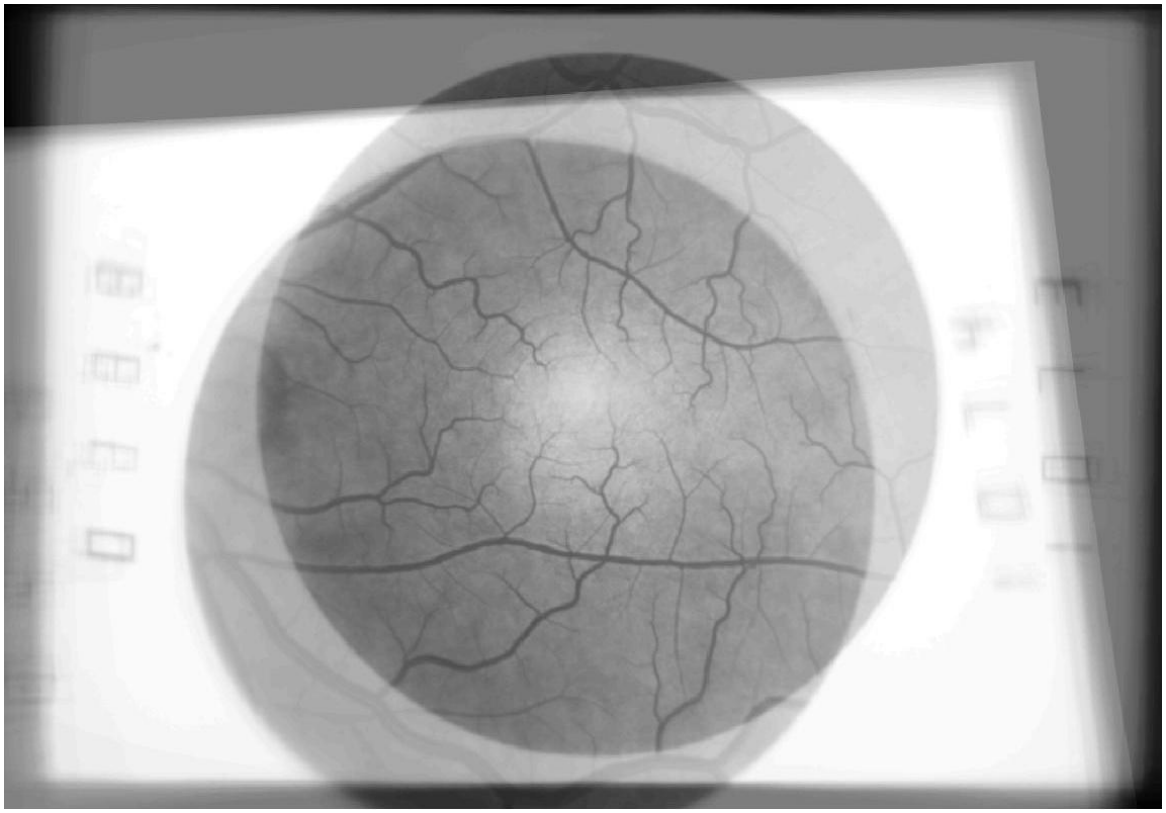


Figure 3.18: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 3.

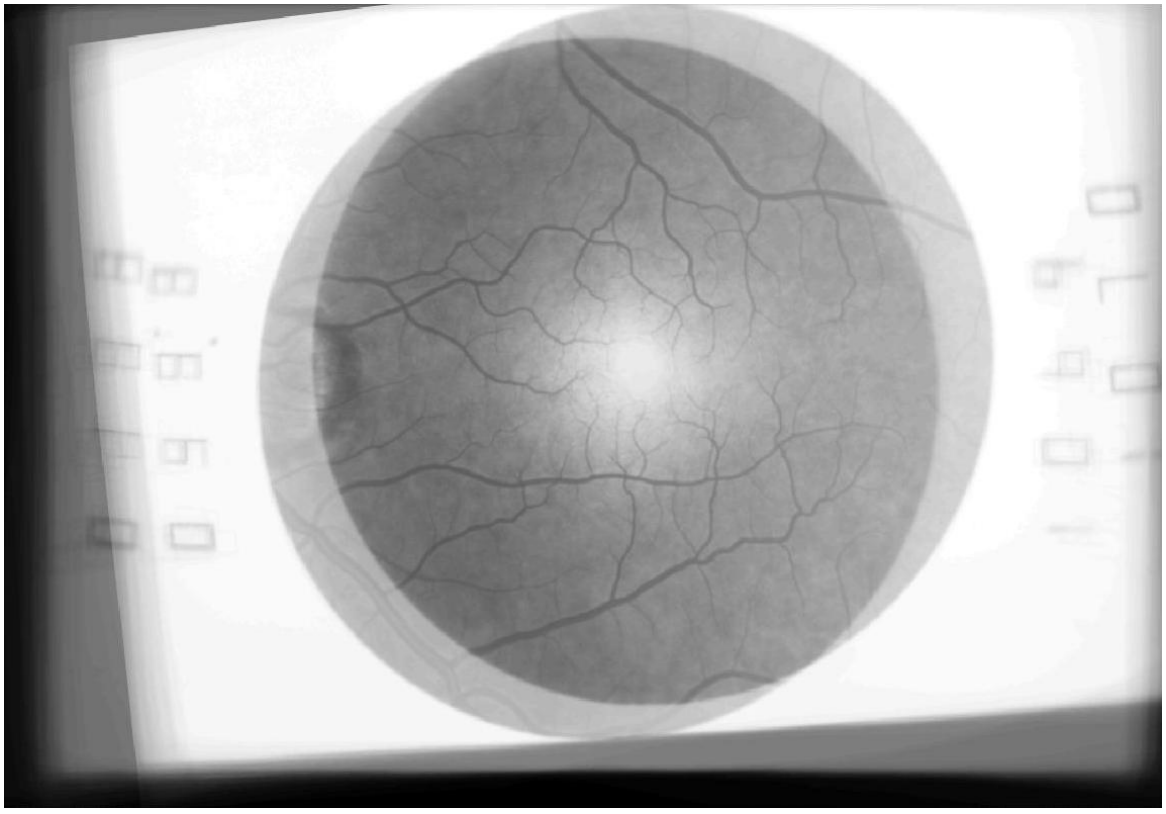


Figure 3.19: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 4.



Figure 3.20: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 5.

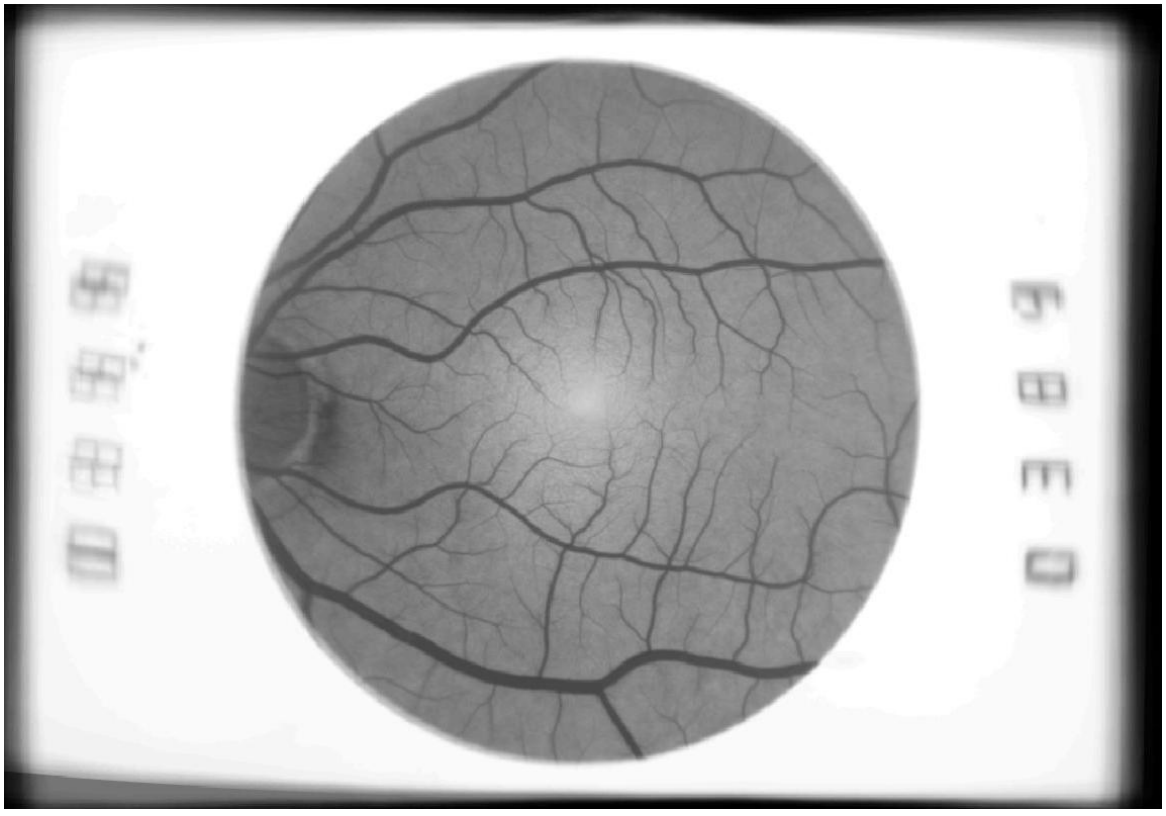


Figure 3.21: Temporal average $f_{average}$ of registered pre- and post-op sequences for Patient 6.

Chapter 4

2D+t Sequential Fluorescein Angiogram II: Centreline Extraction and Tracking

4.1 Introduction

The morphological features of retinal tree contains beneficial and diagnostic information in biomedical and clinical research (refer back to Section 1.2.1 in Chapter 1 for detailed review), and it is found to be unique for each individual and can be used for biometric identification [117]. The manual measurement of the retinal vasculature is a time consuming process that entails training and the skill of the operator, thus the use of automated analysis is becoming more popular among both research and clinical communities. However, fine microvasculature often exhibit poor contrast-to-noise ratio, posing a bigger challenge for segmentation [118].

In this chapter, we first explain the concept of curvilinear structure ("line") in the context of image processing in Section . We then review the existing techniques for (micro-)vessel centreline detection and tracking in 4.3. This is then followed by the introduction of our scale-space approach for search-based centreline detection, tracking and extraction in Section

4.4. Lastly, we discuss the benefits of our approach and conclude this chapter in Section 4.5.

4.2 Context

Detection and extraction of curvilinear structures ("line") is a crucial task in image processing. An earlier simplified view of only taking into account the pixel intensity difference on and off the "line" often leads to false hypotheses for line points. An alternative view regards "lines" as structures with parallel edges is the basis of common Gaussian-based methods, where the edge detection filters are applied in the direction perpendicular to the line and the non-linear combination of each filer responses yield the final response of the operator. An extension of this view give rises to the assumption that ridges and ravines in an image indicates the presence of "lines" [49]. Therefore an earlier definition for vessel centerlines as the connected sets of pixels which correspond to intensity minima/maxima (depending on the image modality) computed from the intensity profiles of the vessel cross sections has also been updated. The view that centreline is the geometric middle of the cross-sectional interception orthogonal to the line orientation is often considered as the basis for more modern centreline extraction techniques [49]. Figure 4.1 demonstrates the intensity variations present in patch taken from a high resolution fluorescein angiogram. A simple intensity-threshold is prone to errors for vessel segmentation and centreline tracking.

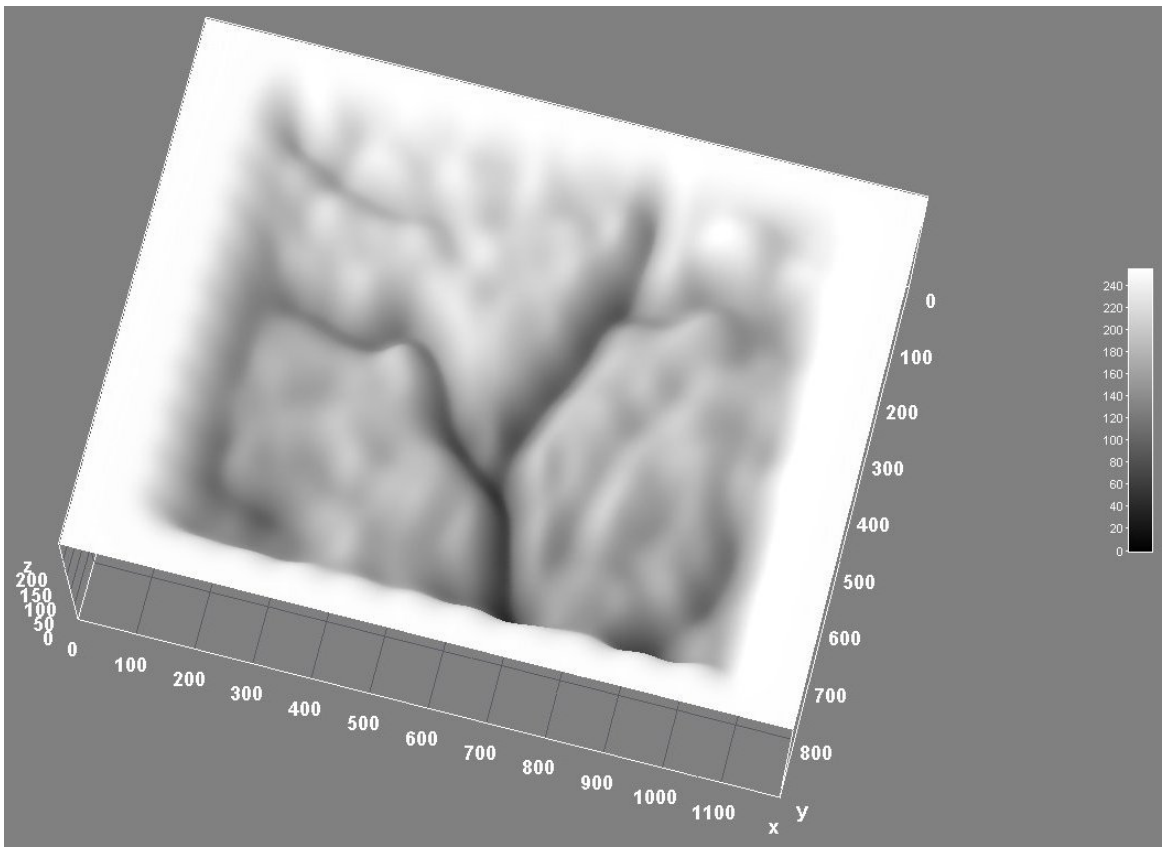


Figure 4.1: Three-dimensional intensity profile (in 256-level gray scale) of a branching retinal microcapillary near macula region, obtained by fluorescein imaging. Here troughs refer to vessel pixels, while peaks refer to background pixels.

4.3 Literature Review on Retinal Microvascular Centreline Segmentation

Generic retinal vessel tracking algorithms are often **locally additive** and **link-after-extract**. This means they often start by extracting "seed" points on the vascular centrelines either automatically or manually. Given the location and direction of a vessel, the algorithms then take a small step in the direction of the vessel from the initial seed point, and search for the centreline points within the local neighbourhood. If one such point is found, an new estimate of vessel direction in the new neighbourhood of the new "seed" point can be made, and a new step can be made in the new direction until the vessel endpoint is reached.

As existing methodologies for (micro-)vessel centreline detection and tracking often rely on approaches such as matching filter, multiscale or morphological processing (detailed

categorization was introduced in Section 2.3 on retinal vascular segmentation) to supply "seed" points, here we review these methodologies on the basis of the approaches they combine with.

Combined with Matching filter: Difference of Offset Gaussian filter (DoOG) kernel, derived by subtracting the binomial kernel from itself at a distance along the cardinal axes and is an approximation of the Gaussian derivative, was applied to extract the vessel centerlines [119]. Tramline filter was also suggested for vessel centerlines labelling [120].

Combined with Multiscale: "Minimum cost path" method, using the vector-valued multiscale representation of features, was used to extract central axis [121]. A likelihood ratio test that combines multiscale filter responses, confidence measures and vessel boundary measures was also suggested for vessel centerlines extraction [122]. Ridge, defined as an extremum in the direction of the largest surface curvature, was extracted using Hessian matrix, which was then grouped using region-growing algorithm to form affine convex sets of approximate vessel centerline pixels [2]. Alternatively, tracking strategy in combination with Gaussian and Kalman filters was also developed for retinal centreline extraction [123].

Combined with Morphological processing: Employing a combination of multiscale morphological enhancement (thresholding and morphological thinning), the fuzzy filter and the watershed transformation, the vascular tree in angiograms could be extracted [124]. Multiscale morphological reconstruction was also used to segment the vessel centrelines [119].

However, the nature of these **local-additive** and **link-after-extract** strategies means a separation of the vascular extraction and linking logic. Hence, if there is no seed points present, vessels or vessel segments are often hard to detect, and missed bifurcation points often result in undetected sub-level microvascular trees.

4.4 Suggested Method

4.4.1 Proposed "Link while Extract" Piece-wise Centreline Segmentation

Here, we propose a "**link while extract**" piece-wise segmentation approach that simultaneously extracts and tracks (sub-)pixels positioning at the centreline of (micro-)vasculature. We combine our "phase-invariant orientation field" containing local neighbourhood information (detailed definition and explanation see Section 2.4) with an additional "local phase measure" which estimates local phase for each individual pixel. If the piece-wise orientation, represented by its phase, of the (sub-)pixel aligns with the local neighbourhood dominant orientation, then this is a good estimate that the (sub-)pixel contributes to the construction of the local dominant orientation field, hence more likely to locate on the vascular centreline. The computation of this scheme is parallel as the nature of the strategy treats the two orientation measures as two variables that are temporally independent and spatially correlational.

Construction of (Phase-invariant) Orientation Field of the Local Spatial Neighbourhood

We define our *phase-invariant orientation field* as a condensed representation of the *collective* structural information in the local neighbourhood patch, the area of the neighbourhood is dependent on the "scale" of the base filers. The performance of the wavelet filter banks has been demonstrated to be stable in the presence of illumination change [66], hence highly practical in medical image processing.

To take into account the sequential nature of the globally registered angiogram stack(s), we make some minor adjustment here to Equation 2.1 as follows, where the complex field O_n (where n refers to the n^{th} image in the sequence) is constructed using a vectorial summation of the complex outputs of assigned directional base filters:

$$\mathrm{O}_n^{(l)}(x_n, y_n) = \frac{\sum_{k=0}^{K/2-1} |g_k^{(l)}(x_n, y_n)| \overrightarrow{u_k}}{p + (\sum_{k=0}^{K/2-1} |g_k^{(l)}(x_n, y_n)|^2)^{\frac{1}{2}}} \tag{4.1}$$

for $g_k^{(l)}(x, y)$ where $k = 0, 1, 2, ..., \frac{K}{2}$-1 denotes the output of the k^{th} oriented bandpass complex analysis filter from image f_n at level l.

Again, we use the directional vector set (where $K = 8$): $\overrightarrow{u_k} = [1, 0], [0, 1], [\text{-}1, 0], [0, \text{-}1]$. And p is a conditioning constant set at 1.25% of the maximum value of the image [66].

Construction of Local Phase Estimate of Each Pixel in the Neighbourhood from Steerable Wavelet Filter Output

The "local phase estimate" for each pixel is obtained from the (angular) **steering** of the complex wavelet filter outputs. Direction steering arises immediately from the use of the $cos^2(\theta)$ angular tapering in the frequency domain, whereas exact angular steerability over Fourier half space is achieved by using 4 rotated directional filters uniformly spanning the space $(0, \pi)$. Intermediate orientations can then be synthesized by combining the filter outputs with appropriately chosen weights [65]. In order to steer the filters to arbitrary angles, some means of functionally parameterising these weighting vectors is required. The learned steering coefficients from radial response exemplars in the spatial domain is parameterized through regressive (least-squares) tting of polynomial functions to each of the weight vector components, with steering angle as the regressor. This approach has the advantage of reducing the weighting vectors to a relatively small number of coefficients, but relies on the weight vector tracing out a reasonably continuous manifold in the weight space as a function of steered angle. In the context of regions full of tortuous fine vessels, the performance of this approach would benefit from using more directional filters to increase the angular resolution required to differentiate twists and turns very close to each other.

This estimate, $\Psi_n^{(l)}$, is defined as the angular component of a weighted vectorial combination of the complex wavelet filter bank where n refers to the n^{th} image in the sequence. The polynomial steering functions $s_p(\phi, k)$ and $s_q(\phi, k)$ are optimized weighting first described in

[66] where in-depth discussion on the method can be found.

The phase estimate on image f_n at level l can thus be defined as:

$$\Psi_n^{(l)} = \angle \left(\sum_{k=0}^{K/2-1} s_p(\phi, k) g_k^{(l)}(x_n, y_n) + \sum_{k=0}^{K/2-1} s_q(\phi, k) (g_k^{(l)}(x_n, y_n))^* \right) \tag{4.2}$$

The phase estimate can be viewed as an angular "feature" at the pixel level, and has been demonstrated to work well in constructing a likelihood image for edges of fine structures [66], such as the boundary ("edge likelihood") of rice grains (see Figure 4.2).

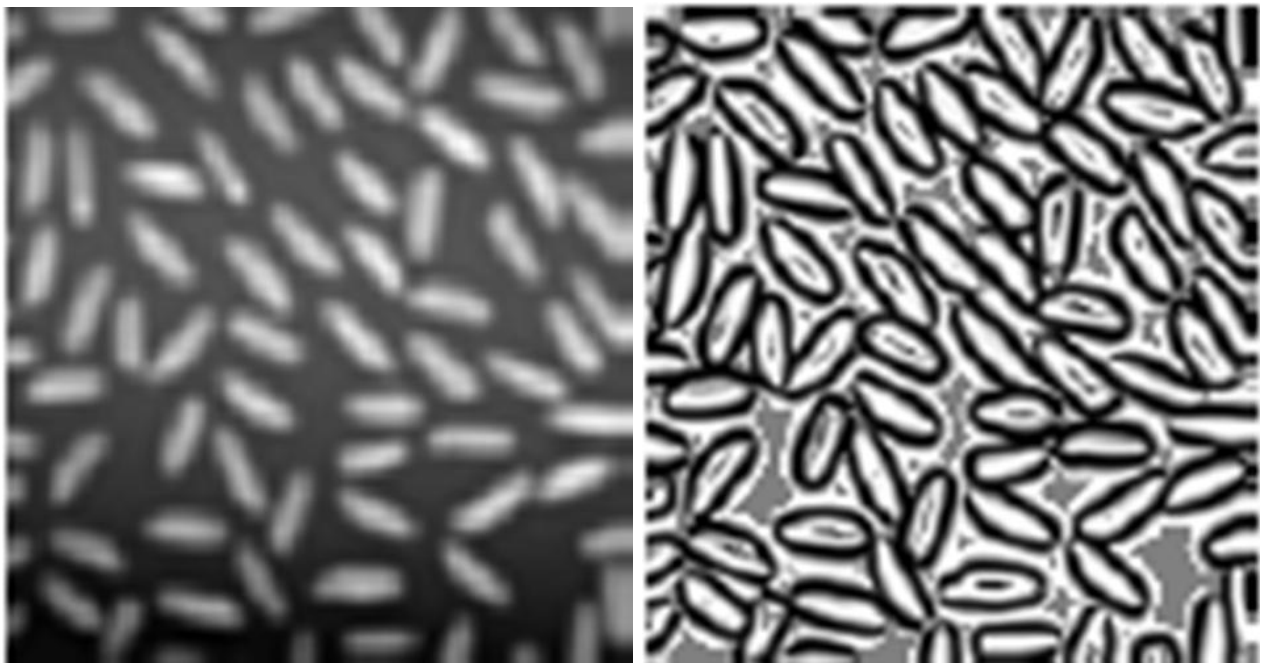


Figure 4.2: Left: Input image of rice grains at courser scale ($l = 2$); Right: Measure of edge likelihood computed using $cos(\Psi^{(l=2)})$.

Non-training Approach to Locate and Link Vessel Centreline (sub-)Pixels

In a neighbourhood $\Theta(p, b)$ from image f_n, take pixel $p = (p_x, p_y)$ located at the centre of the square $b \times b$ patch. From the prior two subsections, we now know the local vascular (neighbourhood) orientation at scale l as $\mathrm{O}_\Theta^{(l)}(x \in \Theta, y \in \Theta)$, and the phase estimate at pixel p is $\Psi_n^{(l)}(p_x, p_y)$. Since it can be assumed that the local orientation field yields a fairly accurate estimate for the local vascular direction, the phase difference between the two

indicates the (mis-)alignment of the angular components of the two measure. If the difference lies within a small tolerance, then we can assume that pixel p is more likely to locate on the vessel centreline.

To improve the efficiency the scheme, say if the local vascular (neighbourhood) orientation $\mathrm{O}_{\Theta}^{(l)}(x \in \Theta, y \in \Theta)$ at scale l is in the interval $[-\frac{\pi}{8}, \frac{\pi}{8}]$, then it is computationally sensible that only the points $(p_x + 1, p_y + 1)$, $(p_x + 1, p_y)$, $(p_x + 1, p_y - 1)$ are considered as "locally aligned" pixels, and if its local phase estimate $\Psi_n^{(l)}(p_x, p_y)$ matches the phase component of its neighbourhood orientation $\mathrm{O}_{\Theta}^{(l)}$, then pixel p can considered as a "probable" vessel centreline pixel.

If p is a vessel centreline pixel, then the angle difference between p and its "locally aligned" pixels determines whether these neighbouring pixels are also likely to be located along the vessel centreline. The combination of both allows a soft classification of the centreline (sub)pixel locations. The Euclidean distance between two (sub)pixel centreline neighbours is also calculated as an indicator for the continuity of the vessel.

Meanwhile, for centreline pixels, the angular difference between its local phase estimate $\Psi_n^{(l)}(p_x, p_y)$ and its neighbourhood orientation $\mathrm{O}_{\Theta}^{(l)}$ can be used to "correct" and improve the resolution of its actual pixel location, where:

$$
\begin{aligned}
p_x &= p_x + |sign(\angle(\mathrm{O}_{\Theta}^{(l)})) \cdot \pi - \angle(\mathrm{O}_{\Theta}^{(l)})| + |sign(\angle(\Psi_n^{(l)}(p_x, p_y))) \cdot \pi - \angle(\Psi_n^{(l)}(p_x, p_y))| \\
p_y &= p_y + \frac{|sign(\angle(\mathrm{O}_{\Theta}^{(l)})) \cdot \pi - \angle(\mathrm{O}_{\Theta}^{(l)})|}{|sign(\angle(\mathrm{O}_{\Theta}^{(l)})) \cdot \pi - \angle(\mathrm{O}_{\Theta}^{(l)})| + |sign(\angle(\Psi_n^{(l)}(p_x, p_y))) \cdot \pi - \angle(\Psi_n^{(l)}(p_x, p_y))|}
\end{aligned}
$$

Here we would like to note we achieve "sub-pixel" (centreline) pixel location by "correction" rather than "interpolation" using its neighbourhood information. Interpolation is a technique that applies a shift to the signal such that the subpixel value is moved to an integer location. To interpolate a subpixel value at a position ϵ, one needs to find the function $f'(x - \epsilon)$ from extracted discrete samples and obtain the interpolated value at $f'(0)$. In the Fourier domain, this is accomplished by applying the shift theorem as:

$$\hat{f}' = exp(-i\epsilon\pi\hat{k})\hat{f} \tag{4.3}$$

Thus the applied interpolation filter has a unit magnitude but introduces linear phase shift $\epsilon\pi\hat{k}$ to bring the subpixel value to the integer position [125].

4.5 Discussion and Conclusion

Each sub-steps in our "**link while extract**" piece-wise segmentation approach can be computed simultaneously and combined in one final step, and the direction of vessel centreline pixel extraction can be bi-directional (the nature of orientation description does not not differentiate directional vectors with 180 degrees angular difference in 2D). And more importantly, the entire algorithm is completely automated requiring no manual "seeding" of the start and end (vascular centreline) pixels.

One application of this method is to automatically evaluate (sequential) registration accuracy without prior knowledge of ground-truth. In-depth description and detailed discussion are presented in Section 6.3. As the method itself relies on few thresholds, does not require region growing or manually-segmented ground-truth, we can conclude that our suggested error metric ("Centreline Error") is promising in evaluating the registration accuracy in fine-resolution microvasculature-rich regions, ie. the macula., often omitted by standard grid-based evaluation approaches.

In addition, our centreline extraction method, coupled with our novel spatio-temporal registration framework (proposed in Chapter 3), allows comparison and non-invasive monitoring of fine-resolution microvasculature from fluorescein angiography. It provides the potential for detecting temporal changes in the circulation (possibly caused by (micro)embolism) in real-time. This allows early preventative measures to be taken to reduce aggravated blood-clotting, thus improving the post-operative recovery of the patients.

Figure 4.3 and 4.4 illustrate the extracted microvascular centrelines in sub-regions extracted

from two different retinal fluorescein angiogram using our proposed method. Figure 4.5 and 4.6 illustrates the comparison results of the retinal vasculature of one patient before and after major surgery, likely occluded (disappearance of microvasculature) by microemboli.

In this chapter, we first present a literature review on existing methods for (micro)vascular centreline segmentation (extraction and tracking) in retinal fluorescein angiography. We then present our proposed "link while extract" piece-wise segmentation method up to subpixel accuracy relying on phase information extracted after steering complex wavelets according to the phase-invariant orientation field. Our method is computational efficient due to the temporally independent and spatially correlational nature of the measures. Finally, we conclude the chapter by presenting two potential applications of our method: the automated evaluation of (microvasculature) registration accuracy by quantifying the (sub-)pixel centreline offset and the automated occlusion detection by spatio-temporal assessment and comparison of patient's retinal vasculature tree. We would like to note the rationale behind keeping the Chapter 4 and Chapter 2 separate is largely because Chapter 4 focuses on fine vessel (centreline) segmentation in both temporal and spatial domain whereas Chapter 2 solves segmentation problem only in spatial domain but on the entire vessel bed within the field of view (FOV). The results based on techniques described in this chapter are largely dependent on the success of the high-resolution sequential registration (see Chapter 3).

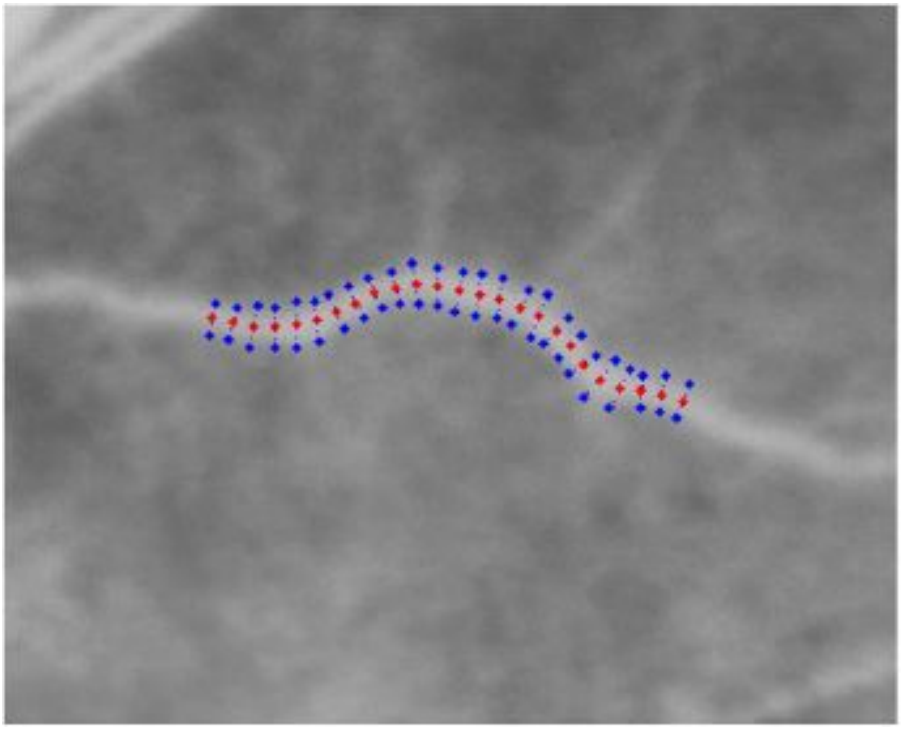


Figure 4.3: Tracked retinal vascular centreline (in red) with the boundary pixels in blue, from a sub-section of a retinal angiogram.

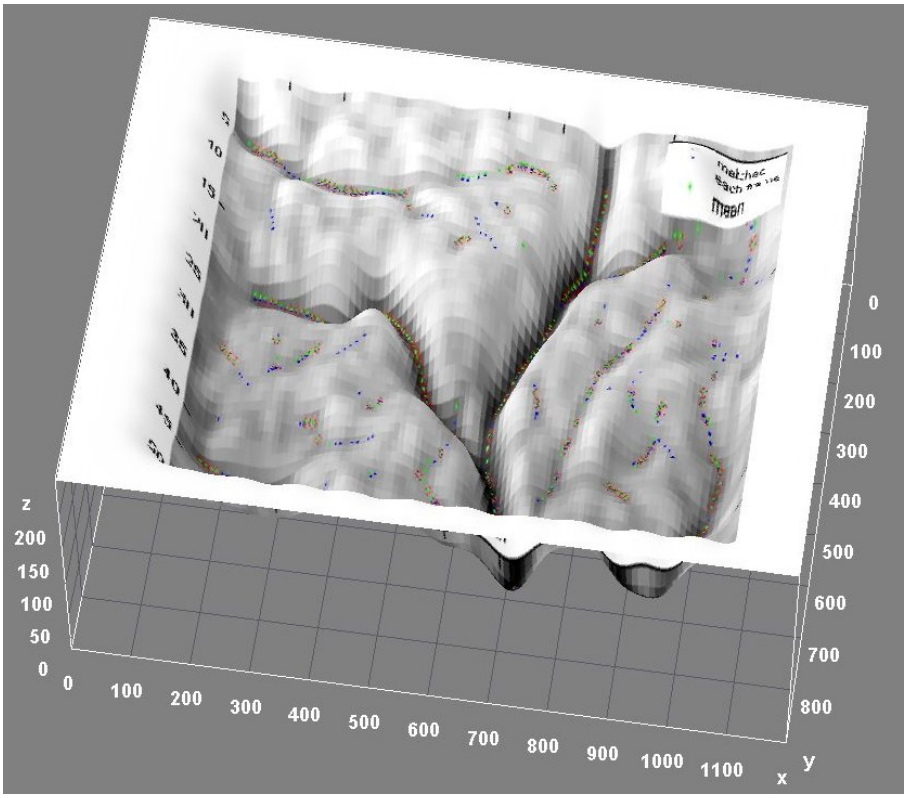


Figure 4.4: Three-dimensional intensity profile (in 256-level gray scale) of a branching retinal microcapillary near macula region with extracted centreline (in blue and green) superimposed on the profile. Here troughs refer to vessel pixels.

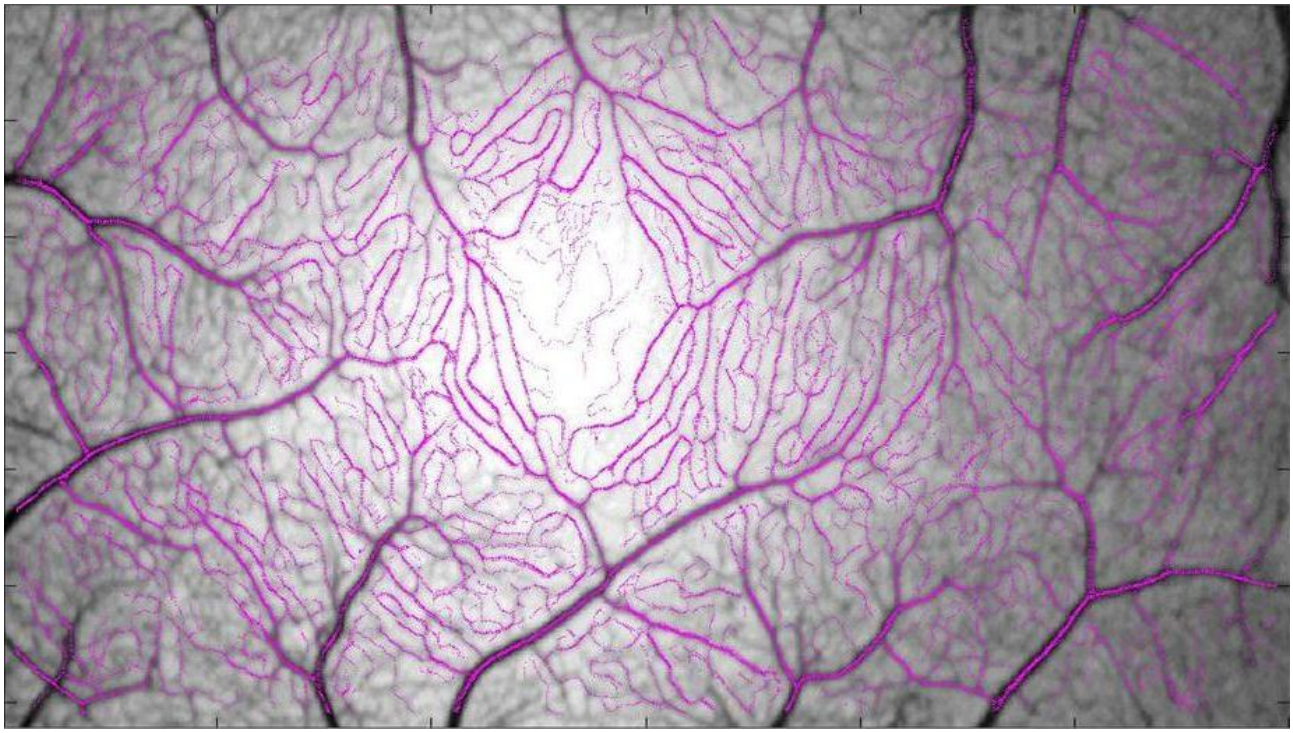


Figure 4.5: Macular vasculature (centreline in magenta) imposed on pre-operative fluorescein angiogram.

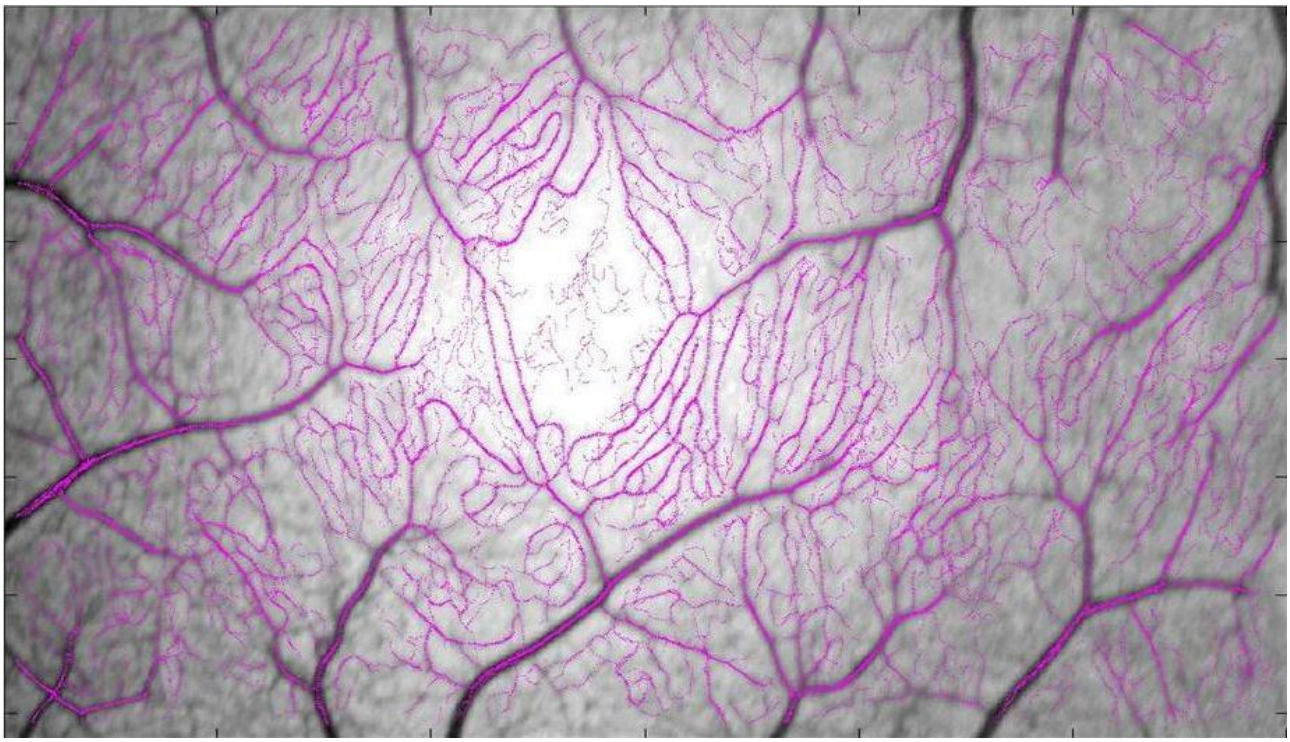


Figure 4.6: Macular vasculature (centreline in magenta) imposed on post-operative fluorescein angiogram; with visible disappearance of microvascualtures (in red) compared to the pre-op centrelines above, possibly occluded by microemboli.

Chapter 5

3D Microscopic Stack of Cerebellar Cortex: Segmentation

5.1 Introduction

Detailed understanding of structure, function, and development of the nervous system at the molecular, cellular, and systematic levels is crucial for neuroscience, as these processes are often considered the bio-physiological building blocks of higher-order cognitive functions. Recent progress in neuroscience is increasingly relying on imaging, giving rise to massive amounts of heterogeneous and complex data collected at multiple scales of observation. This, in turn, leads to increasing demand for advanced bioimage informatics and neuroinformatics solutions to integrate and analyzing these data.

In particular, the recent revival of computational methods and tools to study neuronal anatomy is due to the fact that the morphological properties of the axonal and dendritic arborizations is often highly significant in network connectivity, signal transmission and neural circuitry dynamic. Quantitative analysis on neuronal morphology could potentially lead to the discovery of factors influencing neuronal interaction and the functional roles afflicted by certain structural alterations [126].

The difficulty for image analysis in neural application is that the large and often sparse data sets are acquired using varying image techniques and equipments aimed to capture specific structures of statistical significance fitted to individual experimental/clinical study. This breath of scientific investigations give rise to non-standardization in image archive, thus make it difficult to analyze and compare across the industry. There have been constant drive towards industry-wide agreement on a common representation of the neuronal topology and geometry.

What we aim to achieve here is not a standardized point coordinate system or contingent local thickness measurement. Rather, we aim to offer a more rounded solution that seeks structural connectivity between points in a 3D data stack using neighbourhood orientation information. The difference in our approach is that the system can be "trained" to reflect application-specific imaging and hence makes the approach more inherently tailored for digital segmentation with little prior knowledge.

The order of the chapter is organized as below: Section 5.2 describes the context of segmentation in neuro-imaging. Section 5.3 is the literature review of common methodologies used for microvascular segmentation of neural structures. Section 5.4 is our proposed pixel-level patch-based segmentation with MLP training. Lastly, we discuss the benefits of our approach and conclude this chapter in Section 5.5.

5.2 Context

To investigate the molecular mechanisms involved in neurite outgrowth and differentiation, accurate and reproducible segmentation and quantification of neuronal processes are a prerequisite.

The beginning of modern neuroscience is often associated with the first drawings of complex neuronal arbors since the 1873 publication of Golgis staining [126]. However, as computational approaches are necessary to quantify the intricate relationship between neuronal structure and physiological activities, digitization of neuronal morphometry became

the preferred approach since its inception in 1960s [127] [128]. It often consisted in using computers to interact with the microscope and to store point coordinates indicated manually by a human operator. Many attempts were made to reduce the amount of manual labor involved, but the level of automation remained very limited due to a lack of computer power and sophistication in computerized image analysis methods [128].

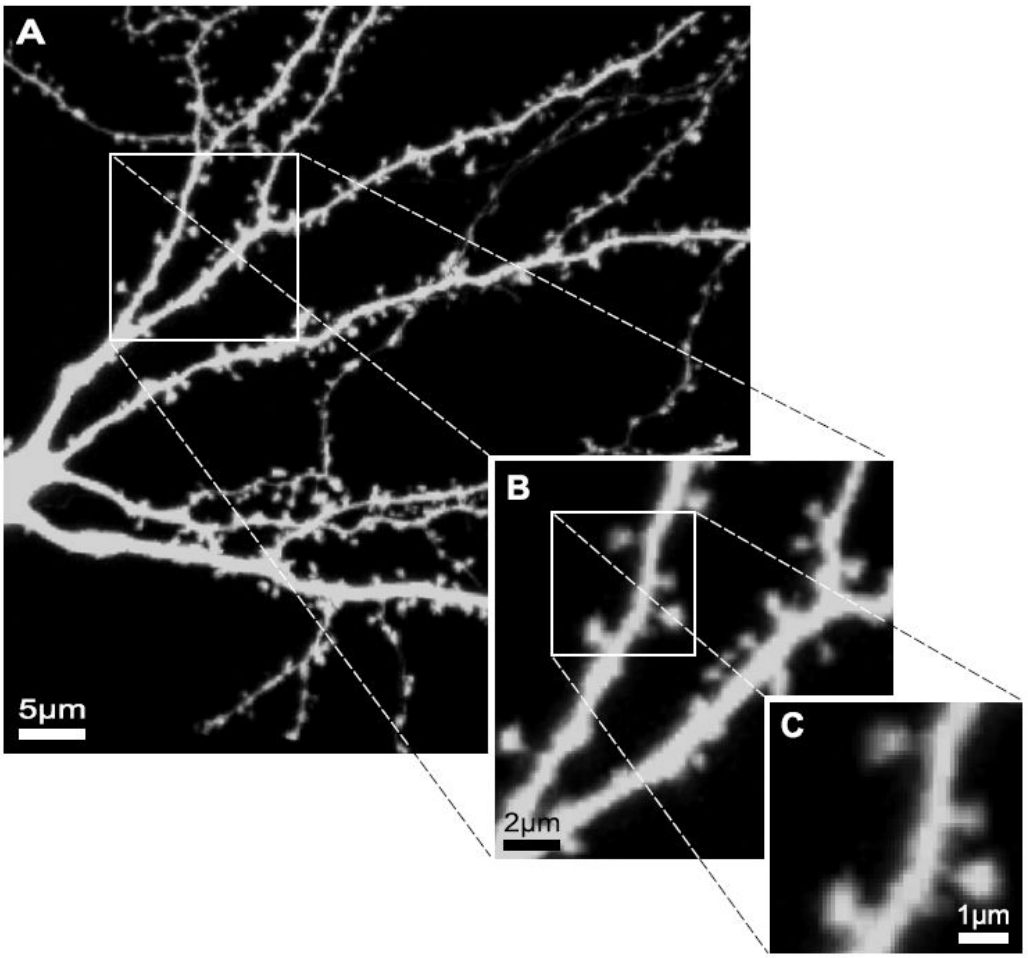


Figure 5.1: A partial pyramidal cell represented by maximum intensity projection of different portions of the original 3D fluorescence microscopy image stack. At neuron level (A), at the level of dendritic branches and bifurcations (B), at the level of individual spines (C), adapted from [11].

Over the past two decade, open source release of digital/digitized neuro-informatics resources [7] initiated the growth of a more mature, diverse, and vibrant ecosystem of users and applications. As more and more academic tools are being developed, the problem has now shifted from lack of neuro-anatomical tools back in the early days to the lack of *general applicability* of existing tools. Thus, in practice, neuronal analysis (neuron tracing and

reconstruction) are still mainly carried out manually by human experts.

Figure 5.1 illustrates the complexity of the neural segmentation problem. For a high-quality 3D microscopic data set, it is commonly observed that some branches can poorly stained and demonstrate low contrast, causing visual ambiguities. In practice, this local poor signal-to-noise ratio is especially problematic for segmentation of the spines, which due to limited resolution may easily appear as detached blobs.

5.3 Literature Review on Microvascular Segmentation in Neural Imaging

The current state-of-the-art for neural segmentation can often described as the process of assigning each volume element (voxel) a "label" indicating which part of the neural structure it belongs to. This process often involves three distinctive steps: image preprocessing, segmentation of the "soma" (cell body) and segmentation of the "neurites" (axon and dendrites). Noticeably, common approaches tend to approach the 3D problem in a 2D manner, ie. using 2D projections of the data (in x-y, x-z, and y-z planes) instead rather than the entire 3D stack to save computational cost and degrees of freedom in modelling.

5.3.1 Image Preprocessing

This step is intended to remove to some extent factors that hamper automated image segmentation, such as noise caused by statistical fluctuations or other *irrelevant* structures, resolution-constraint, inhomogeneous contrast from non-uniform spatial and temporal distribution of the dye, and background gradients from non-uniform illumination. Image processing operations such as (nonlinear) smoothing [129] [8], deconvolution [130] [131], shading correction [129] [8], and morphological filtering [132] are often applied to minimization artifacts described above.

5.3.2 Soma Segmentation

The soma is a crucial component of a neuron where dendritic signals are joined and passed on. As it constitutes the root of the axonal and dendritic trees, its segmentation are often served as a starting point for segmenting the latter [133]. In fluorescence microscopy, additional experimental techniques can be made to distinguish soma by simply stain them separately, ie. DAPI (4',6-diamidino-2-phenylindole) nucleic acid stain, so that the nuclei light up in a different complementary image channel [134]. For the cases when only one stain is used, the somas may have a higher dye concentration than the neurites, intensity thresholding [135] can be applied. More commonly, as it is difficult to distinguish soma solely based on intensity, (binary or grayscale) morphological filtering are often applied to tackle the problems [136] [137].

5.3.3 Neurite Segmentation

Neurite segmentation can be categorized into *global* or *local* approaches.

Global approaches often start with **binarization** to yield a "rough" image structures using adaptive thresholding [132] [133] [138] [139]. However, intensity thresholding can be erroneous and it is highly reliant on homogeneity of the staining. This is then followed by **skeletonization** to extract the centerlines of the segmented areas for a more compact description of the neurites [132] [133] [138] [139]. Alternatively, Hessian [11] [140] or Jacobian [141] based analysis of critical points, or nonmaximum suppression [142] [137] can also be used to extract neurite centerlines directly from the grayscale images. The neurite centreline from the skeletonization often contains spurious gaps, loops, crossings or branches. Various filling and pruning strategies have been developed to try to correct the errors [143] [11] [140] [132] [133] [138] [139] [142] [137] [141] .

Local approach traces neurites by using the information on the current point, iteratively predicts the next point on the neurite and simultaneously updates this estimate by fitting a shape or a profile model [136]. Region-growing methods can be used to segment the neurites

running largely in the axial direction whose centroid positions can serve as seed points to initiate segmentation [144] [145]. Alternatively, energy minimization approaches are also proposed to constrain the search to given start and end points by defining a "cost" or energy function that assigns a penalty to connecting any two neighboring points (computed from local image features at these points), and minimizing the cumulative cost from start to end point [146] [147].

5.4 Proposed Pixel-level Patch-based Segmentation with MLP Training

5.4.1 3D to 5D Orientation Mapping to Preserve Angular Information and Line Curvature

Simple neighborhoods in images are commonly defined as areas that are "shift invariant in at least one direction and not shift invariant in at least one other direction" [148]. Such areas are the basis in the description of directional information of the local structures. With the first order intensity variation being the gradient, a collection of local gradients is needed to compute a **dominant orientation**. The accompanying intensity variations and that of orthogonal directions can be used to describe lines, surfaces and edges as well as texture. A characterization of simple neighborhoods is by the dominant orientation [148].

A *line* can be defined a symmetric separator of two regions that no relation between the regions can be inferred, whereas an *edge* is an asymmetric separator that an ordering of the two regions is often implied. Thus, lines can only be unambiguous when associated with an angle spanning over an interval of 180° whereas edges can be unambiguous when associated with an angle spanning over an interval of 360° [13] [14].

In other words, representing orientation of a line using vector angles can be troublesome as it is discontinuous. Vectors pointing at opposite directions can have the the same orientation

5.2. Coherent circular vectors, on the other hand, does not necessarily result in coherent orientation 5.3.

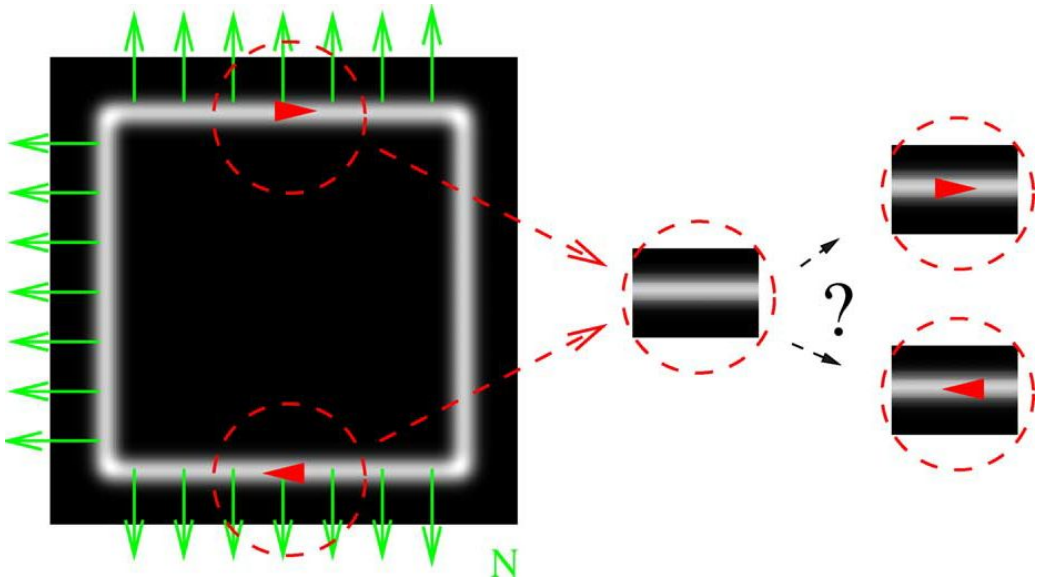


Figure 5.2: The normal vectors (in green) in the two red-dashed windows point in opposite directions even though the orientation of content in both windows is the same.

To remove this discontinuity a **mapping** is required for a continuous distance preserving representation of structural orientation. Knutsson [14] proposed three properties to stand should such mapping exists:

1. **Uniqueness**: Antipodal vectors should be mapped onto one point to remove the phase jump. This can be expressed as:

$$\overrightarrow{V}(\overrightarrow{X}) = \overrightarrow{V}(\overrightarrow{-X}) \tag{5.1}$$

   where $\overrightarrow{X}$ is a vector in the original space; and $\overrightarrow{V}$ is the mapping function.

2. **Polar Separability**: The norm of the mapped vector should be rotation invariant, thus information carried by the magnitude of the original vector does not normally depend on the angle. This can be expressed as:

$$|\overrightarrow{V}| = R = f(r) \tag{5.2}$$

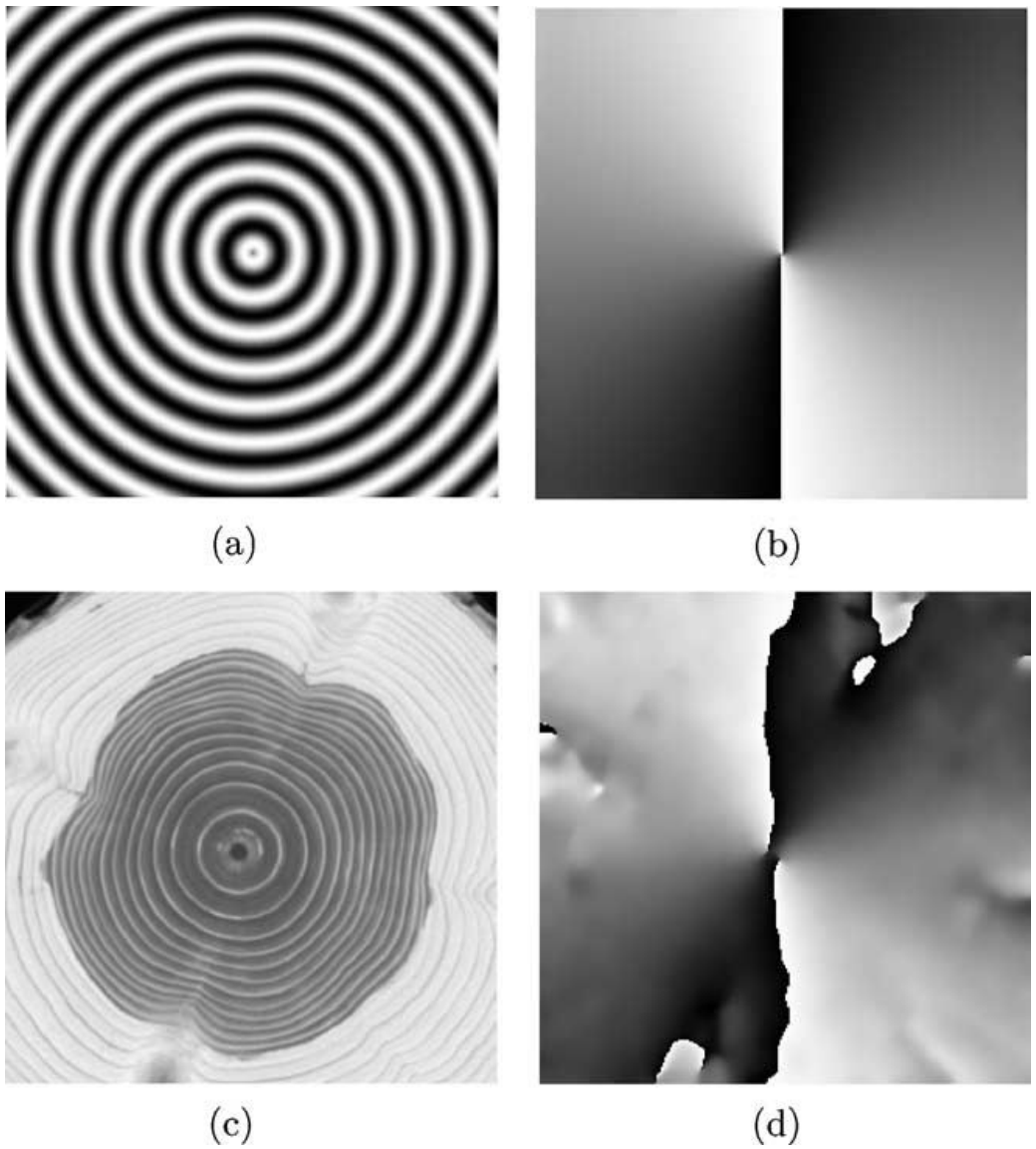


Figure 5.3: (a) Synthetic image with concentric circles; (b) the angular representation with interval $[-\pi/2, \pi/2]$ (light, dark) with a discontinuity mod π; (c) a CT cross-section of a tree trunk; (d) the angular representation of the CT slice, adapted from [12].

where $r = |\overrightarrow{X}|$ and f is an independent function relating the magnitude between $\overrightarrow{V}$ and $\overrightarrow{X}$.

3. **Uniform Stretch**: The mapping should locally preserve the angle metric of the original space, ie. it carries implicitly information about the distances in the original space that is rotation invariant and scales linearly with the angle between two hyper planes. This can be expressed as:

$$|\delta(\overrightarrow{V})| = C|\delta(\overrightarrow{X})|; r = const. \tag{5.3}$$

where $r = |\overrightarrow{X}|$ and C is the "stretch" constant.

A generalized form of mapping that fulfills above requirements can be defined as such that $\overrightarrow{V}: \quad \Re^n \rightarrow \Re^{n \times n}$:

$$\overrightarrow{V}(\overrightarrow{X}) = \frac{\overrightarrow{X}\overrightarrow{X}^T}{\|\overrightarrow{X}\|} \tag{5.4}$$

In the case of 3D, a 3D-to-5D mapping, denoted by M_3, was first proposed by Knutsson [13] that fulfills such requirements, where:

$$M_3: \quad \begin{cases} s = rsin^2\theta cos2\phi \\ t = rsin^2\theta sin2\phi \\ u = rsin2\theta cos\phi \\ v = rsin2\theta sin\phi \\ w = r\sqrt{3}(cos^2\phi - \frac{1}{3}) \end{cases} \tag{5.5}$$

where

$$\overrightarrow{V} = (s, t, u, v, w) \tag{5.6}$$

$$\overrightarrow{X} = (r, \theta, \phi) \tag{5.7}$$

The definition of $\overrightarrow{X}$ in polar coordinates is illustrated in Figure 5.4.

In a way M_3 mapping can be viewed as simultaneous realization of the 2D-to-2D mapping of all possible 3D orientation of the plane, which in turn can be visualised by Figure 5.5. Each plane in 3D plus its axis is mapped into a separate 3D subspace in a 5D space. The subspace is mostly empty as the plane is actually mapped onto the surface of a cone while the axis is mapped into a line.

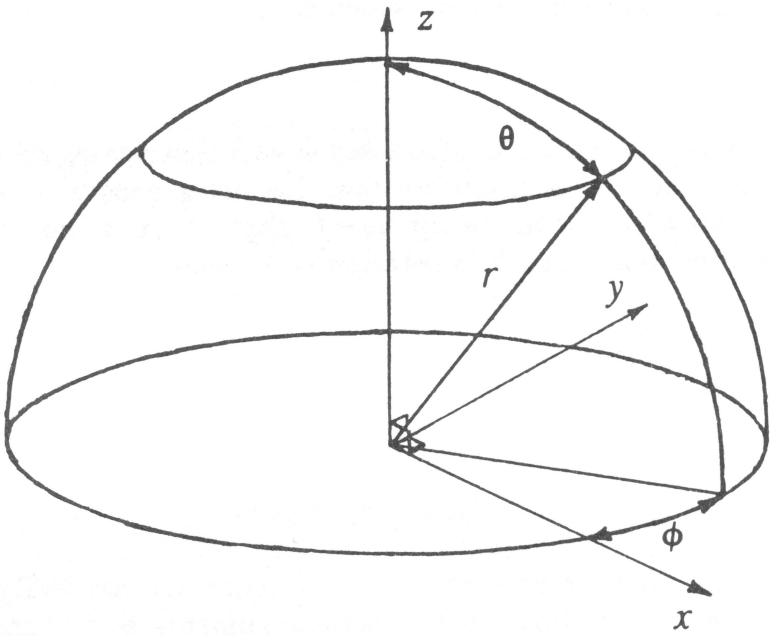


Figure 5.4: Definition of X in polar coordinates.

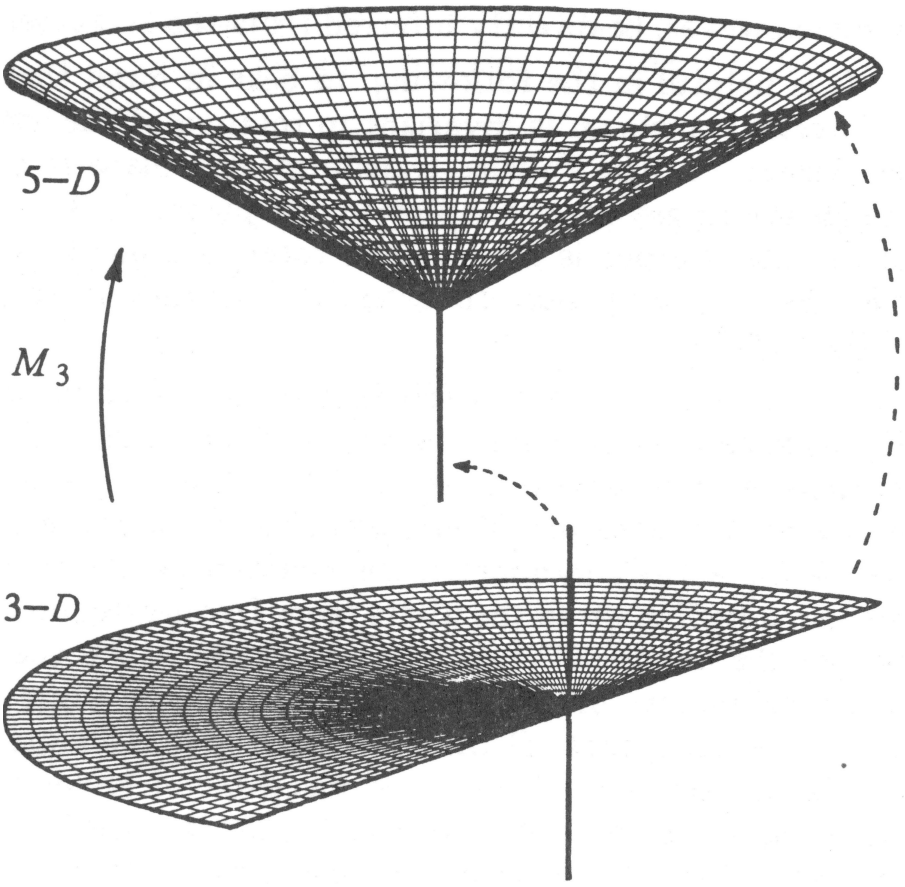


Figure 5.5: A 2D orientation half plane in 3D is mapped onto a 2D cone in 5D with opening angle of 120 degrees, adapted from [13].

5.4.2 Filter Realization of 3D to 5D Orientation Mapping

We now need to implement the mapping in the prior section in a meaningful way such that the representation implied by the mapping can realized using measurements on actual image data, where structures are represented as local grayscale correlations. It has been shown that by combining the outputs from polar separable quadrature filters, it is possible to produce a representation corresponding exactly to Equation 5.5. The exactness relies on the image data being locally one-dimensional [14], such that there exists a locally well defined orientation. The case where the one-dimensionality assumption does not hold can be found in [14] where a best approximation is used to estimate the local orientation.

We follow a three-step procedure adopted by methods first proposed by Knutsson to realize our filter.

1. The **polar-separable** filter shape

   As mentioned in the prior section, this is one of the three properties required for our orientation mapping. The frequency response of the **even** filters can be expressed as:

   $$F_e(\overrightarrow{\omega}) = g(|\overrightarrow{\omega}|)(\overrightarrow{\omega} \cdot \overrightarrow{n_k})^2 \qquad g(|\overrightarrow{\omega}|) \geq 0 \tag{5.8}$$

   where $\overrightarrow{\omega}$ is the frequency and the $\overrightarrow{n_k}$ is a unit vector determining the filter-orientation. The radial part of the above function is arbitrary but positive (usually some type of bandpass function) and the angular part varies as $cos^2(\varphi)$, where φ is the angular difference between ω and the filter-orientation $\overrightarrow{n_k}$. A complementary **odd** filter in the frequency domain,$F_o(\overrightarrow{\omega})$, is also required, which can be written as:

   $$F_o(\overrightarrow{\omega}) = F_e(\overrightarrow{\omega}) sign(\overrightarrow{\omega} \cdot \overrightarrow{n_k}) \tag{5.9}$$

   where $\overrightarrow{\omega}$ is the frequency and the $\overrightarrow{n_k}$ is a unit vector determining the filter-orientation.

2. The **quadrature** filter concept

A quadrature filter is defined in Fourier space as a filter being positive over half of the Fourier space and zero over the other half, which can be written as:

$$\begin{cases} F(\overrightarrow{\omega}) > 0 \quad if \quad (\overrightarrow{\omega} \cdot \overrightarrow{n_k}) > 0 \\ F(\overrightarrow{\omega}) = 0 \quad otherwise \end{cases} \tag{5.10}$$

where $\overrightarrow{\omega}$ is the frequency and the $\overrightarrow{n_k}$ is a unit vector determining the filter-orientation.

Thus, this forms the basis for obtaining phase invariant information as the output q_k of the corresponding quadrature filter will be a complex number, where the magnitude of q_k will be phase invariant (implying local shift invariance) and the argument $\arg(q_k)$ represents the local phase.

To retain the phase invariant quality, the quadrature filter output Q is taken to be:

$$Q = (q_o^2 + q_e^2)^{1/2} \tag{5.11}$$

where q_e and q_o are the outputs from the filter F_e and F_o defined in Equation 5.8 and 5.9.

3. The **directional** filter set

Consider the case of $2N1$ quadrature filters, having symmetry axes passing through the corners of a cube in N dimensions, giving a fully symmetric distribution of filters. Consider the contribution to the filters from frequencies on a line through the center of two opposing cube faces. Since the angle between the line and any filter axis will be the same, it is clear that all the filters will give the same output. Consequently the filter set is incapable of giving information sufficient to determine which pair of cube faces the line passes through and thus, the orientation of the signal is undecidable. It can thus be concluded that more than $2N1$ quadrature filters must be used to differentiate among them.

For $N = 3$ the number of filters must be greater than 4, but the possible number for filters to distribute in a fully symmetrical fashion are those given by half the number of

vertices (or faces) of a diametrically symmetric regular polyhedron, leaving **6** as the minimal amount of directional filters requires to encrypt 3D orientation.

The orientations of the 6 directional filters are given by vectors pointing to the vertices of a hemi-icosahedron (see Figure 5.6).

The six unit vectors can be expressed in Cartesian coordinates as:

$$\begin{cases} \overrightarrow{n_1} = c(a, 0, b) \\ \overrightarrow{n_2} = c(-a, 0, b) \\ \overrightarrow{n_3} = c(b, a, 0) \\ \overrightarrow{n_4} = c(b, -a, 0) \\ \overrightarrow{n_5} = c(0, b, a) \\ \overrightarrow{n_6} = c(0, b, -a) \end{cases} \tag{5.12}$$

where

$$\begin{cases} a = 2 \\ b = 1 + \sqrt{5}c = (10 + 2\sqrt{5})^{-1/2} \end{cases} \tag{5.13}$$

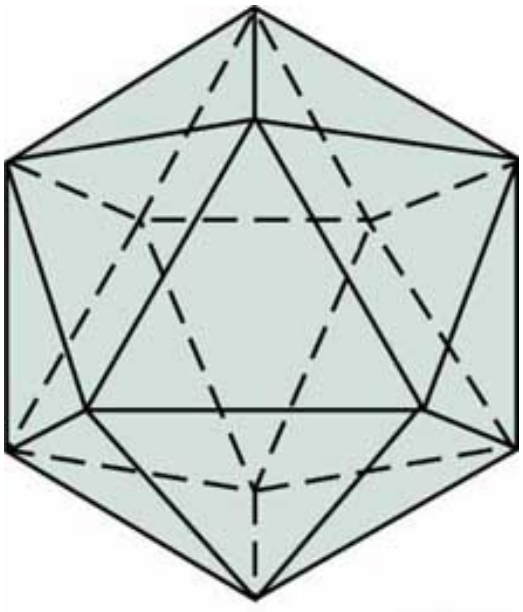


Figure 5.6: An icosahedron (one of the five Platonic polyhedra), adapted from [13].

Above 6 filers,$\overrightarrow{n_k}$ where $k = 1, 2, .., 6$, are then mapped using Equation 5.5 to find the corresponding 6 vectors $\overrightarrow{m_k}$ in the 5D space.

The polar form for the 3D sphere (the original space) can be written as:

$$\begin{cases} x = rsin\theta cos\phi \\ y = rsin\theta sin\phi \\ z = rcos\theta \end{cases} \tag{5.14}$$

This leads to Equation 5.5 re-written as:

$$\begin{cases} s = r^{-1}(x^2 - y^2) \\ t = r^{-1}2xy \\ u = r^{-1}2xz \\ v = r^{-1}2yz \\ w = r^{-1}\frac{1}{\sqrt{3}}(2z^2 - x^2 - y^2) \end{cases} \tag{5.15}$$

This, the resulting 5D vectors, $\overrightarrow{m_k}$ can be written as:

$$\begin{cases} \overrightarrow{m_1} = c^2(a^2, 0, 2ab, 0, \frac{1}{\sqrt{3}}(2b^2 - a^2)) \\ \overrightarrow{m_2} = c^2(a^2, 0, -2ab, 0, \frac{1}{\sqrt{3}}(2b^2 - a^2)) \\ \overrightarrow{m_3} = c^2(b^2 - a^2, 2ab, 0, 0, -\frac{1}{\sqrt{3}}(a^2 + b^2)) \\ \overrightarrow{m_4} = c^2(b^2 - a^2, -2ab, 0, 0, -\frac{1}{\sqrt{3}}(a^2 + b^2)) \\ \overrightarrow{m_5} = c^2(-b^2, 0, 0, 2ab, \frac{1}{\sqrt{3}}(2a^2 - b^2)) \\ \overrightarrow{m_6} = c^2(-b^2, 0, 0, -2ab, \frac{1}{\sqrt{3}}(2a^2 - b^2)) \end{cases} \tag{5.16}$$

Figure 5.7– Figure 5.12 below uses 6 uniform smooth lines in the $65 \times 65 \times 65$ volume to represent the 2D plane projection of the 6 filter directions, both as the original 3D vector $\overrightarrow{n_k}$ and as the mapped 5D vector $\overrightarrow{m_k}$ where each component of the complex 5D

vector are all represented by its 2D plane projection. The lines and volumes are represented in the format of Matlab 3D matrices.

4. The **combination** scheme

   The final result can be obtained by projecting the output of filters onto fixed unit vectors and summing the projections. The linear summation of the quadrature filter set output magnitudes can be written as:

$$\overrightarrow{\Psi} = \sum_k Q_k \overrightarrow{m_k} \tag{5.17}$$

   where

$$\overrightarrow{m_k} = \overrightarrow{V}(\overrightarrow{n_k}) = M_3(\overrightarrow{n_k}) \tag{5.18}$$

   for Q_k is the output from the filter oriented by $\overrightarrow{n_k}$.

   A way to visualise our filter set is to convolve the filter sets with uniform smooth test spheres and visualise the filter responses. Figure 5.13 illustrates the combined 6 directional filter responses on 3 test spheres.

5.4.3 Supervised Machine Learning on Constructed High-dimensional Feature Patch

As proposed in prior Section 2.4, a multilayer perceptron (MLP) can then be trained over a small sample of labeled pixels to effect the segmentation. The MLP system we use here is a multi-layer feed-forward network with backpropagation and Levenberg-Marquardt training algorithm [73]. The network uses a scaled conjugate gradient descent training function, a mean squared error performance function, and hyperbolic tangent sigmoid transfer functions for both hidden and output layers.

The high-dimensionality of the microscopic data stack compounded with our complex feature space results in huge input feature vector if we simply extend our existing method in 2D segmentation to 3D: For each voxel, its $3 \times 3 \times 3$ neighbourhood intensity combines with the

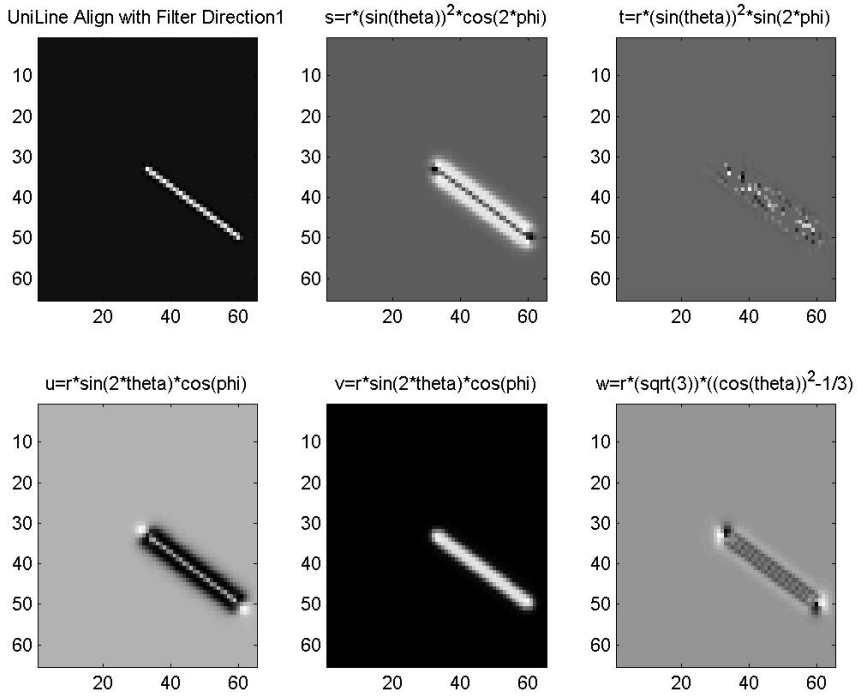


Figure 5.7: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_1}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_1}$. Note, the slice is taken at the Z-X plane at the centre of the volume (Y=33).

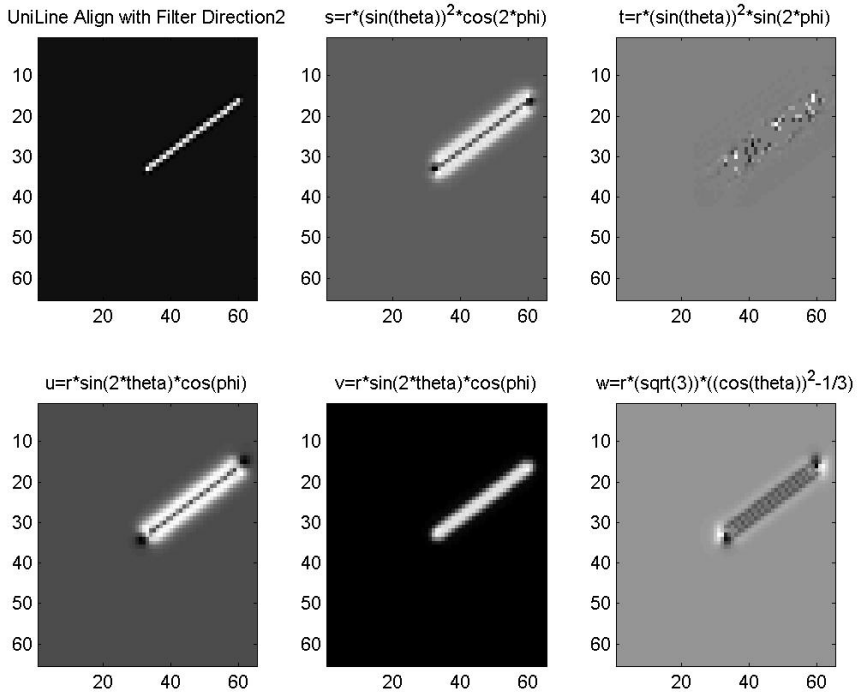


Figure 5.8: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_2}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_2}$. Note, the slice is taken at the Z-X plane at the centre of the volume (Y=33).

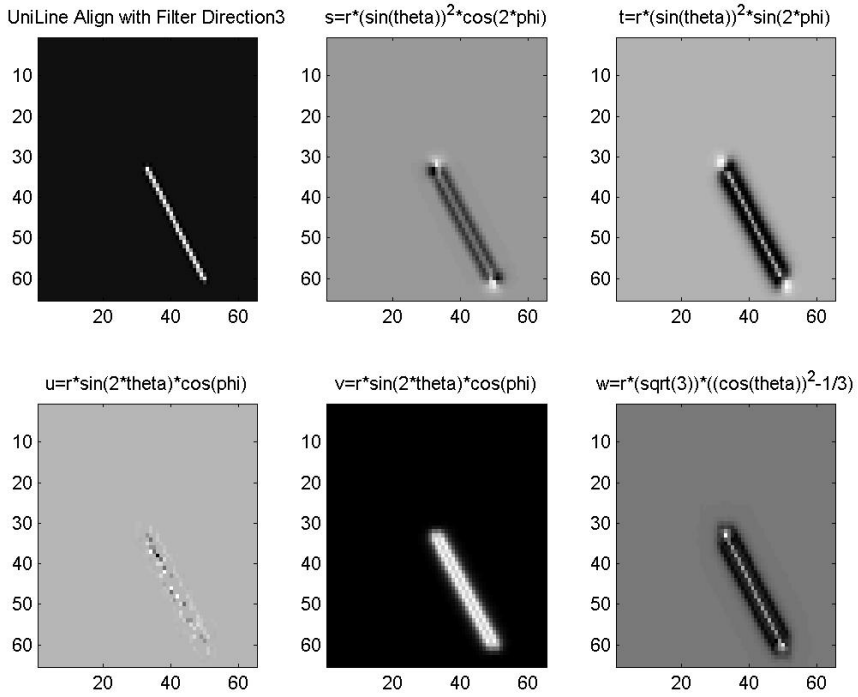


Figure 5.9: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_3}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_3}$. Note, the slice is taken at the Y-X plane at the centre of the volume (Z=33).

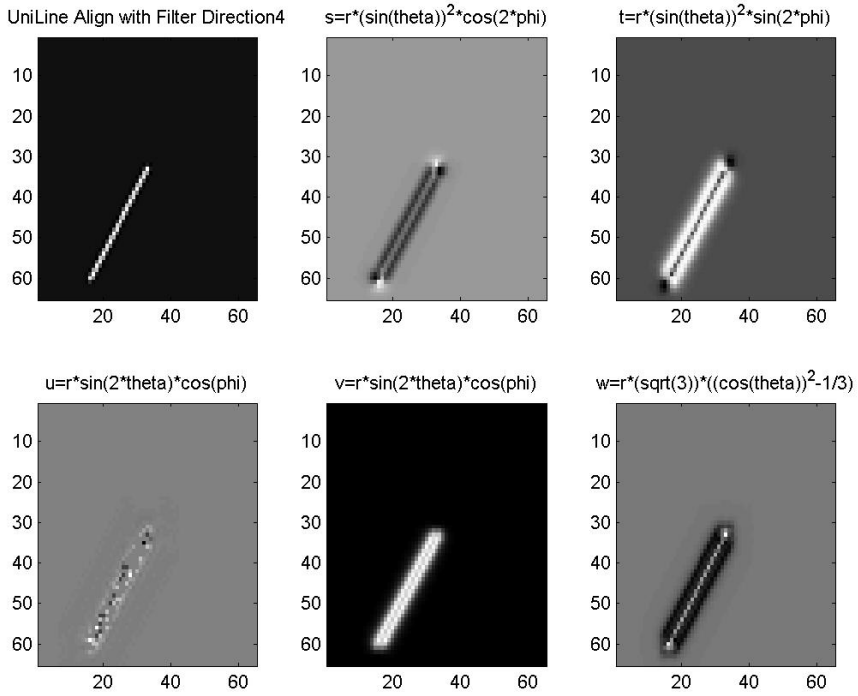


Figure 5.10: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_4}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_4}$. Note, the slice is taken at the Y-X plane at the centre of the volume (Z=33).

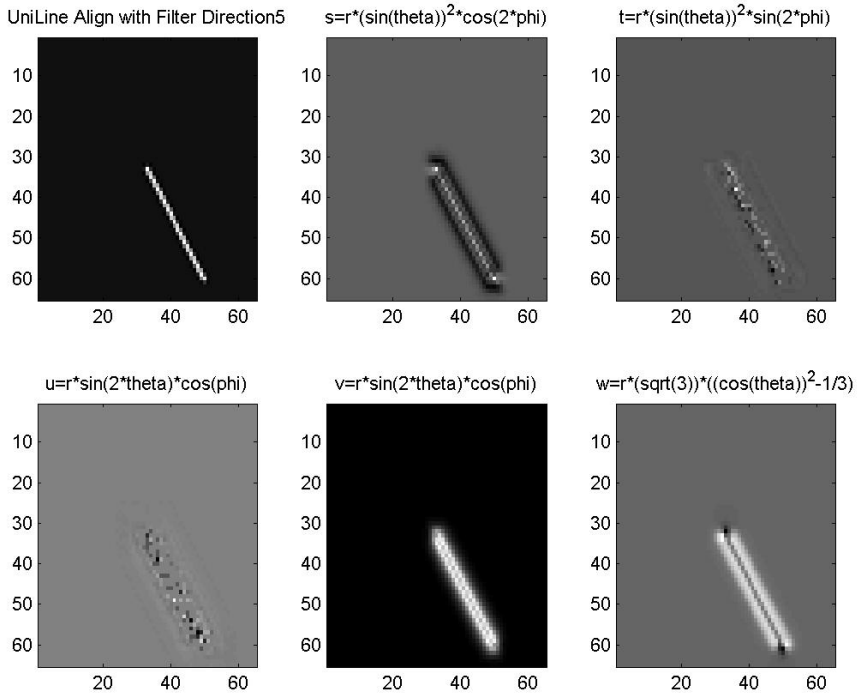


Figure 5.11: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_5}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_5}$. Note, the slice is taken at the Z-Y plane at the centre of the volume (X=33).

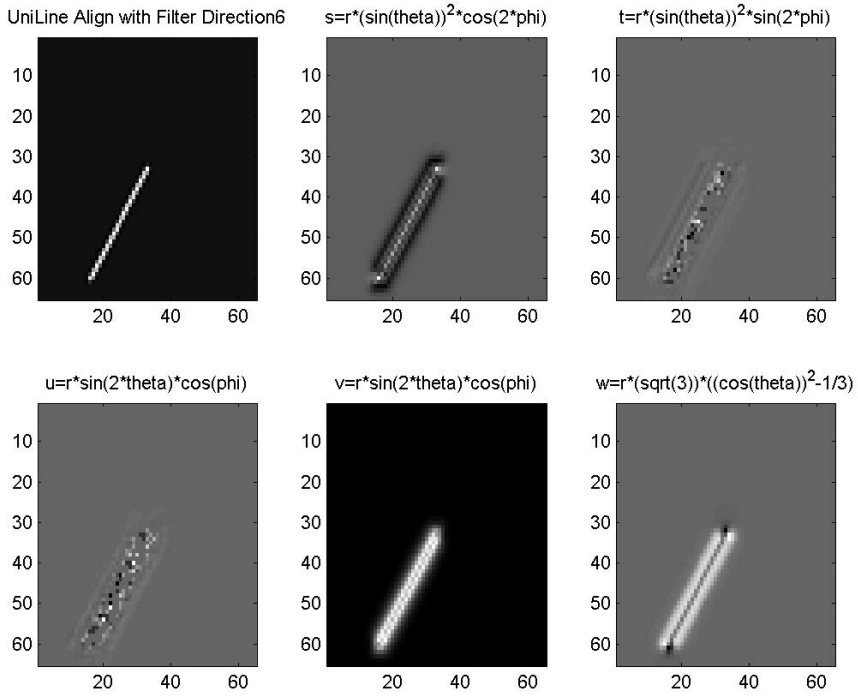


Figure 5.12: A uniform smooth line in a $65 \times 65 \times 65$ volume aligning with Filter $\overrightarrow{n_6}$, with the 5 components of its mapped 5D vector $\overrightarrow{m_6}$. Note, the slice is taken at the Z-Y plane at the centre of the volume (X=33).

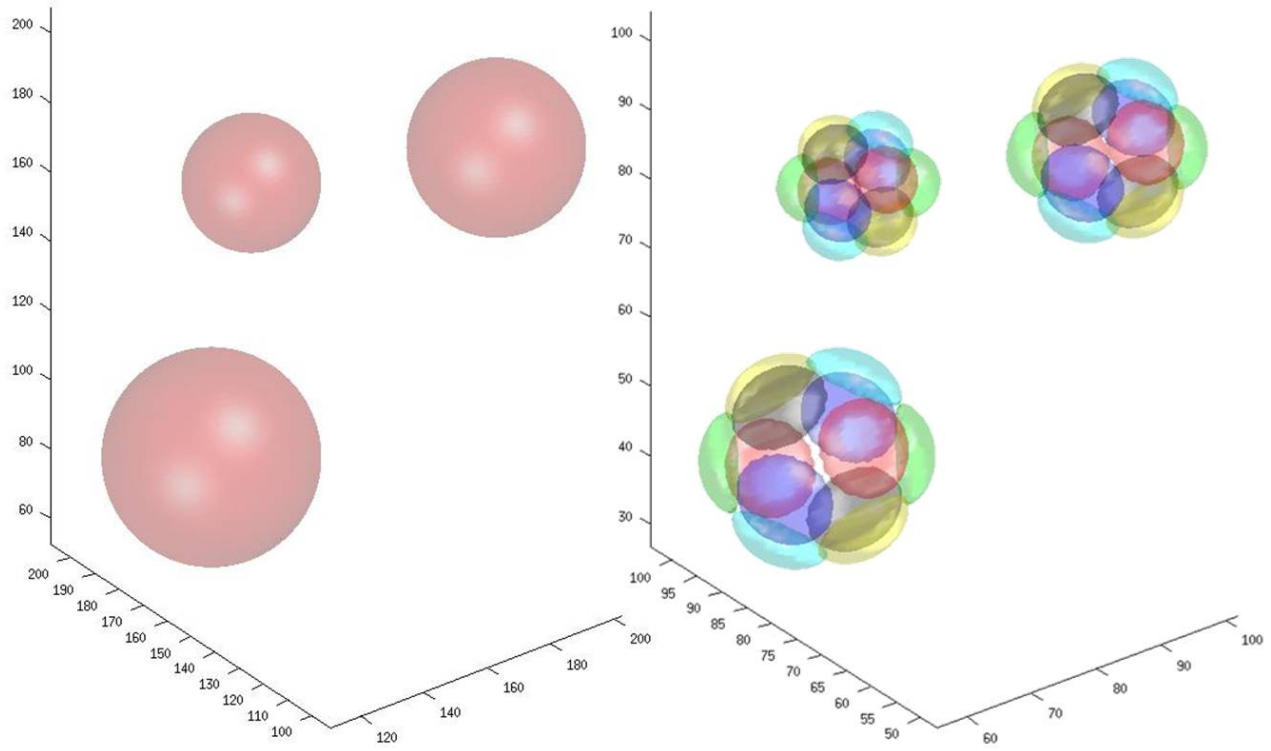


Figure 5.13: Directional filter set response on test spheres. Left: The surface visualisation of three test spheres with varying radius, hollow interior and Gaussian blurred surface in a $256 \times 256 \times 256$ test volume. Right: Directional filter set response (surface representation with maximum filter response) on the down-sampled (by a factor of 2) test volume for faster computation. The 6 different colours represent 6 different directional filters.

five components of the orientation feature maps (in the scale of $3 \times 3 \times 3 \times 5$) results in a total of 162 feature vector space!

Huge input feature space has the danger of over-training and low computational efficiency. To avoid the "Curse of Dimensionality" and re-affirm the dimensionality of our feature space is within bound, we first train our MLP system by using a simple $3 \times 3 \times 3$ (27-feature) neighbourhood intensity-only 3D patch with manual-labelled "ground-truth". As manual-labelling is time-consuming and requires expertise in differentiating the mixture of neural structure, we start by manually segmenting blood vessels (relative larger scale in comparison to neurites and parallel fibres) in the experimental rat cerebellar cortex microscopic tissue stack (see Figure 1.15). The segmentation results after training is visualized in Figure 5.14.

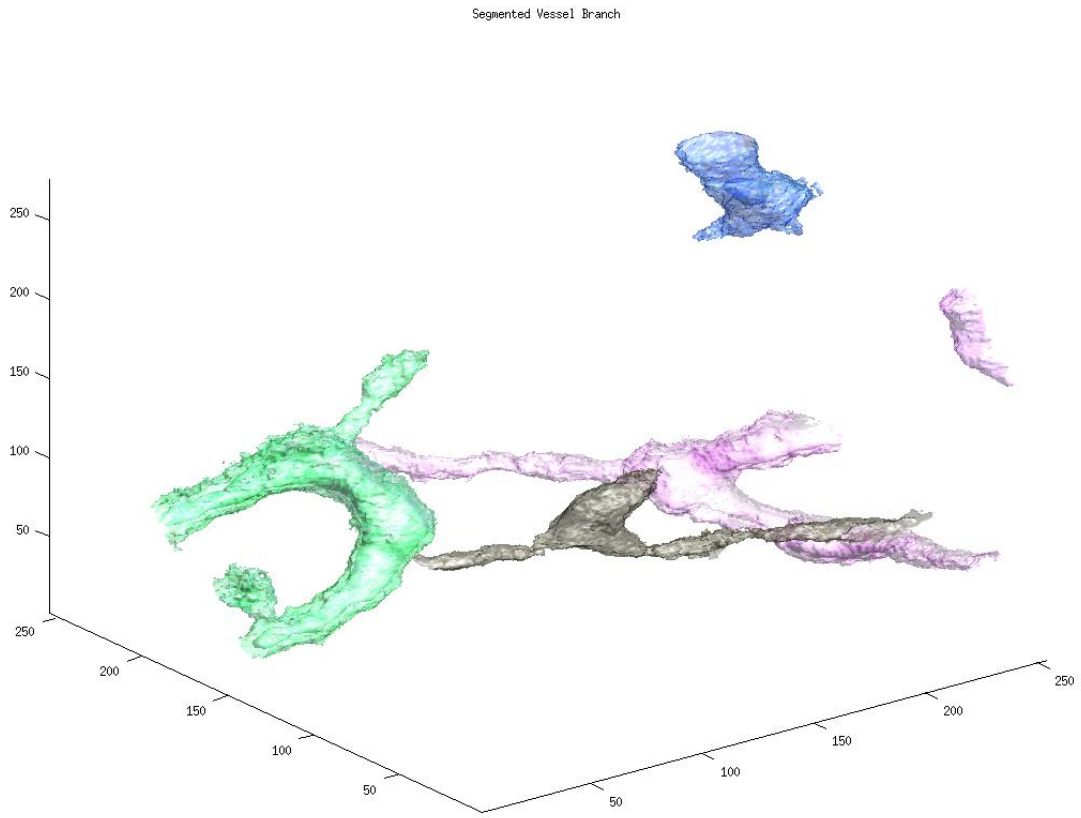


Figure 5.14: Segmented blood vessels from rat cerebellar cortex microscopic tissue stack using intensity-only volume patch for training.

We then increase the dimensionality of the feature space by adopting the **Gradient Structural Tensor (GST)** representation [14]. It is defined as the dyadic product of the gradients on image I:

$$\overline{G} := \overline{\nabla I \nabla I^t} \quad G := \nabla I \nabla I^t \tag{5.19}$$

where $\overline{(\cdot)}$ stands for averaging the tensor elements across the neighbourhood. Interpretation of the filtering requires eigenvalue decomposition of the G, for the eigenvalues and eigenvectors encode the local structure. The gradient vector ∇I is computed by convolutions with Gaussian derivatives at the scale σ_g. The size of the Gaussian weighted tensor smoothing σ_T defines the neighborhood of the image structure, and is typically 3 to 10 times larger than the

scale of $signma_g$. The operation suppresses gradient orientation contributions from noise and yields a smoother and more robust overall gradient orientation. Compared with our prior definition and properties of the Knutsson mapping, GST is different in the sense that though it is unique and polar separable, it is not distance preserving [149].

Highly structured neighbourhoods are complex thus require higher rank GST representations. The relations between the local spatial auto-correlation function and the corresponding energy distributions in the Fourier domain is illustrated in Figure 5.15.

For *planar* neighbourhood ($Rank = 1$), the orientation is constant on parallel planes. The normal vector of the planes is given by the eigenvector corresponding to the dominant eigenvalue. For *linear* neighbourhood ($Rank = 2$), the orientation is constant on parallel lines. The orientation of the line is given by the eigenvector corresponding to the smallest eigenvalue. For *isotropic* neighbourhood ($Rank = 3$), there is no dominant orientation but rather a linear combination of all eigenvectors scaled by a combination of eigenvalues.

The segmentation results after MLP training using three eigenvalues (as biological structures are defined in 3D spatial domain) from GST ($3 \times 3 \times 3 \times 3$, a 81-dimensional feature space) is visualized in Figures 5.16 – 5.19.

Lastly, we further expand our input feature vector by including both the intensity measure and the 5-D orientation mapping in an 19-connect neighbourhood patch to create a 114-dimensional feature space that we think is a compromise between depth of features encoded and the dimension of feature space.

Our unique descriptor for segmentation is assembled as follows: For each pixel, we first assemble its 19-connected neighbourhood intensity. We then separate the 5-dimensional components of the structural orientation feature, and assemble the $5 \times 18 \times 3$ neighbourhood orientation field. Here, we do not train the MLP on blood vessels, rather we train the system on artificial surrogate data sets that contain elongated neurites and Purkinje cell body (in-depth description on the method in Section 6.4).

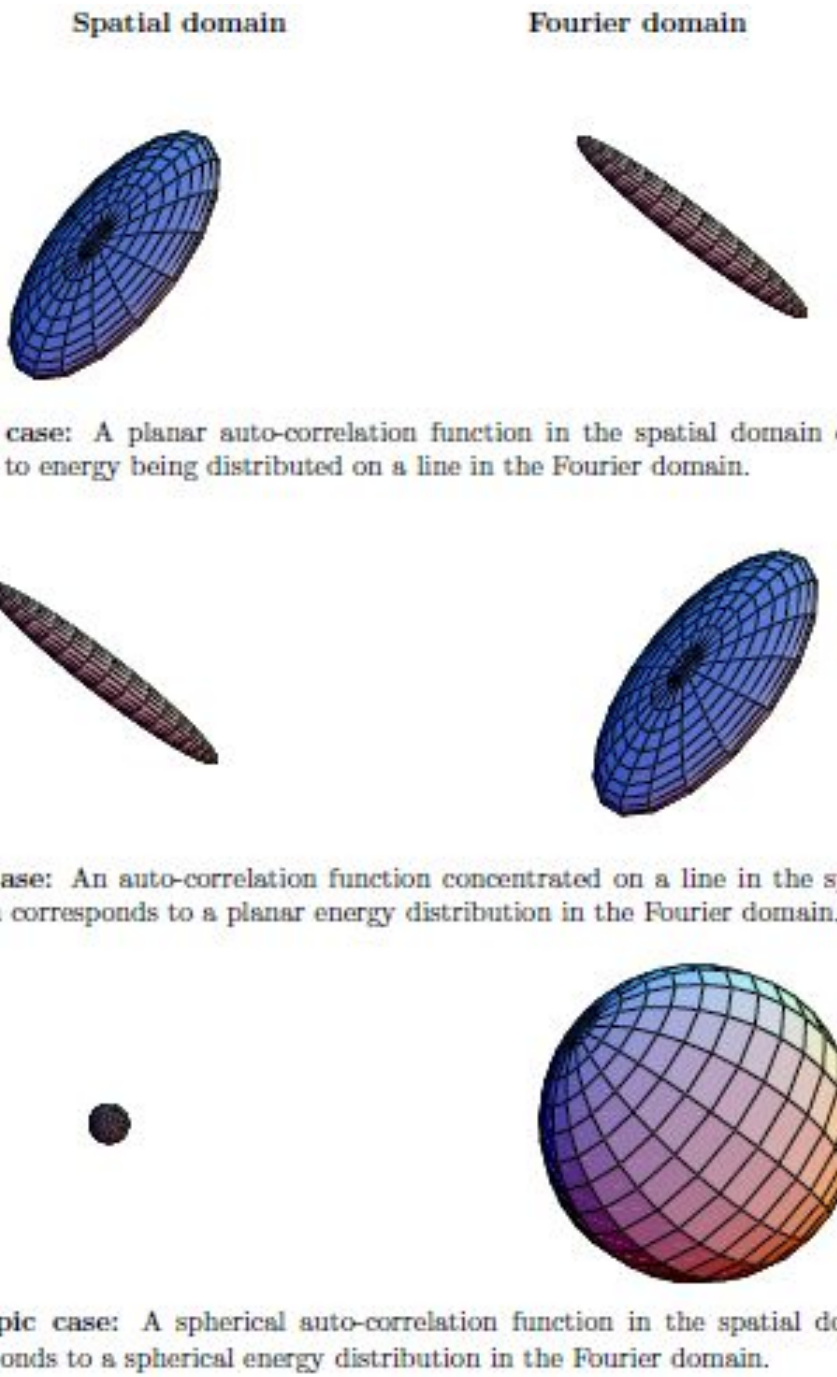


Figure 5.15: Iso-surface plots of spatial auto-correlation functions and corresponding energy distributions in the Fourier domain. Adapted from [14].

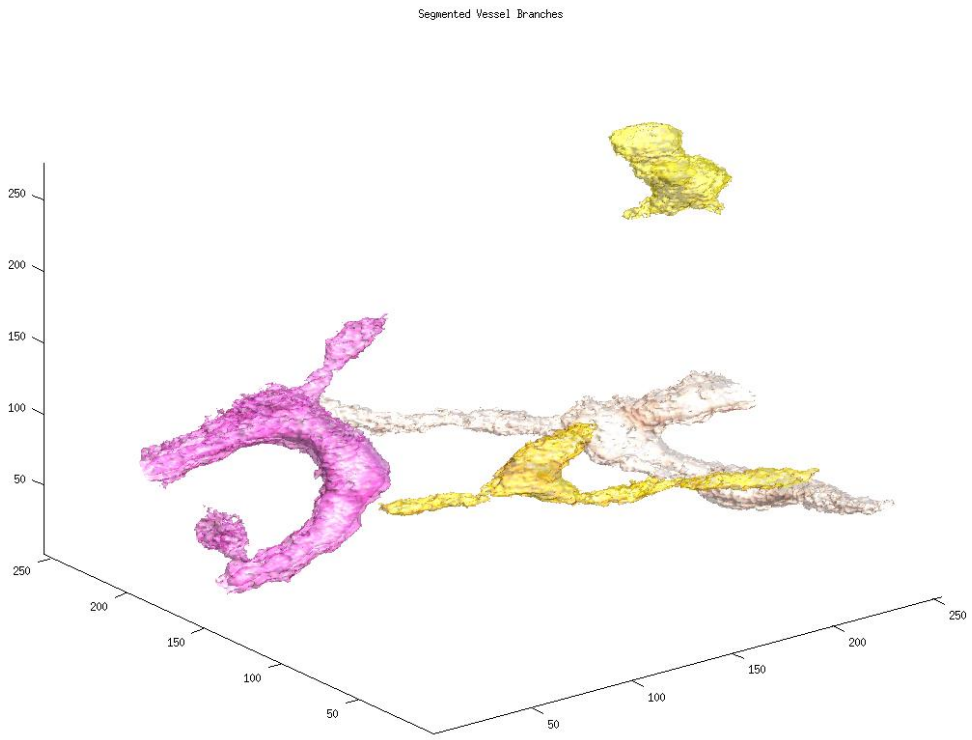


Figure 5.16: Segmented blood vessels from rat cerebellar cortex microscopic tissue stack using patch consist of 3 eigenvalues from GST for training.

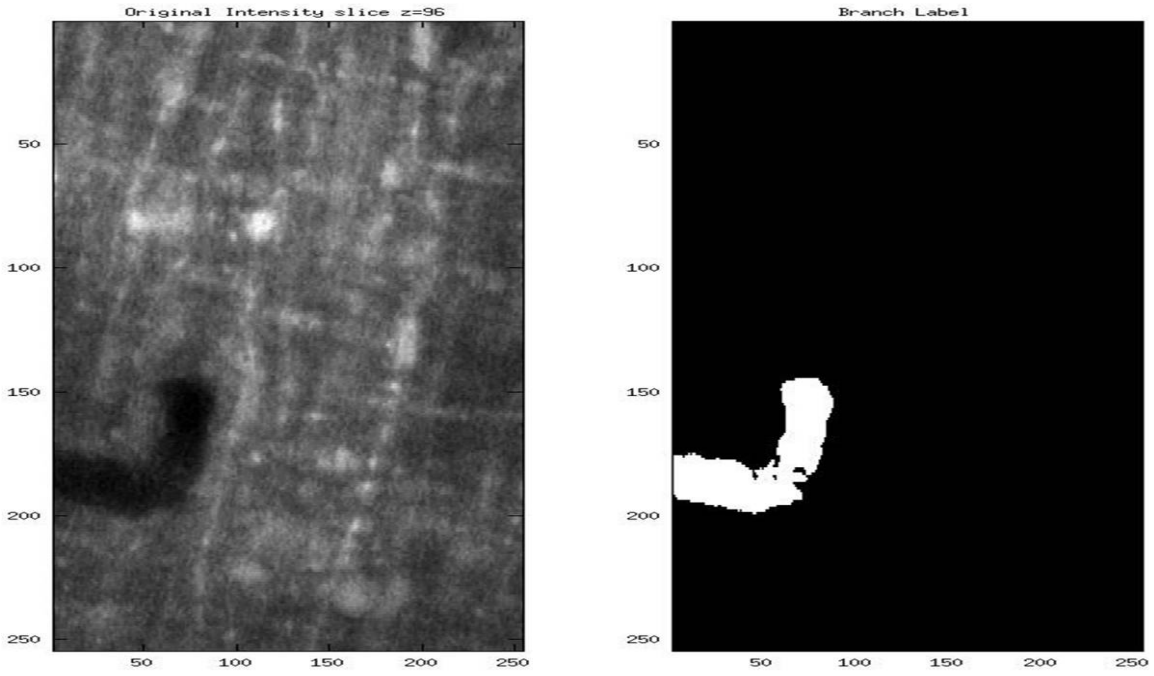


Figure 5.17: A dissection (z=96) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity).

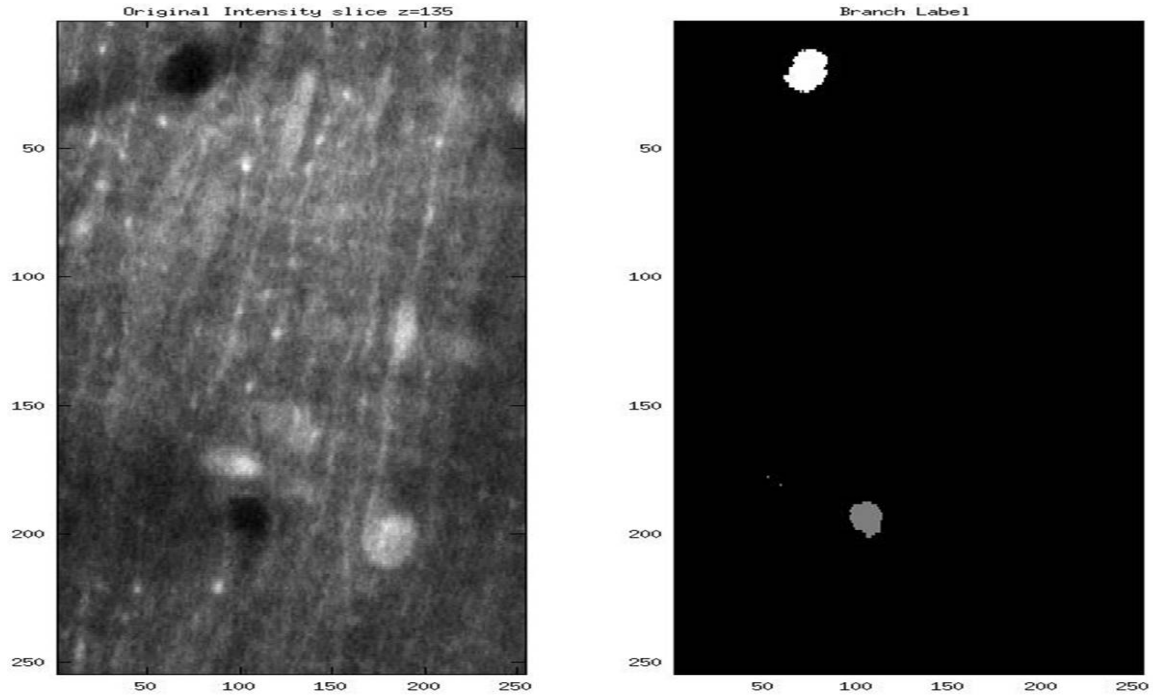


Figure 5.18: A dissection (z=135) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity).

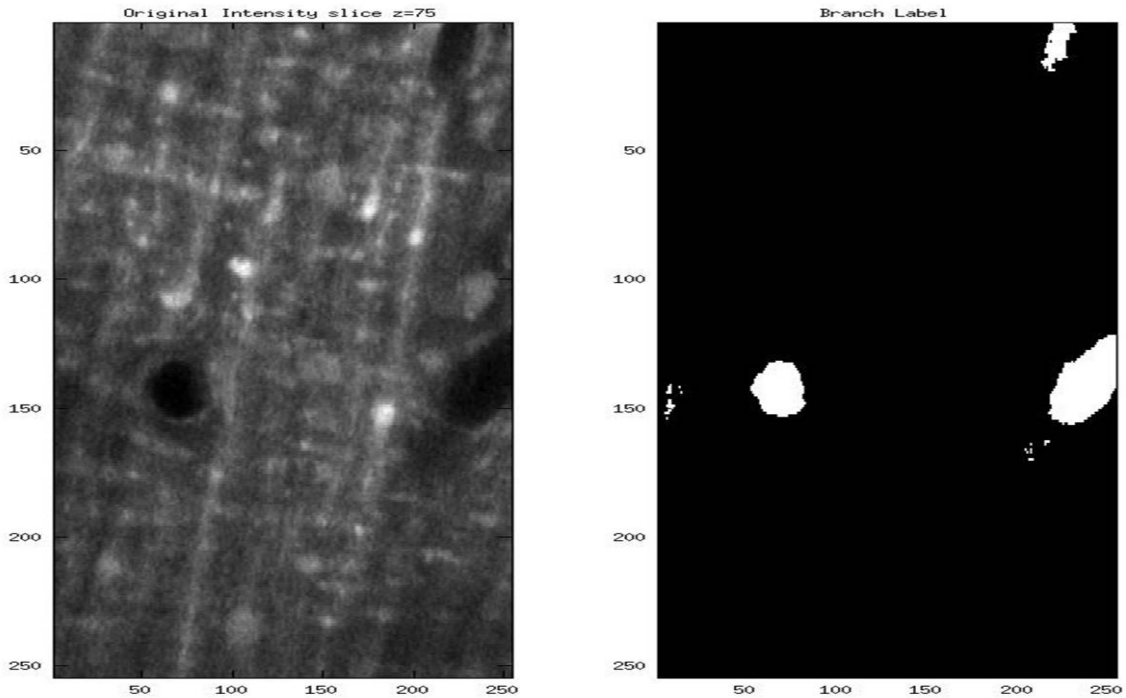


Figure 5.19: A dissection (z=75) of segmented and labelled blood vessel (right) from the two-photon microscopic stack (left, original intensity).

5.5 Discussion

In a local neighbourhood enriched with fine neurites, interneurons and parallel fibres, the neighbourhood orientation is often multi-directional. The presence of more than one dominant orientation and the discontinuous representation of eigenvector orientations in GST often result in difficulties applying structure tensor approach for neural segmentation [149]. Moreover, intensity alone or GST-based eigenvalues may not encode sufficient information to differentiate among cellular structures, such as Purkinje cells, neural dendrites and interneurons, all display certain elongation along their medial axes, yet each class has a characteristic shape captured by an orientation field that distinguishes it from other structures. As expected, even though all three input feature space (27-d intensity-only, 81-d eigenvalue from GST, 114-d intensity measure with Knutsson orientation mapping) are successful in segmenting prominent blood vessels in the microscopic stack, our 114-d feature vector yields much more successful results in segmenting Purkinje cell bodies and associated dendrites (see Section 6.4 for comparison) that are much smaller in scale and more prone to the presence of noise.

In this chapter, we first present a literature review on existing methods for (micro)vascular segmentation of neural structures. We then present our proposed pixel-level patch-based segmentation with MLP training applied in 3D neural segmentation. We start by introducing our implementation of 3D to 5D orientation mapping in order to preserve angular information and line curvature in a local neighbourhood. Our feature descriptor is constructed in an 18-connected neighbourhood that combines the intensity measure with the 5D Knutsson-mapped orientation representation. Our 114-d feature space is then used for classification on a MLP Leaning Machine to differentiate among the complex and inter-connected network of neural structures. Our descriptor, when compared to Gradient Structure Tensor approach, is able to perform better in a multi-directional complex neighbourhood as our 5D orientation representation offers more accurate encoding of the structural orientations.

Chapter 6

Evaluation and Validation

6.1 Introduction

This chapter details the extensive performance evaluation and validation of each of the techniques presented in prior chapters. In Section 6.2, we compute Receiver Operating Characteristic (ROC) curves on existing public retinal fundus image databases ("DRIVE" & "STARE") to assess and compare our proposed retinal vascular segmentation "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method with other benchmark methods. In Section 6.3, we first review existing error measures that assess registration accuracy in current literature. This is followed by the introduction of an error metric ("Centreline Error") that we believe uses a more objective representation of "ground-truth" microvasculature in fluorescein angiography and is a fairer assessment of registration quality such that the clinically interesting fine-scale microvasculatures contribute more strongly to the error metric than a spatially-averaged global measure. We then benchmark our "Joint Global RANSAC" registration method against a Global Affine registration approach on our 2D+t sequential retinal angiograms ("FASA" database). Lastly, in Section 6.4, for 3D two-photon microscopic cortical data stacks, we present segmentation results on our constructed surrogate data with known ground-truth as well as experimental data stacks. Additionally, we evaluate the benefit of our proposed 114-dimensional feature

descriptor (constructed in an 19-connected neighbourhood combining the intensity measure with the 5D Knutsson-mapped orientation representation) by comparing the segmentation results with using an intensity-only neighbourhood 3D feature patch or a "Gradient Structural Tensor" (GST) based 3D feature patch prior to the complex pixel-level MLP classification. Lastly, we conclude this chapter in Section 6.5.

6.2 2D Retinal Funduscopy: Receiver Operating Characteristic (ROC)

6.2.1 Definition: Receiver Operating Characteristic (ROC)

A Receiver Operating Characteristic (ROC) curve, is a graphical plot illustrating the performance of a classifier with varying discrimination threshold [150]. It is a plot of the fraction of true positives out of the positives ("True Positive Rate") against the fraction of false positives out of the negatives ("False Positive Rate") at various threshold settings. "True Positive Rate" (TPR) is also known as the "Sensitivity", and "False Positive Rate" (FPR) is also known as one minus the "Specificity". The ROC analysis is a standard statistical tool to select possibly optimal models and to discard suboptimal ones independently from the cost context or the class distribution [150].

Summary measures of accuracy derived from the ROC curve describe the inherent accuracy of a classifier because they are not affected by the choice of the decision threshold or the prevalence of anomalies [151]. The "Area under Curve" (AUC) is one such popular summary measure of accuracy. It is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative') [151]. It ranges in value from 0.5 (chance) to 1.0 (perfect discrimination or accuracy). A reliable and valid AUC estimate can be interpreted as the probability that the classifier will assign a higher score to a randomly chosen positive example than to a randomly chosen negative example, hence often used to evaluate classifier

performance in machine-learning. It can be calculated by using trapezoidal approximations on the ROC curve.

In addition to ROC, Accuracy (*Acc*) is also used as a statistical measure of how well the classification method correctly identifies or excludes a condition. It is defined as the proportion of true results (both true positives and true negatives) in the entire population:

$$Acc = \frac{\text{number of true positives + number of true negatives}}{\text{number of true positive + false positives + true negatives + false negatives}} \tag{6.1}$$

However, the accuracy paradox for predictive analytics states that predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy as accuracy (*Acc*) only reflects the ratio of correct predictions relative to the total number of cases evaluated.

6.2.2 DRIVE Database: Results

We compared our "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method with 10 previously benchmarked methods [60] [11] [50] [2] [57] [48] [119] [152] [52]. Figures 6.1 and 6.2 are the "Posterior Probability Map" or "soft decision" segmentation results of the global vessel-bed of two test images, whereas Figures 6.3 and 6.4 demonstrate the segmentation results of the entire test image set, a total of 20 images.

For training, we generated 4 different batches of random patch (similar to a discretized Wiener process) on each training image: 4000 pixels (0.89% of total image pixels), 8000 pixels (1.78%), 12000 pixels (2.68%) and 16000 pixels (3.57%). Each patch contains half vessel pixels and half background pixels. We compared MLP models constructed from training based on a total of 8×10^4 pixels, 1.6×10^5 pixels, 2.4×10^5 pixels and 3.2×10^5 pixels. Figure 6.5 represents the ROC curves of the four MLP models with incrementing numbers of training pixels. The notable feature of this series of ROC curves is that the model "learned" from the **most** training pixels (in thick dotted aqua line) does not statistically perform better classification than the the model "learned" from the **fewest** training pixels (in thin dotted

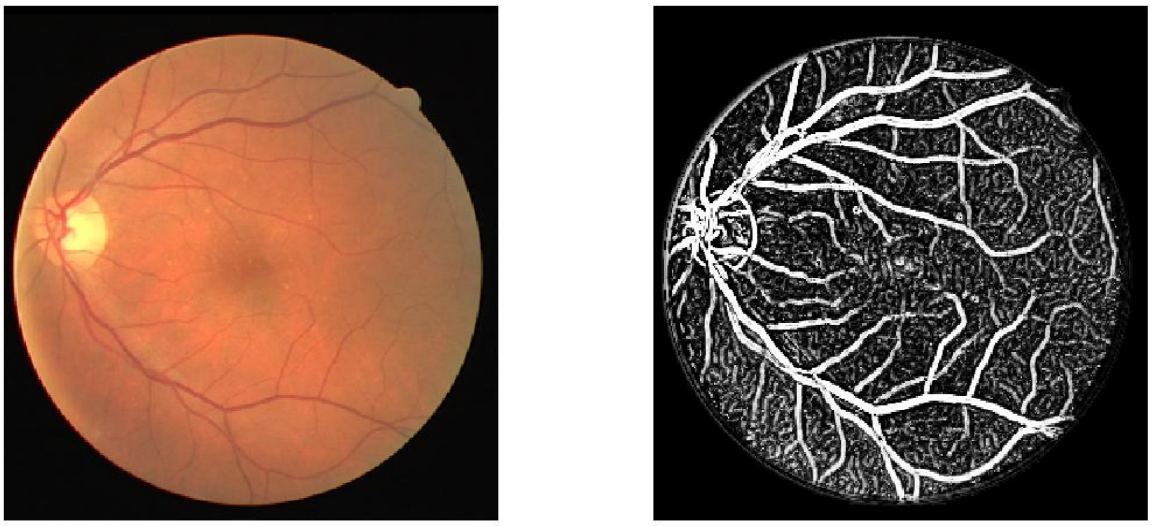


Figure 6.1: The segmentation results (right) of our **PP-MLP** algorithm on Test Image 1 (left).



Figure 6.2: The segmentation results (right) of our **PP-MLP** algorithm on Test Image 20 (left).

faint blue line). We can conclude that the models we employ here are **not** "over-trained" (common in machine-learning approaches), and yield meaningful and statically stable classification results.

Additionally, to assess the variability of the high dimensional classification using a complex decision surface, we apply the same MLP model on each of the 20 test images in the DRIVE database (see Figure 6.6). We can observe the majority of ROC curves cluster tightly in the upper left corner of the ROC space. Four curves shift slightly away from the cluster, this is

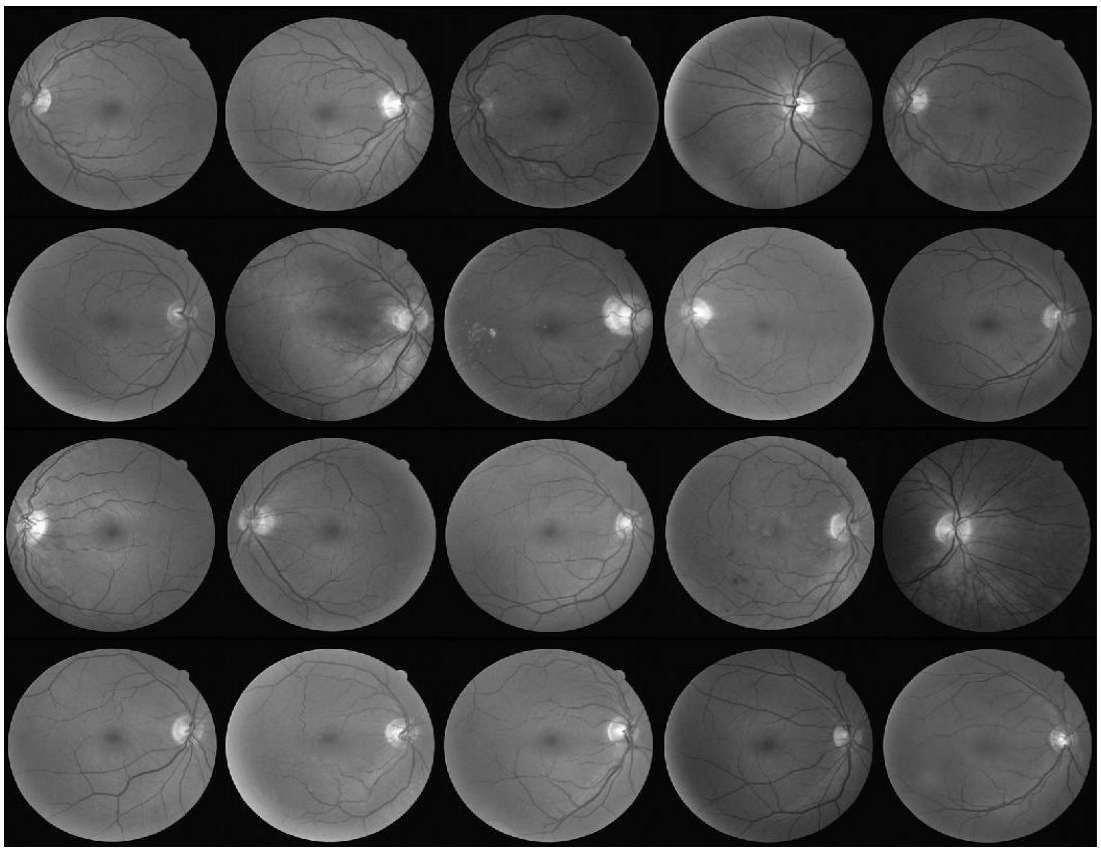


Figure 6.3: 20 Test images from the DRIVE Database.

due to the fact that DRIVE test database include 4 images (Image 3, 7, 8 and 14) containing pathology, namely exudates, hemorrhages and pigment epithelium changes. As the diseases-related retinal vasculature pixels only contribute to 20% of overall pixels for training, the classifier is naturally more successful in segmenting healthy vasculature than the pathological ones.

The ROC for the entire test data superimposed on the ROC lower and upper bound are depicted in Figure 6.7. The algorithm performance on individual images is an indicator of the technical usability, especially in a clinical setting.

Table 6.1 compares the area under the Area under curve (AUC) and the Average Accuracy (*Acc*) for all ten segmentation methods. Our PP-MLP method achieved 0.9779 for AUC and 0.9270 for average accuracy. The higher AUC and slight lower *Acc* from our method is because our classifier returns **fewer** false positives and false negatives, while capturing slightly fewer true positives and true negatives, when compared to methods with higher Acc but lower AUC.

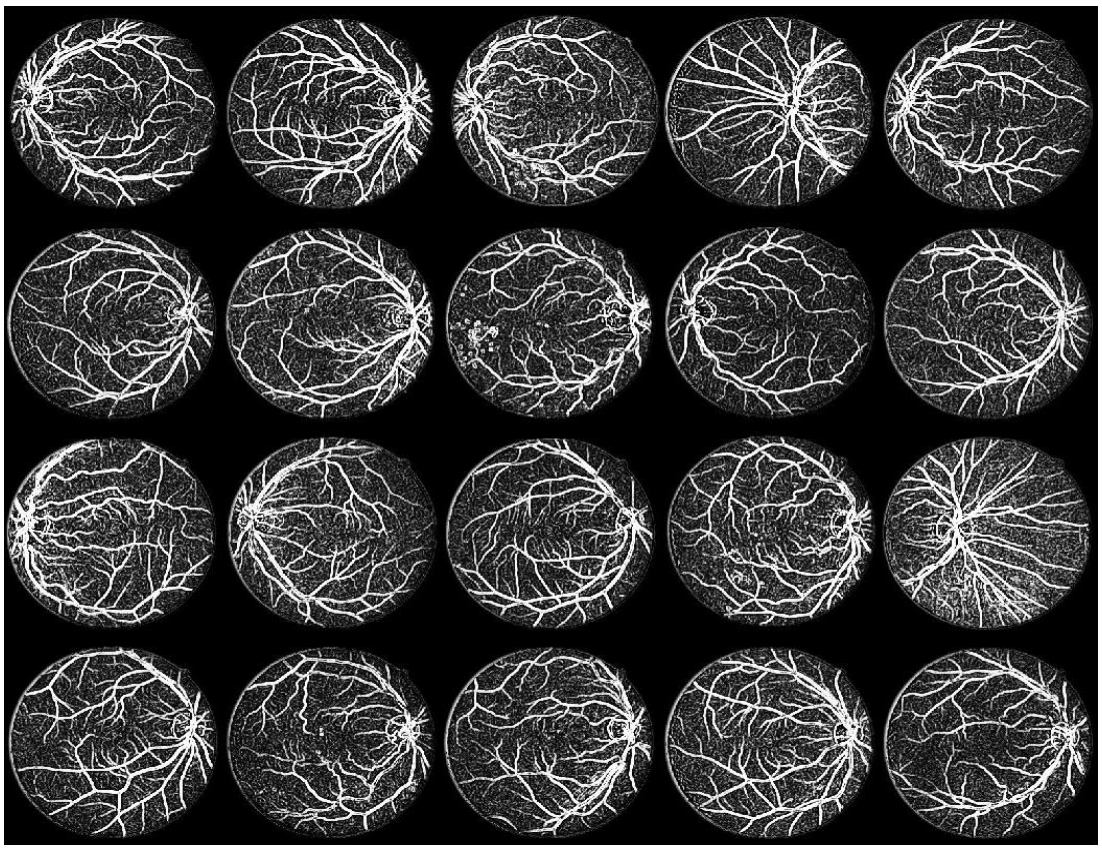


Figure 6.4: The segmentation results (posterior probability map, or "soft decision") of our **PP-MLP** algorithm on all the test images.

6.2.3 STARE Database: Results

We compared our "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method with 8 previously benchmarked methods [60] [2] [50] [3] [119] [52].

For training, we generated 5 different batches of random patch (similar to a discretized Wiener process) on each training image: 2000 pixels (0.47% of total image pixels), 4000 pixels (0.95%), 6000 pixels (1.42%), 8000 pixels (1.89%) and 10000 pixels (2.36%). Each patch contains half vessel pixels and half background pixels. We compared 5 MLP models constructed from training based on a total of 2×10^4 pixels, 4×10^4 pixels, 6×10^4 pixels, 8×10^4 pixels and 1×10^5 pixels.

Figure 6.8 represents the ROC curves of the five MLP models with incrementing numbers of training pixels. We intentionally added an additional model trained with fewer than 0.5% total image pixels to compare with the rest of the four models. The notable feature of this

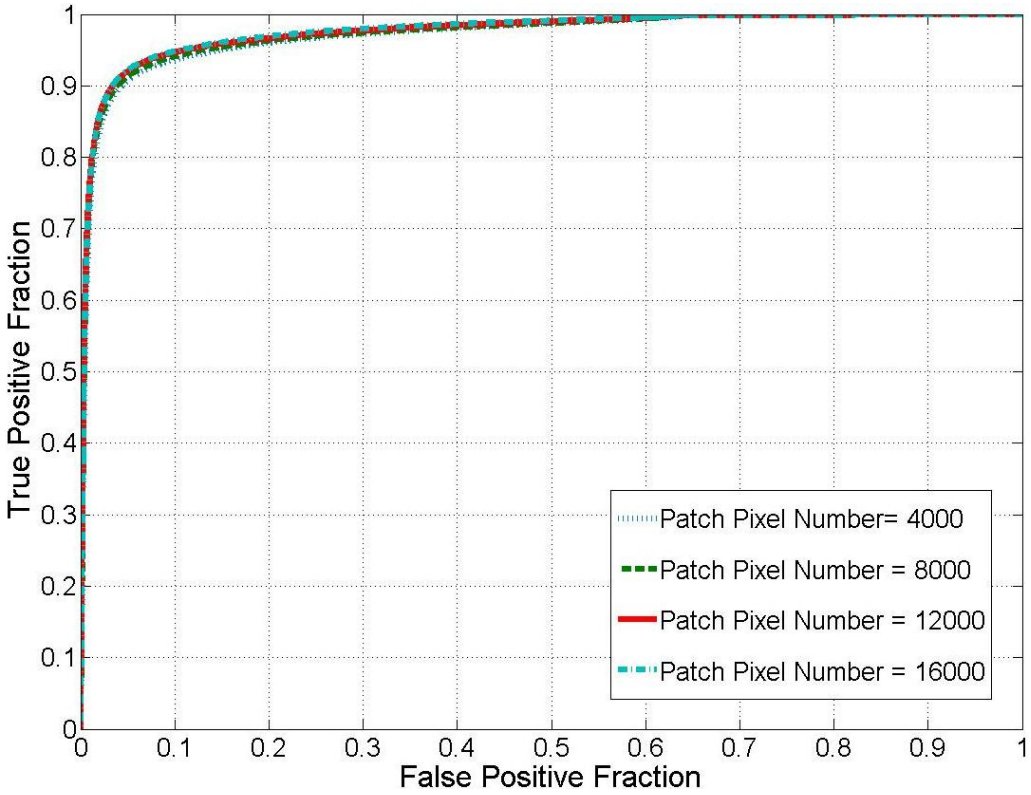


Figure 6.5: The ROC curves of our proposed "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method evaluated on the DRIVE database with different training patch size, incrementing from less than 1% to 3.6% of the total image pixels.

No.	**Method**	**Area under curve** (*AUC*)	**Average accuracy** (*Acc*)
1	Proposed Method (PP-MLP)	0.9779	0.9270
2	Soares at al. [60]	0.9598	0.9466
3	Niemeijer et al. [11]	0.9294	0.9416
4	Jiang et al. [50]	0.9114	0.9212
5	Staal et al. [2]	0.9441	-
6	Zana et al. [57]	0.8984	0.9377
7	Chaudhuri et al. [48]	0.7878	0.8773
8	2^{nd} human observer [11] [2]	-	0.9473
9	Mendonca et al. [119]	-	0.9452
10	Al-Diri et al. [152]	-	0.9258
11	Martínez-Pérez et al. [53]	-	0.9181

Table 6.1: *AUC* and *Acc* comparison for 10 different segmentation methods and human observer on 20 test images from DRIVE database.

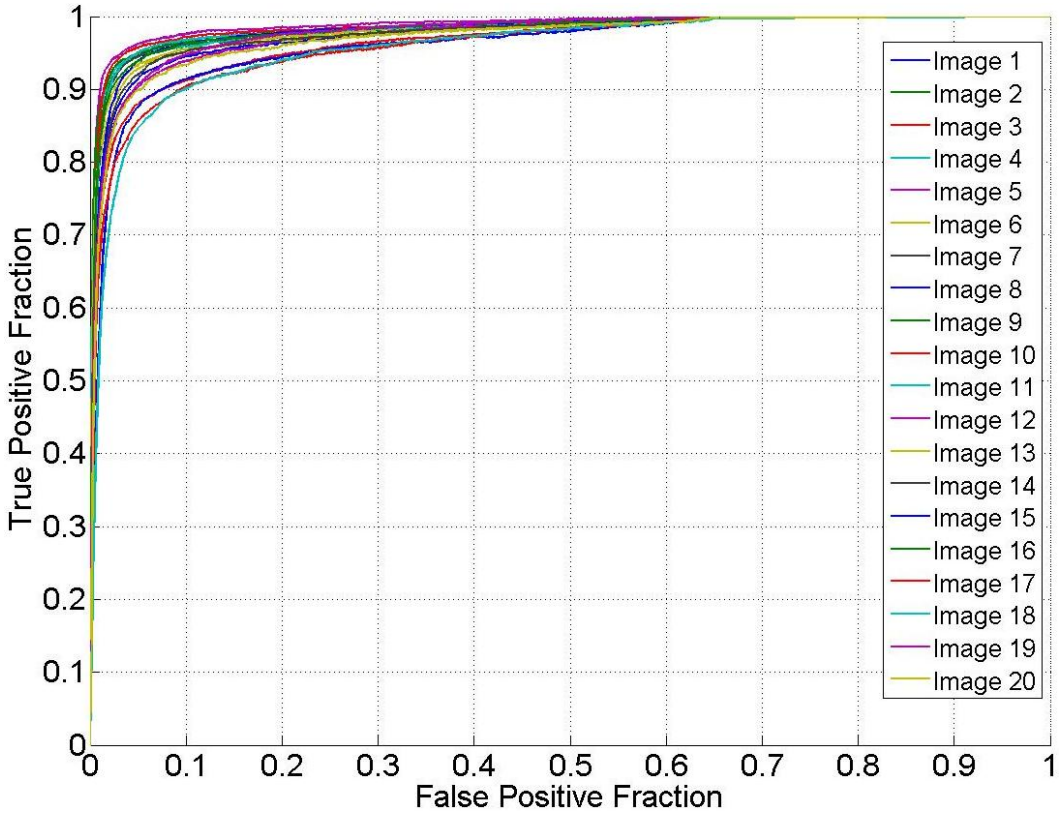


Figure 6.6: The ROC curves for 20 test images using one MLP model.

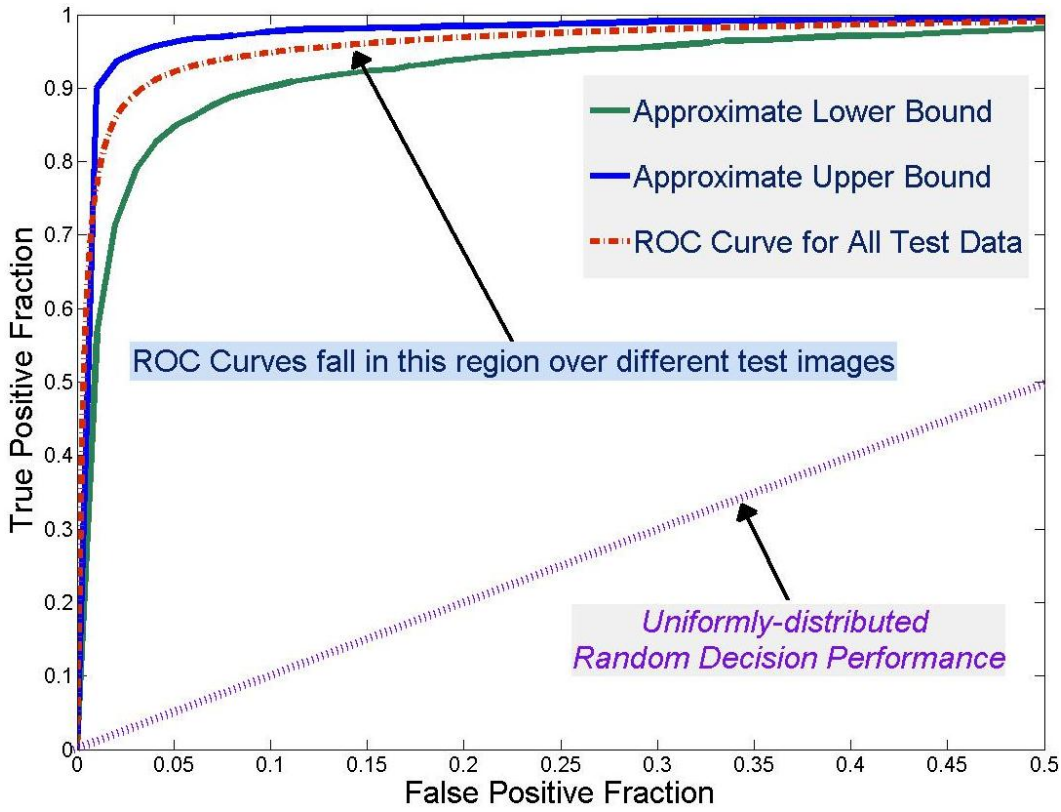


Figure 6.7: The ROC curves for lower and upper bound, and for the entire test data.

series of ROC curves is that the model "learned" from the **most** training pixels (in green line) does not statistically perform better classification than the model "learned" from fewer training pixels (in blue, yellow and pink lines). However, the model "learned" from the **fewest** training pixels (in purple line) does perform noticeably less well than the others. We can conclude that the 4 models constructed with larger training patches are **not** "over-trained", and yield meaningful and statically stable classification results; whereas the model with the smallest training patch is "under-trained". This is further supported by the AUC and *Acc* values from Table 6.2 with varying patch size.

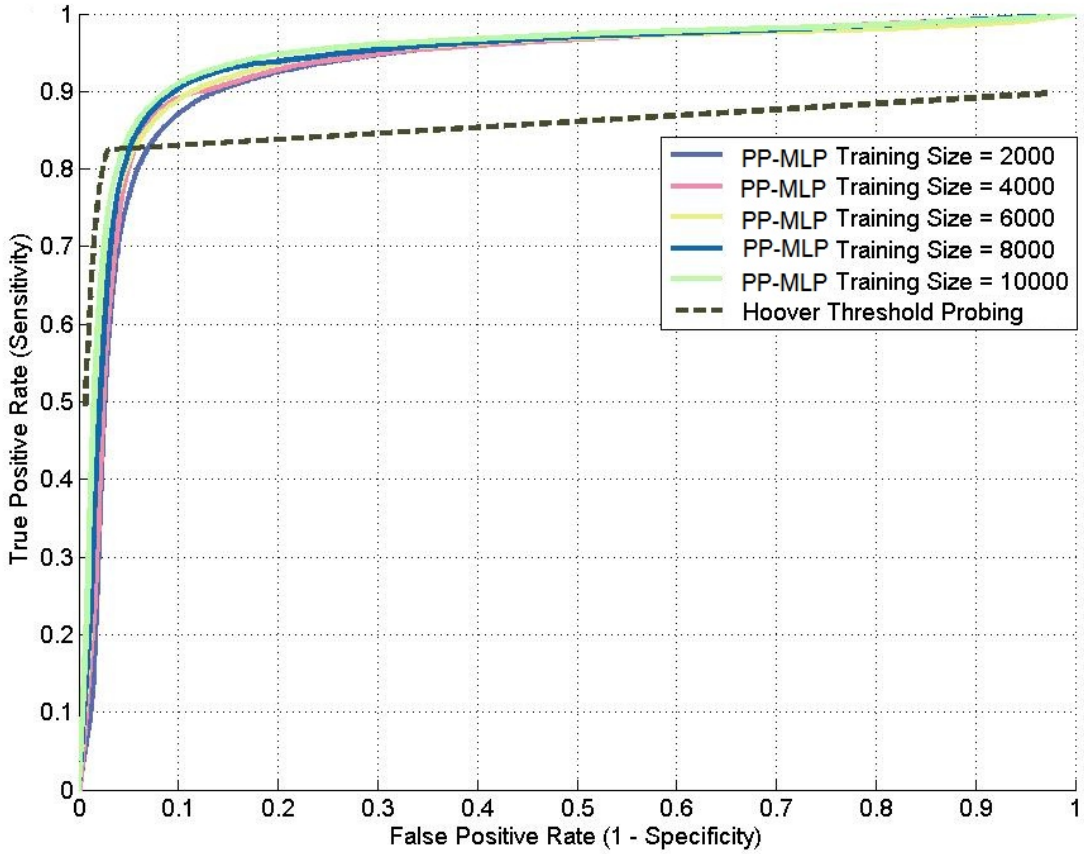


Figure 6.8: The ROC curves of our proposed PP-MLP method evaluated on the STARE database with different training patch size, incrementing from less than 0.5% to around 2% of the total image pixels. The performance of Hoover threshold probing approach is superimposed in dotted black line.

Table 6.3 compares the area under the Area under curve (AUC) and the Average Accuracy (*Acc*) for all eight segmentation methods. Our PP-MLP method achieved 0.9605 for AUC and

Patch Size	**Area under ROC** (*AUC*)	**Average accuracy** (*Acc*)
Sample Size = 10k	0.9490	0.9605
Sample Size = 8k	0.9417	0.9540
Sample Size = 6k	0.9386	0.9556
Sample Size = 4k	0.9349	0.9508
Sample Size = 2k	0.9307	0.9490

Table 6.2: *AUC and Acc comparison for our PP-MLP method evaluated on the STARE database with different training patch size, incrementing from less than 0.5% to 2% of the total image pixels.*

0.9490 average accuracy. After combining the AUC and *Acc* results with prior ROC curve from Figure 6.8, we can deduce that Hoover threshold probing approach has the tendency to "under-segment" the likely vessel pixels. The steeper ROC curve of Hoover's methods indicates high true positive rate and low false positive rate. But if the positive rate is rather low, both measures can be suppressed. However, this would in turn results in lower AUC and average accuracy, as more vessel pixels are classified as background (higher false negative rate).

No.	**Method**	**Area under ROC** (*AUC*)	**Average accuracy** (*Acc*)
1	Proposed Method (PP-MLP)	0.9605	0.9490
2	Soares at al. [60]	0.9651	0.9474
3	Staal et al. [2]	0.9614	0.9516
4	Jiang et al. [50]	0.9298	0.9009
5	Hoover et al. [3]	0.8329	0.9275
6	2nd human observer [3]	-	0.9348
7	Mendonca et al. [119]	-	0.9440
9	Martínez-Pérez et al. [53]	-	0.9232

Table 6.3: *AUC* and *Acc* comparison for 8 different segmentation methods and human observer on 20 images from STARE database.

6.3 2D+t Sequential Fluorescein Angiogram: Centreline Error

6.3.1 Context & Literature Review : Registration Accuracy for 2D+t Sequential Fluorescein Angiography

Registration error can be determined by the extent of misalignment between the registered image and the reference image. Conventional ground truth is obtained from manual registration [57]. This form of reference standard is subject to inter- and intra-observer variability. Therefore, [106] linked lines between feature locations on the original images as relatively "unbiased" ground truth. However, both the linking algorithm that connects vessel fragments to subpixel precision and the similarity measure that optimally matches to subpixel accuracy between different frames vary from case to case in establishing error metrics. Therefore, [109] casted an evenly spaced virtual grid that intersects with vessels to obtain ground-truth pixels. This, however, does not sufficiently take account of the location-dependent nature of pixel information. Pixels located near the capillary-rich macula region generate significantly more clinical interest than those near the "field-stop". Furthermore, their proposed "error tracing" route theoretically favors strategies that pair both the forward and the inverse registration functions to optimize the "net offset" rather than a true assessment of the registration algorithm on its own. Additional processing is also required to address the uneven global illumination.

We believe that the microvascular centrelines are more structurally stable and more patient specific, hence, we favour a more objective representation of "ground-truth" microvasculature for fluorescein angiography. However, the estimate for the centreline location of the vessels requires subpixel resolution and high accuracy. In practice, fine vessel structures may not be captured in all frames within a sequence. It is commonly observed that capillaries may "disappear" from the previous frame then "re-emerge" in the following one. Furthermore, due to the time delay in the passage of dye, angiogram frames at the beginning and the end of a

sequence reveal significantly less information on the detailed microvasculature.

6.3.2 Definition: Centreline Error

Here, we choose the temporal average of the registered sequence(s) as a more perceptually accurate representation of the retinal microvasculature as it takes into account previous and subsequent frames in one or more sequences (Figure 6.9). For globally registered sequence(s) $\{f_n''\}_{n\in\{1,2,\ldots,N\}}$ with N number of frames, the temporal average across the sequence(s) is defined as:

$$f_{average} = \frac{1}{N}\sum_{n=1}^{N} f_n'' \tag{6.2}$$

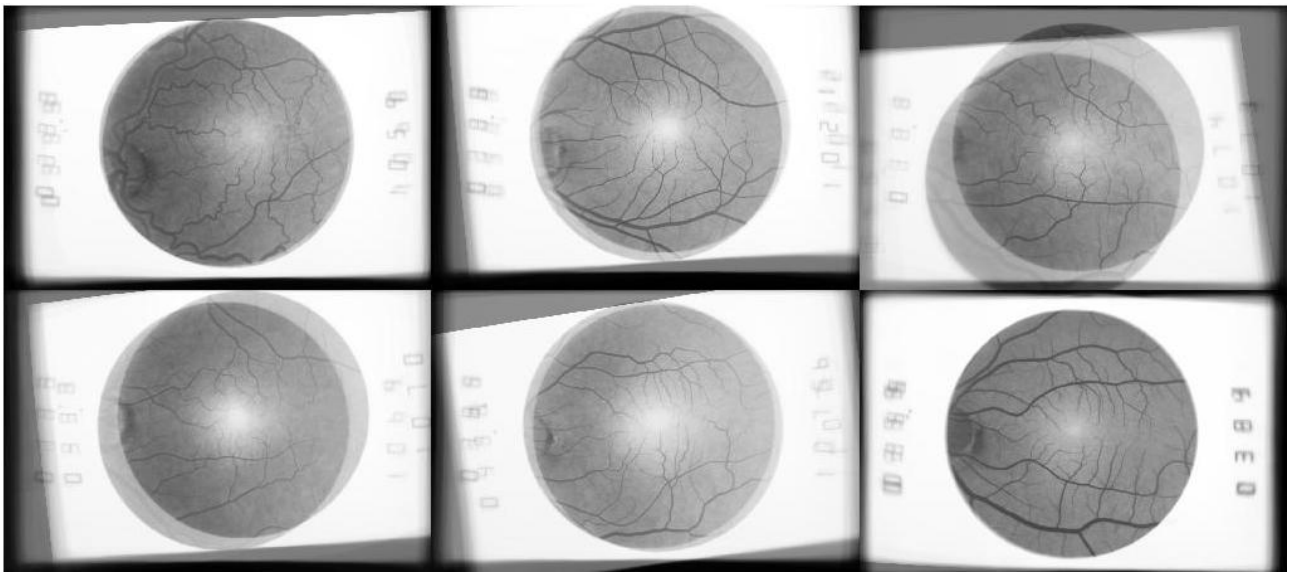


Figure 6.9: Temporal average $f_{average}$ of registered sequences (pre- and post- coronary artery bypass graft surgery) for each patient (a total of 6 patients).

To reduce statistical bias in our ground-truth estimate, we employ the leave-one-out cross-validation (LOOCV) strategy [153]. To evaluate the registration accuracy for frame f_n'', we extract its **ground-truth** microvasculature on a modified leave-one-out average frame g_n:

$$g_n = \frac{N}{N-1}\ f_{average} - \quad \frac{1}{N-1}\ f_n'' \tag{6.3}$$

For our validation method, we first perform centreline extraction of the retinal vessel bed of

all our co-registered pre- and post-op sequences, and each corresponding leave-one-out average $\{g_n\}_{n\in\{1,2,\ldots,N\}}$. We then select a patch (ROI) near the macula region with a high density of microvasculature. For each frame f_n, we establish its ground-truth vasculature as the centreline locations $v_{n(true)}(x,y)$ within the ROI on g_n. Lastly, we establish correspondence by cross-correlation between the intensity fields of the averaged, co-registered centrelines with the per-frame putative centreline estimates. This allows us to compare frame-by-frame the vessel centreline (sub)pixels $v_n(x,y)$ from the same ROI with its matching ground-truth. For a given ROI area with L (sub)pixels on the centreline, we define the **Centreline Error** (CE) as:

$$CE = \frac{1}{L}\sum_{n=0}^{L-1}\alpha_n\|v_n(x,y) - v_{n(true)}(x,y)\| \tag{6.4}$$

$\|\cdot\|$ is the Euclidean distance. α_n is a positive weighting factor that can be adjusted to characterize the neighborhood around the (sub)pixel. Here, we use $\alpha_n = 1$ assuming even contribution from all centreline (sub)pixels. This allows us to compare against a standard affine transformation model. However, in other cases, for instance patients with proliferative retinopathy, α_n can reflect the local vessel tortuosity. This is a fairer assessment of registration quality as the clinically interesting fine-scale microvasculatures contribute more strongly to the error metric than a spatially-averaged global measure.

6.3.3 Results

For image registration and centreline extraction, we compared our algorithms with an affine model [154], that accounts for pairwise rotation, scale, translation and shearing. This model is known as a type of "shape-preserving mapping", as it preserves angles and curvatures. We optimized our "control" affine model by minimising the first order approximation of the geometric error (Sampson distance) of the fit of the transformation matrix [115].

Our segmentation evaluation used ROIs of size 101×101 pixels near the macula containing complex image structures, and with clearly displayed capillaries ranging from 5-10 pixels in width. The validation and segmentation procedures were controlled, so that the same ROIs

were used in both models. Both algorithms ran on the same machine, and both algorithms used the same local and global references to give a fair comparison.

Figures 6.11 – 6.14 are four such sub-regions near the macula from four different frames in the inter-sequentially registered pre- **and** post-operative sequences of one patient. There are overall 66 frames, 33 in the pre-op sequence and 32 in the post-op sequence. We select at random the $12_{th}, 31_{st}, 49_{th}, 60_{th}$ frame (with first two from the pre-op sequence and latter two from the post-op sequence) to illustrate and visualize how our method works. The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ established are depicted in blue, whereas the frame-specific putative microvascular centreline estimates $v_n(x, y)$ where $n = 12, 31, 49, 60$ are depicted in green. The one-by-one correspondence by cross-correlation between each $v_{n(true)}(x, y)$ and $v_n(x, y)$ where $n = 12, 31, 49, 60$ are highlighted by red diamond shapes. The matched pairs contribute towards the computation of Centreline Error (CE). Lastly, Figure 6.10 is a montage of the same sub-region taken from all 66 frames in the registered sequences.

Table 6.4 presents the comparison results of **Centreline Error** (CE) between the outputs of the two algorithms across all 384 retinal angiograms. For fine structures, the affine model has on average 3.2 pixel misalignments and a much larger range of nearly 13 pixels. This is realistic as we retain the original image resolution at 4288×2848 for the statistical comparison of the microvasculature misalignment between the two methods. Our model, on the other hand, has a much more stable performance at 0.11 mean (distance) pixel misalignment and a much tighter range of 0.29 pixel in all registered frames.

Method	**Mean**	**SD**	**SEM**	**Range**	**95% Confidence Interval**
Joint Global-RANSAC	0.113	0.057	0.003	0.286	[0.108 0.119]
Global Affine	3.205	2.771	0.143	12.980	[2.923 3.487]

Table 6.4: *Statistical comparison of Centreline Error (CE) between our joint Global-RANSAC registration scheme with a controlled Global Affine transformation model on 384 angiograms.*

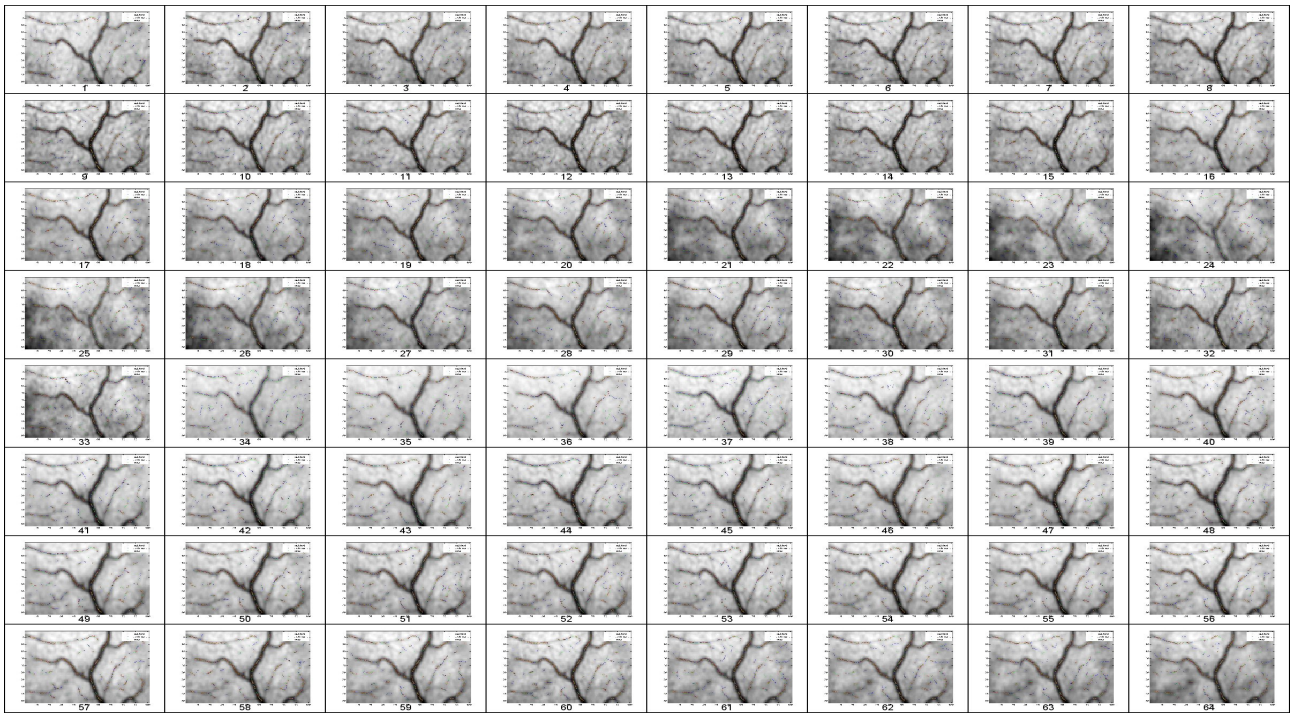


Figure 6.10: Same sub-region in all 66 frames for the inter-sequentially registered pre- **and** post-operative sequences of one patient, all with extracted frame-specific putative microvascular centreline estimates.

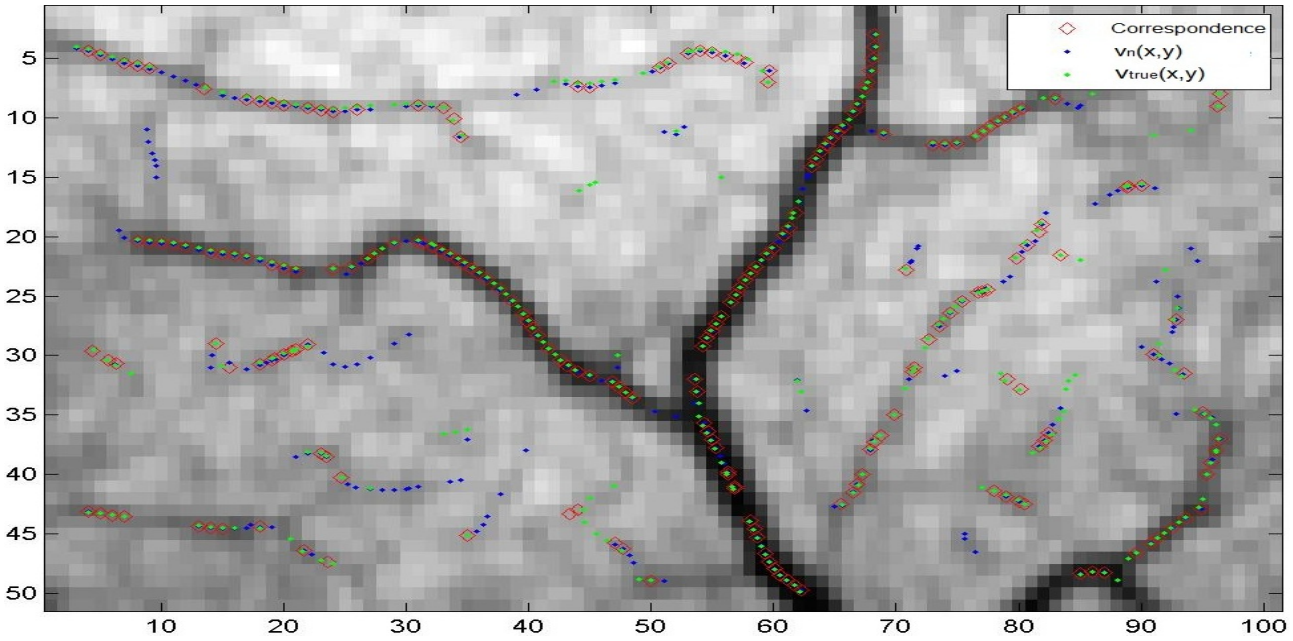


Figure 6.11: The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=12}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 12_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=12}(x, y)$ are highlighted by red diamond shapes.

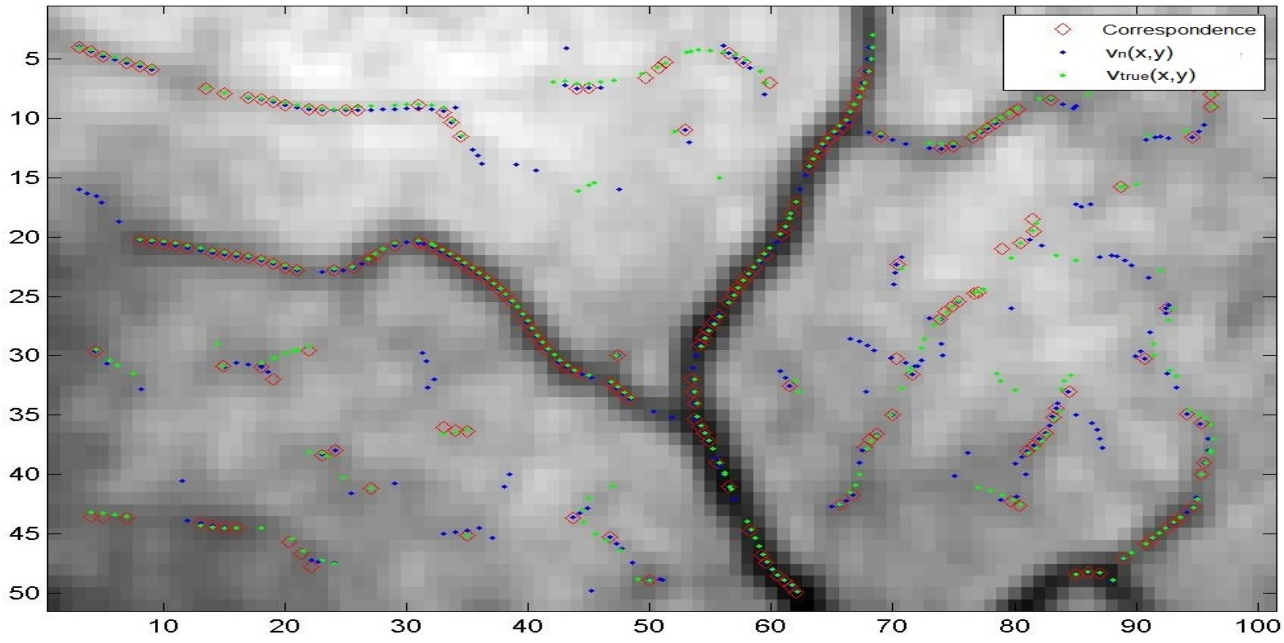


Figure 6.12: The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=31}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 31_{st} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=31}(x, y)$ are highlighted by red diamond shapes.

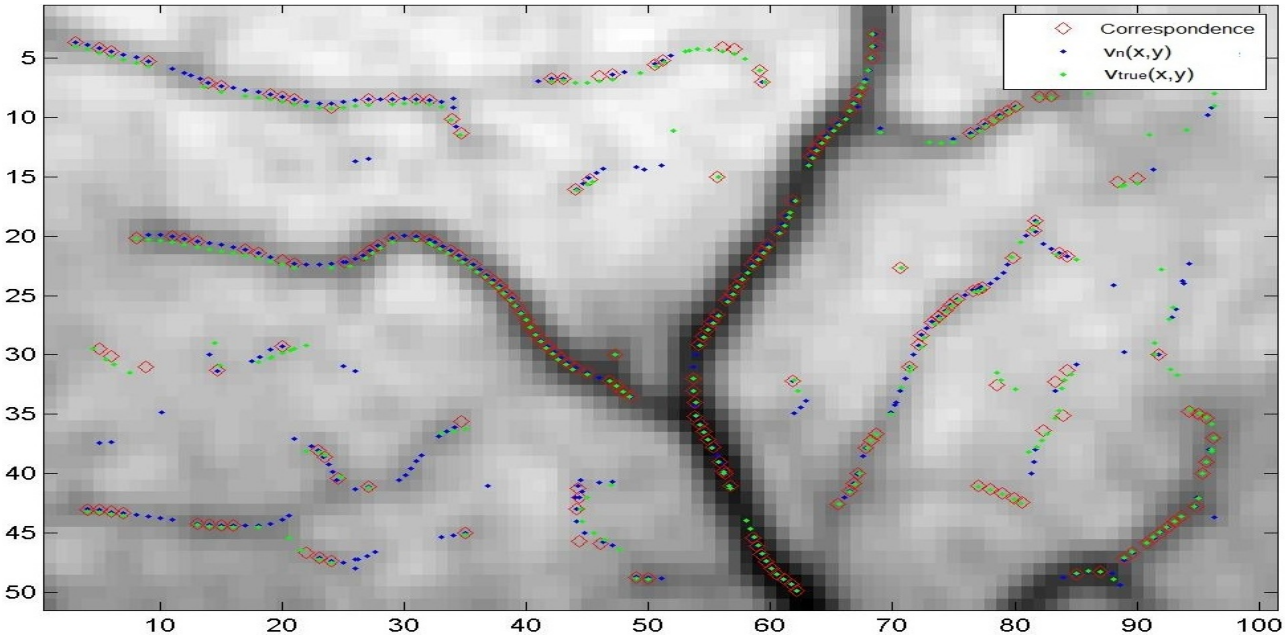


Figure 6.13: The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=49}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 49_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=49}(x, y)$ are highlighted by red diamond shapes.

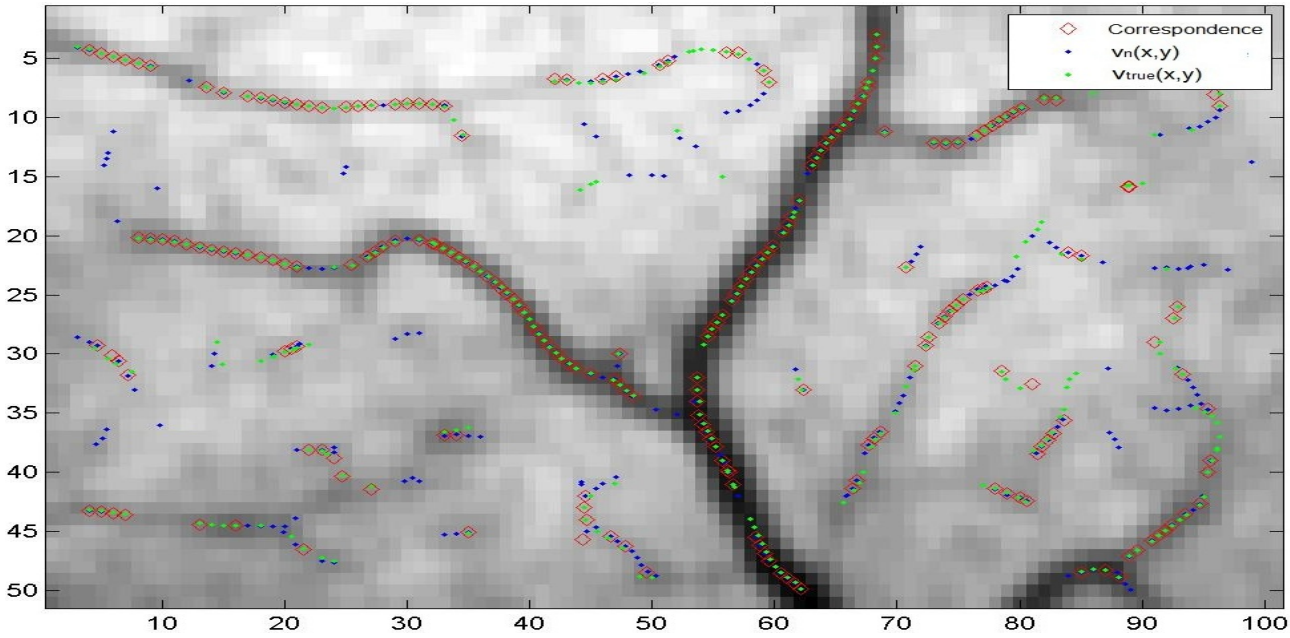


Figure 6.14: The "ground-truth" microvasculature locations $v_{n(true)}(x, y)$ (in blue) and the frame-specific putative microvascular centreline estimates $v_{n=60}(x, y)$ (in green) superimposed on a sub-region (101×51) near the macula on the 60_{th} frame of the inter-sequentially registered pre- **and** post-operative sequences. Note, the matched one-to-one correspondence between each $v_{n(true)}(x, y)$ and $v_{n=60}(x, y)$ are highlighted by red diamond shapes.

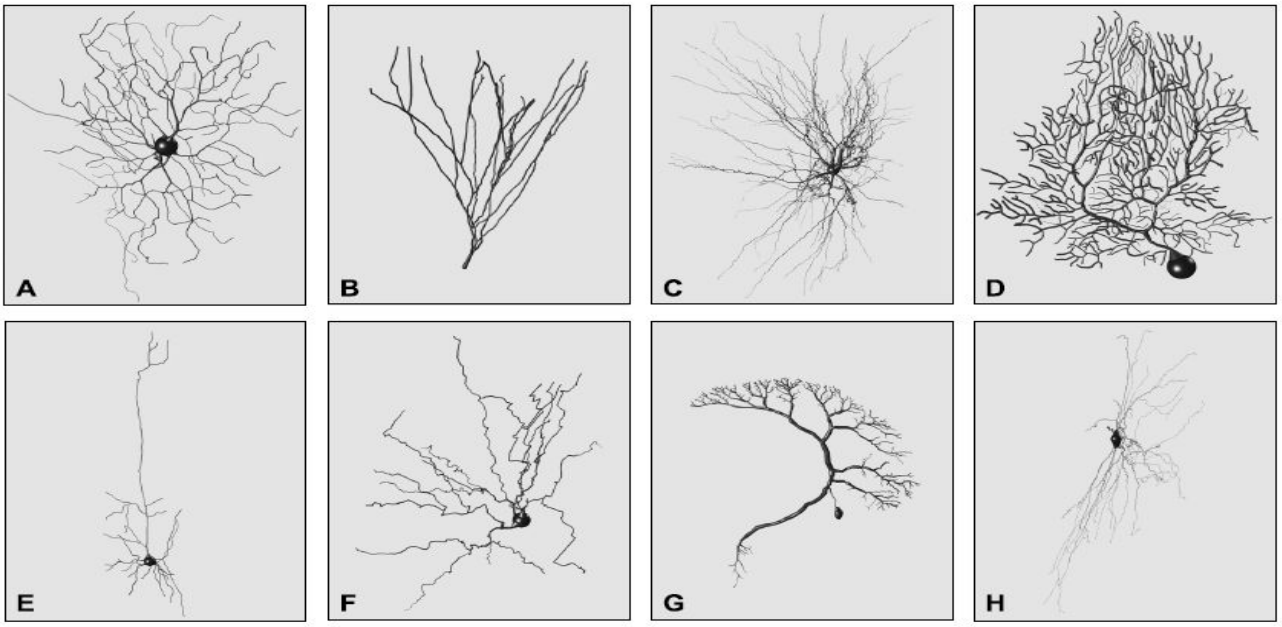


Figure 6.15: Sample neurons from the NeuroMorpho database [7] based on principal cell types: (A) ganglion cell (mouse retina), (B) granule cell (rat hippocampus), (C) motoneuron (cat spinal chord), (D) Purkinje cell (mouse cerebellum), (E) pyramidal cell (rat cerebral cortex), (F) stellate cell (rat cerebral cortex), (G) tangential cell (blowfly visual lobe), (H) relay cell or interneuron (mouse dorsal thalamus).

6.4 3D Microscopic Stack of Cerebellar Cortex

6.4.1 Surrogate Stack: Implementation and Results

NeuroMorpho.org [7] is an open-source online inventory of digitally reconstructed neuron that are presented in a unified format. It contains contributions from dozens of laboratories worldwide and is continuously updated as new morphological reconstructions are collected, published, and shared. Even though in practice, due to imaging and image segmentation artifacts, reconstructions may not always be perfectly accurate representations of the true biological data, the digital neurons accessible from NeuroMorpho.org are peer-reviewed and verified by multiple sources, making them closer to "ground-truth" than manual segmentation by an individual. Figure 6.15 illustrates the various reconstructed cell types available at the repository.

The digitally reconstructed sample Purkinje cells we employed to construct our surrogate data stack were obtained by two-photon laser scanning microscopic imaging of Alexa Fluor 594 fluorescence (at 800 - 840 nm) [155]. Reconstructed Purkinje cells were first cantered on their somata and their main axons were aligned in three dimensions such that the white matter was down and the direction away from the lobule apex was to the right. A two-dimensional axon collateral density projection onto the sagittal plane was then generated. The apex-base direction was confirmed by examining the position of the Purkinje cell within the slice. The 3D collected image stack were then subject to local brightness level thresholding. After 3D skeletonization, manual cylindrical surface (symmetric axis being the 3D skeleton) fitting were carried out between nodes using user-interface of the TREES toolbox (see Figure 6.16) [156]. Though the reconstruction was submitted to NeuroMorpho.org in the format of Neurolucida, we were able to convert the morphology (see Figure 6.17) into cylindrical surface mesh representation using the TREES toolbox. We then further converted the neural surface to triangulated mesh representation using our own Matlab functions. The vertices of the triangulated mesh are then used by Matlab in-built function *patch()* to generate 3D surface fitted with polygons. An iterative discretization process then split the polygons into smaller

pieces until the longest edge reached below 0.5 voxel for the voxel beneath the vertices to be set to one. Thus, we were able to generate a volumetric estimation of the digitally reconstructed Purkinje cells in the format of Matlab volumetric matrix representation.

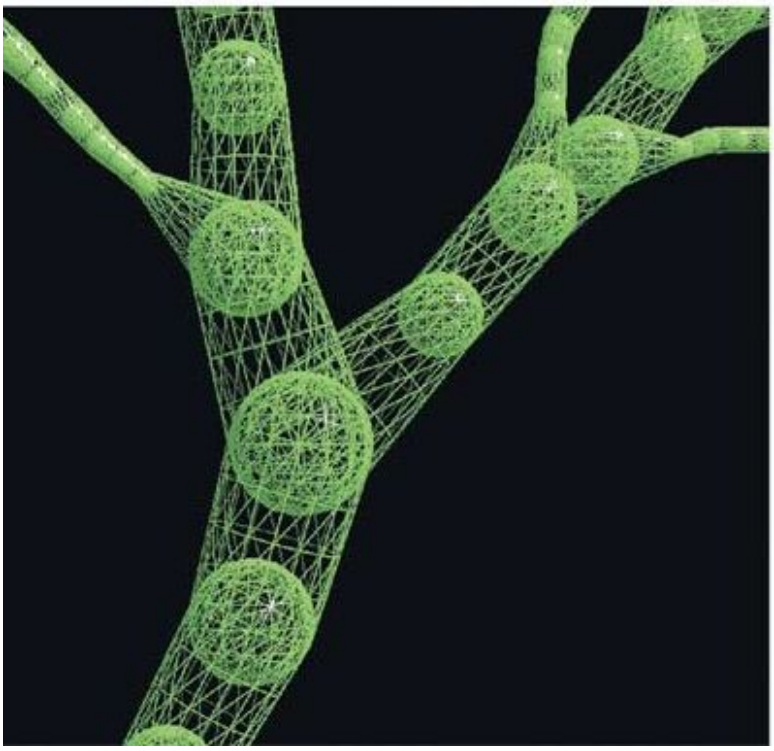


Figure 6.16: Cylindrical mesh-based surface fitting between nodes (represented by spherical structures) along dendrites.

To mimic the noise present at lab environment, we added a range of Gaussian white noise (Var between 0.01 to 0.5) to the "ground-truth" reconstructed Purkinje cells. Figure 6.18 is a montage view (along z-axis) of a sample surrogate stack where Gaussian white noise ($Var = 0.1$) was added on the reconstructed Purkinje neuron. The 180° rotated view of the maximum intensity projection of the this sample surrogate stack can be visualised in Figure 6.20.

We trained our "PP-MLP" system on a sample stack ($213 \times 149 \times 35$) with known variance, ie. $Var = 0.1$. This was then tested on a range of generated stacks with incrementing Gaussian white noise variance (Var from 0.01 to 0.5) to evaluate the noise-resistance of our proposed method. Figures 6.22 to 6.25 illustrate the segmentation results (in green) of our PP-MLP method when trained on a sample stack with $Var = 0.1$. We can see our method

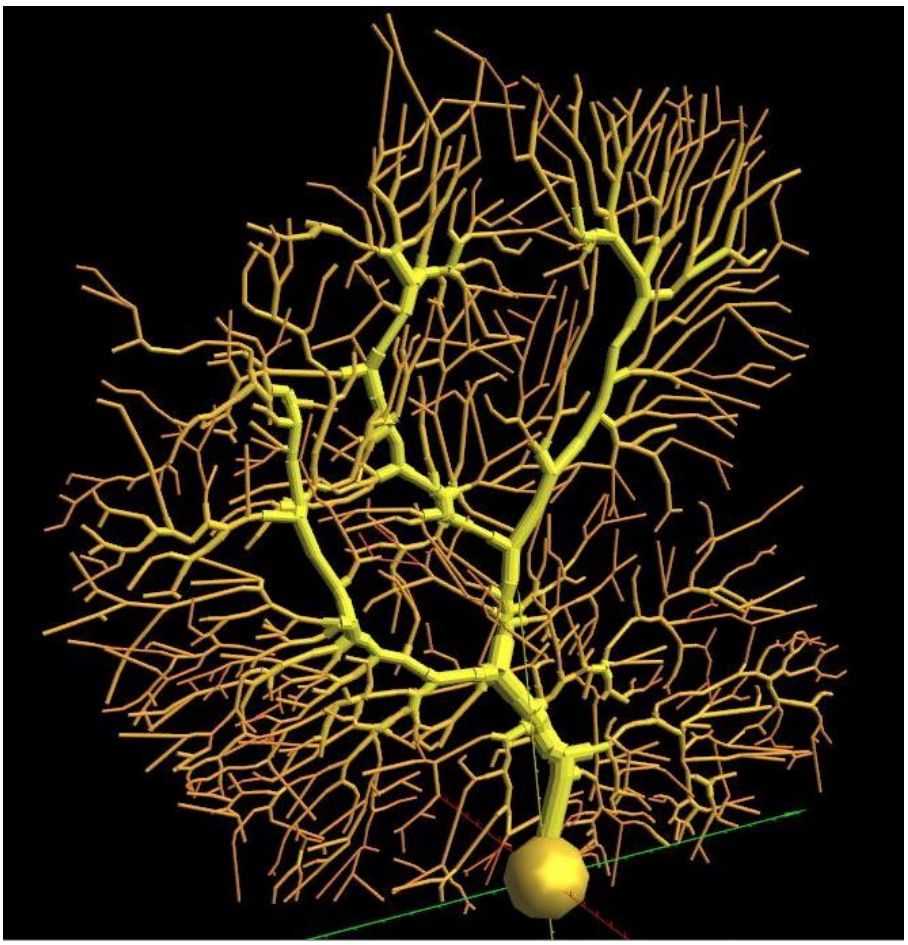


Figure 6.17: A reconstructed rat cerebellar Purkinje neuron morphology.

segmented well (note the coherence of the fine dendritic structures away from the Purkinje cell body in the segmented results) even up to 4 times the noise variance in training. When we applied the same MLP training on the same training stack but using intensity-only 3D neighbourhood patches, the segmentation result (in blue) yielded similar results when less noise were present($Var = 0.1, 0.2$). Though at higher noise level ($Var = 0.3, 0.4$), the segmented dendritic structures became perforated and broken (see Figure 6.25). Thus, we can conclude our method yielded more satisfactory segmentation results at the presence of higher noise interference, as orientation information becomes more crucial in accurate and correct structural differentiation.

We also applied Gaussian blurring operation (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) on the noise-added data stacks to complicate the segmentation problem, as observed in experimental imaging. Figure 6.19 is a montage view (along z-axis) of a sample surrogate

stack where Gaussian white noise ($Var = 0.1$) was first added on the reconstructed Purkinje neuron, which was then followed by Gaussian blurring. The 180° rotated view of the maximum intensity projection of this sample surrogate stack can be visualised in Figure 6.21.

Thus, we trained our "PP-MLP" system on a sample stack ($213 \times 149 \times 35$) with known variance, ie. $Var = 0.1$ and known convolution kernel size of $3 \times 3 \times 3$ and $SD = 7$. This was then tested on a range of generated stacks with incrementing Gaussian white noise variance (Var from 0.01 to 0.5) to again evaluate the resistance to noise and blurring of our proposed method. Figures 6.26 to 6.29 illustrate the segmentation results (in green) of our PP-MLP method when trained on a sample stack with $Var = 0.1$ and convolution kernel size of $3 \times 3 \times 3$. We can see our method segmented well (note the presence and the relative coherence of the fine dendritic structures away from the Purkinje cell body in the segmented results) even up to 4 times the noise variance in training. Again we applied the same MLP training on the same training stack but using intensity-only 3D neighbourhood patches, the segmentation result (in blue) yielded similar results when less noise were present($Var = 0.1, 0.2$). Though at higher noise level ($Var = 0.3, 0.4$), it was not able to segment fine dendritic structures away from cell body (see Figure 6.29). Thus, we can conclude our method yielded more satisfactory segmentation results at the presence of higher noise interference and substantial blurring, as orientation information becomes more crucial in accurate and correct structural differentiation than simply relying on the intensity profile in the 3D neighbourhood.

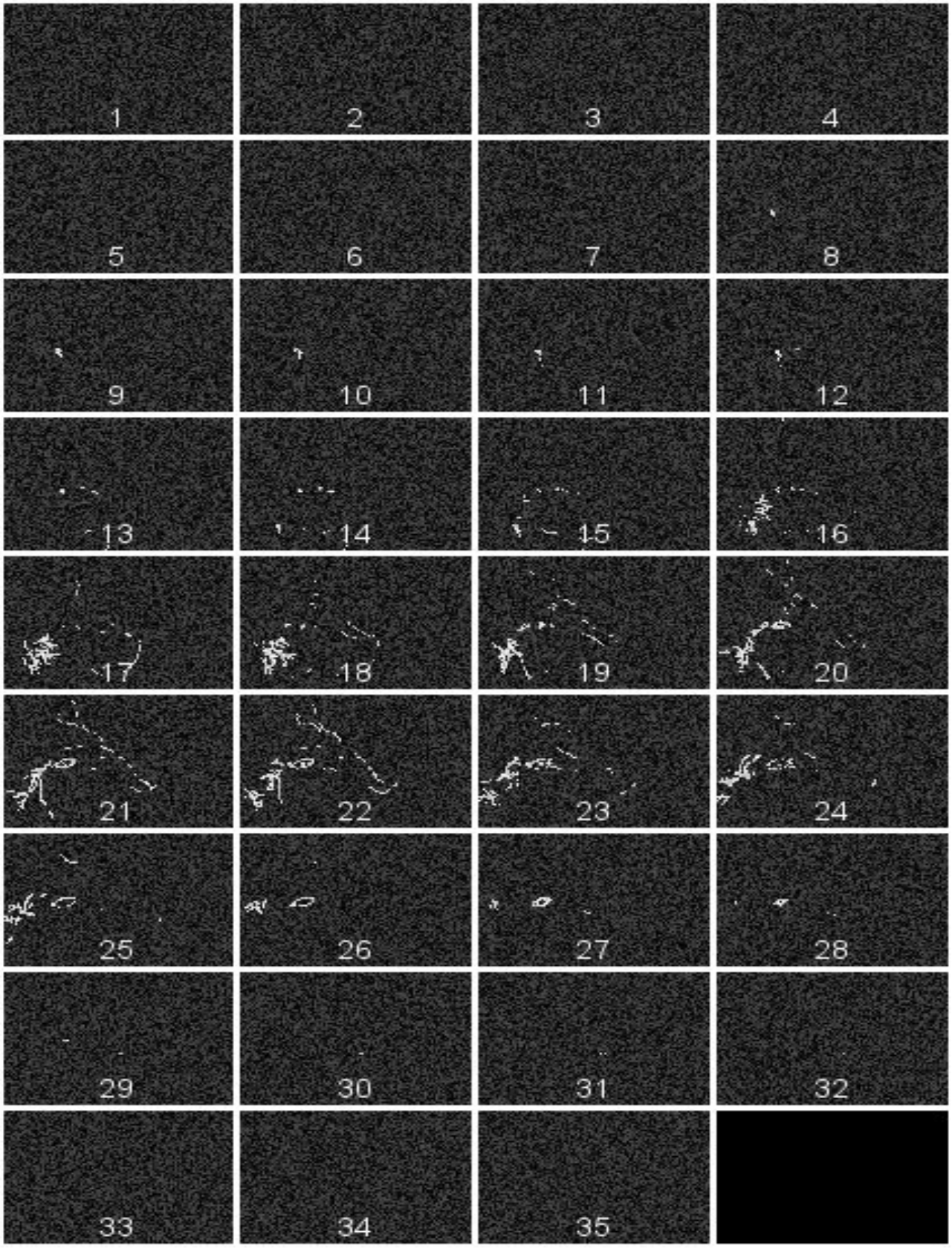


Figure 6.18: A montage view (along z-axis) of a sample surrogate stack ($213 \times 149 \times 35$). Reconstructed Purkinje neuron (stack) with Gaussian white noise ($Var = 0.1$) added.

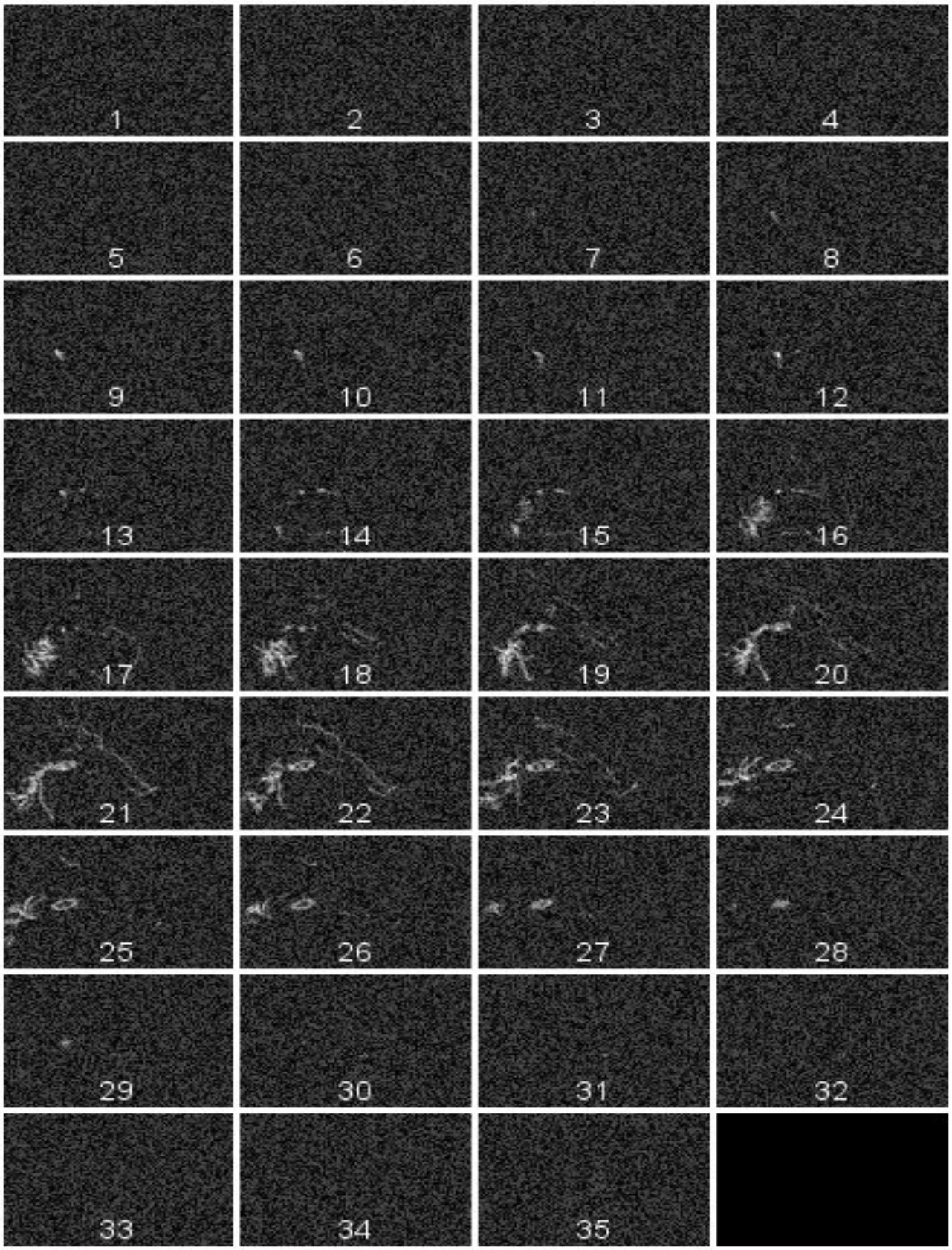


Figure 6.19: A montage view (along z-axis) of a sample surrogate stack ($213 \times 149 \times 35$). Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron (stack).

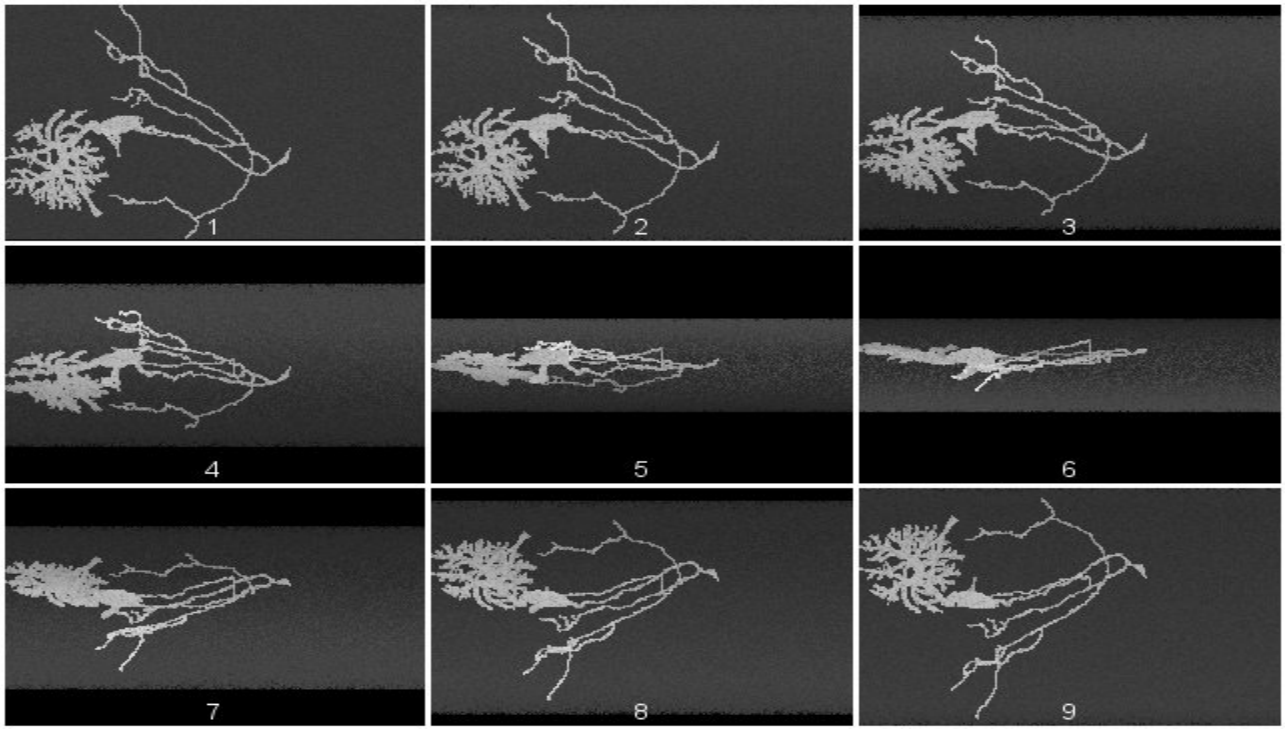


Figure 6.20: A 180° rotated view of the maximum intensity projection of the sample surrogate stack, reconstructed Purkinje neuron (stack) added with Gaussian white noise ($Var = 0.1$).

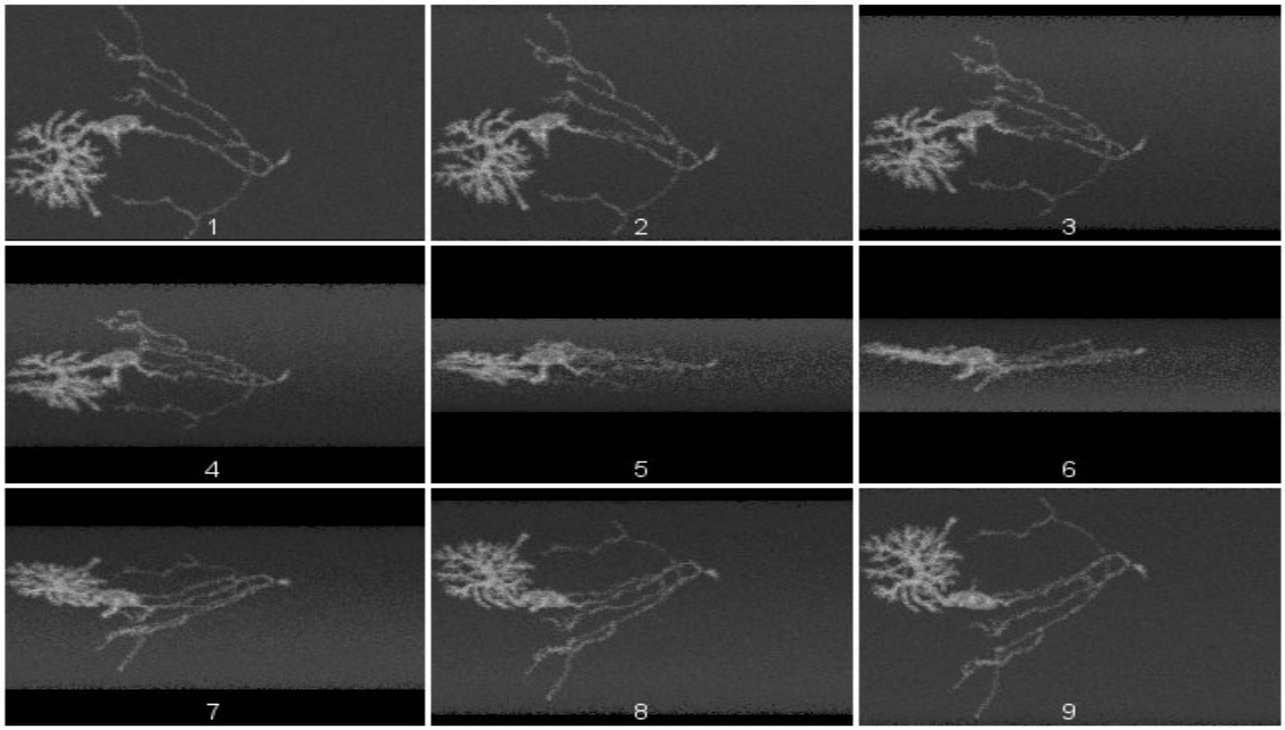


Figure 6.21: A 180° rotated view of the maximum intensity projection of the sample surrogate stack, Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron (stack).

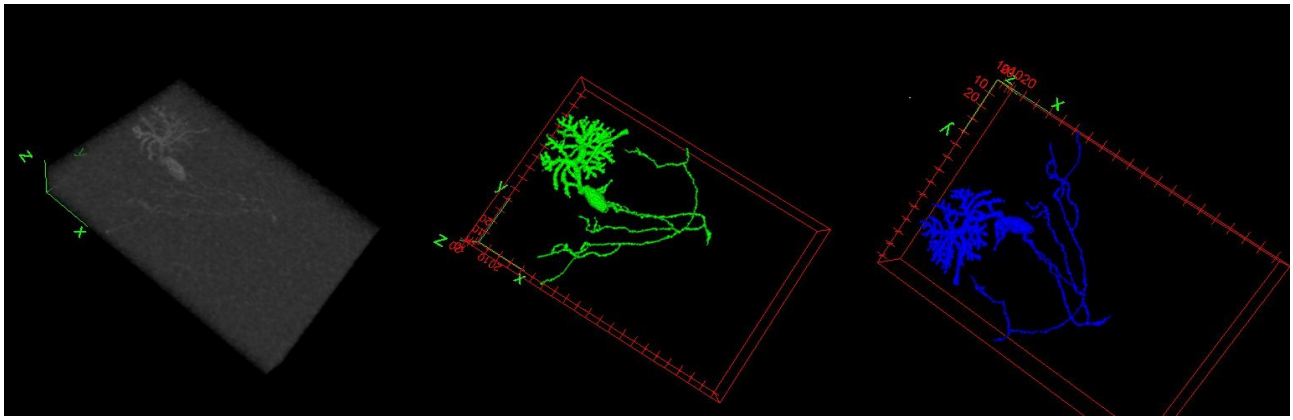


Figure 6.22: Left: A 3D view of the test stack where Gaussian white noise $Var = 0.1$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

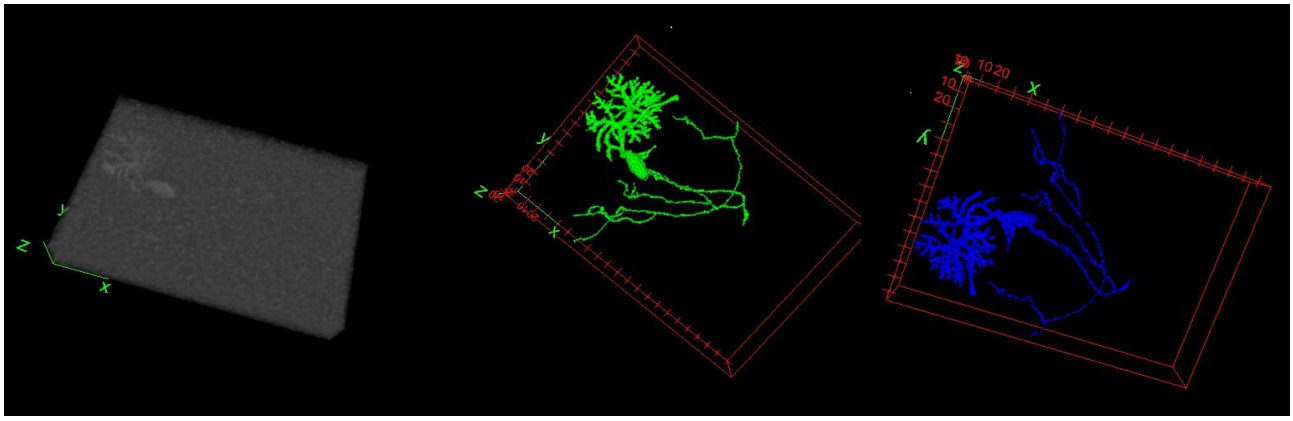


Figure 6.23: Left: A 3D view of the test stack where Gaussian white noise $Var = 0.2$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

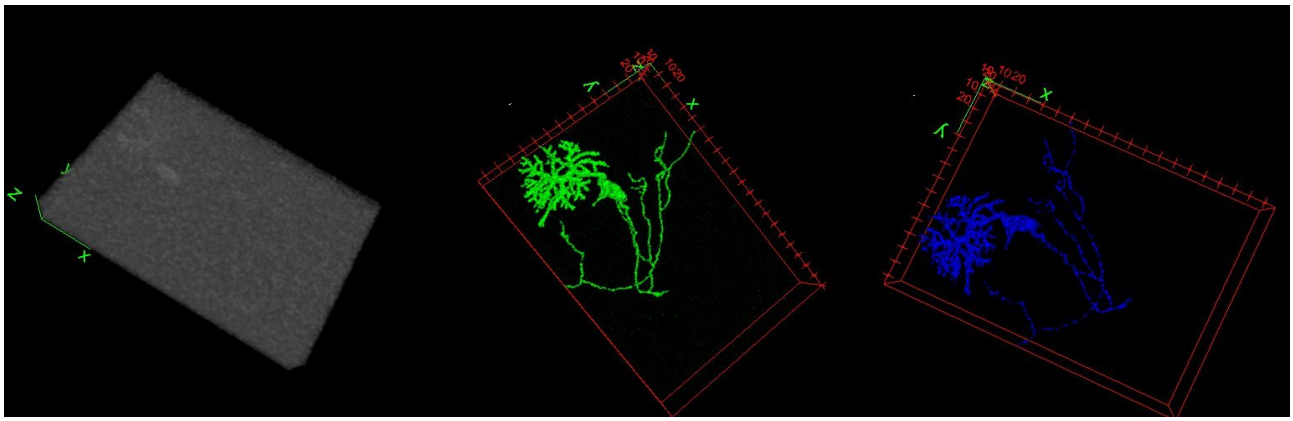


Figure 6.24: Left: A 3D view of the test stack where Gaussian white noise $Var = 0.3$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

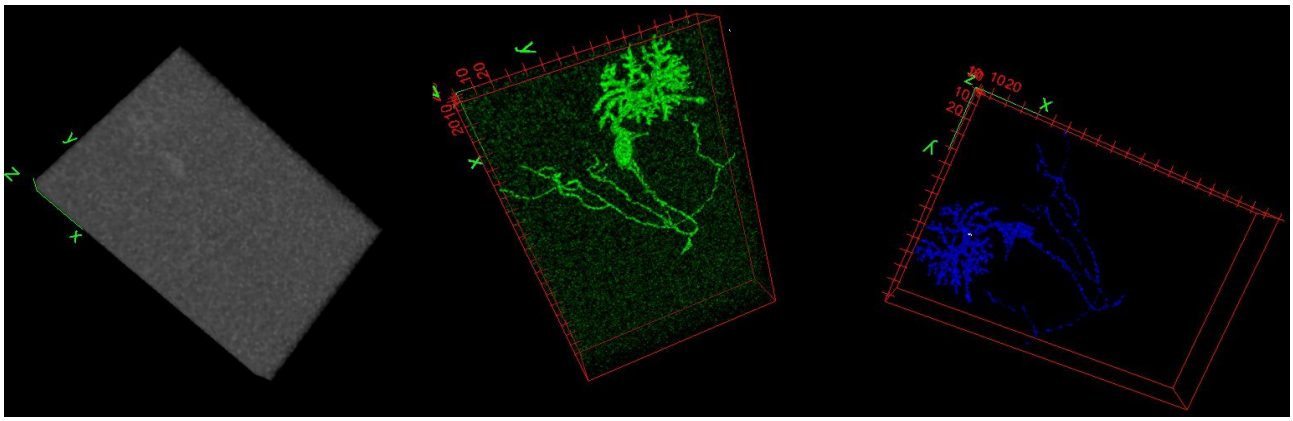


Figure 6.25: Left: A 3D view of the test stack where Gaussian white noise $Var = 0.4$; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

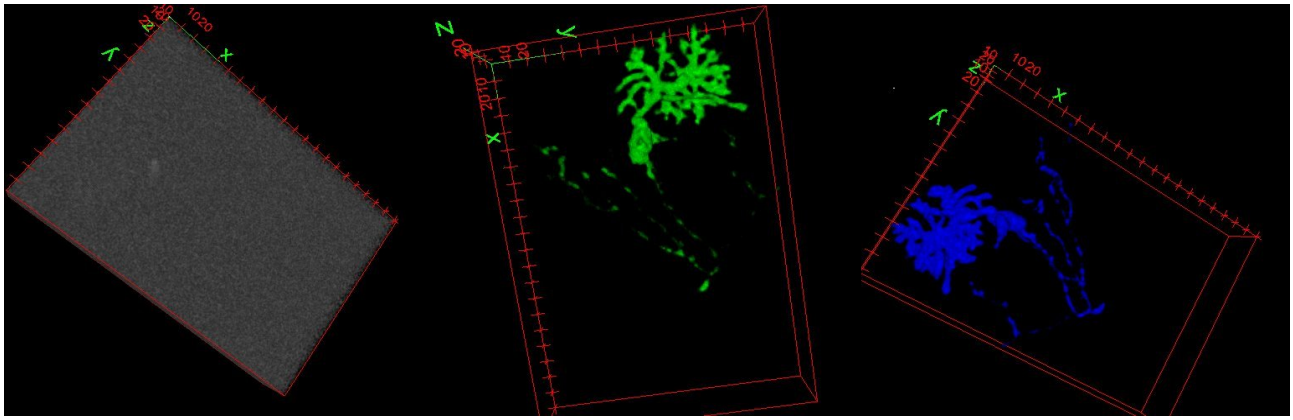


Figure 6.26: Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.1$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

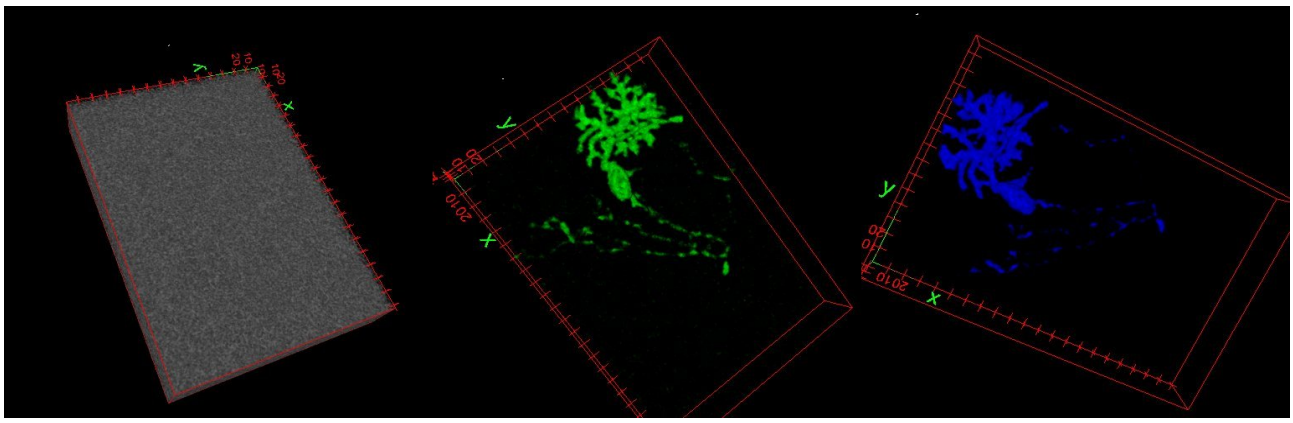


Figure 6.27: Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.2$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

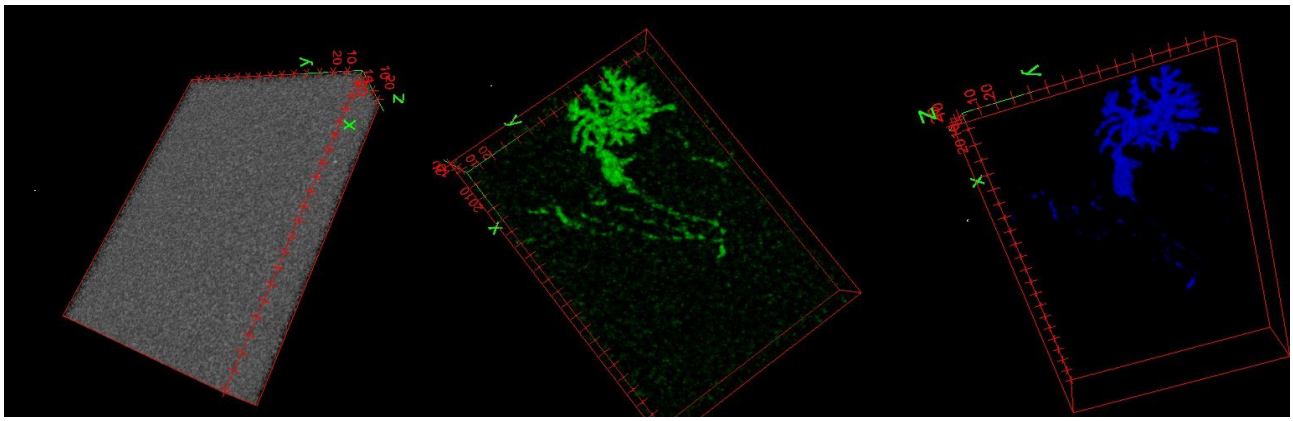


Figure 6.28: Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.3$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

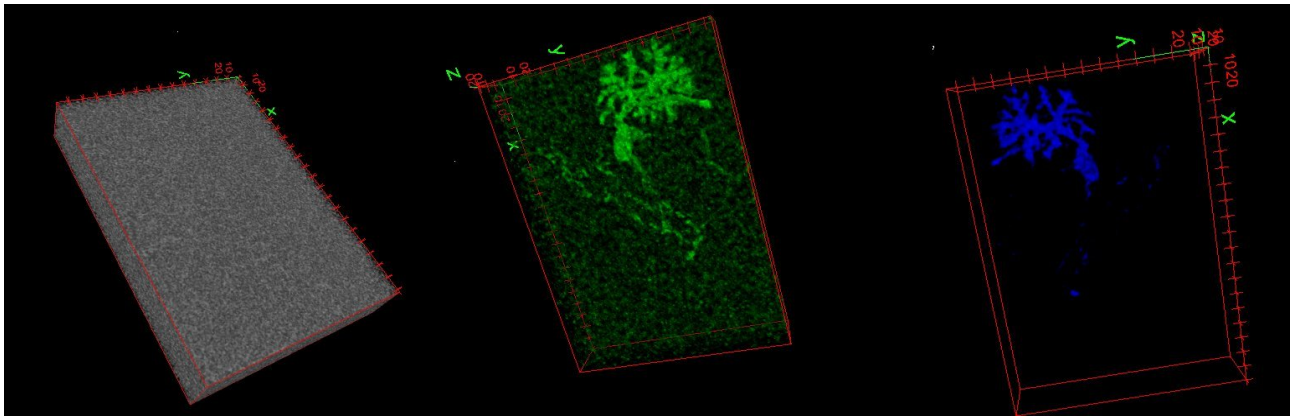


Figure 6.29: Left: A 3D view of the test stack where Gaussian blurring was carried out post Gaussian white noise ($Var = 0.4$) addition; Middle: The segmentation result of our PP-MLP method; Right: The segmentation result of MLP training using intensity-only 3D neighbourhood patch.

6.4.2 Experimental Data : Results

A sample surrogate stack ($213 \times 149 \times 35$) where Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) was carried out post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron was used to train our system where the input feature vector was a combination of intensity measure and our 5-D orientation feature in an 18-connected neighbourhood. The segmentation result of our "PP-MLP" method can be visualized in Figures 6.30 and 6.31, where the input experimental data stack (see Figure 1.15 and 1.16) contains a variety of cellular components. In comparison, the same MLP training when applied on the same training stack but using intensity-only 3D neighbourhood patches, yielded much poorer segmentation results (in blue) as only some cell bodies and some parallel fibres were retained (see Figures 6.32 and 6.33). This comparison illustrates that in lab environment, the presence of noise is highly distorting to complex decision surface when only using intensity information in a structural patch, whereas orientation information are far more effective and indicative in characteristic structural differentiation.

Here, we present another input two-photon microscopic experimental data stack (see Figure 6.34) with more sparse structures, as the fluorescent dye is highly selective to Purkinje cells with associated parallel fibres. Again, a sample surrogate stack ($213 \times 149 \times 35$) where Gaussian blurring (convolution kernel size of $3 \times 3 \times 3$, $SD = 7$) was carried out post Gaussian white noise ($Var = 0.1$) addition on the reconstructed Purkinje neuron was used to train our system where the input feature vector was a combination of intensity measure and our 5-D orientation feature in an 18-connected neighbourhood.

The segmentation result of our "PP-MLP" method can be visualized in Figures 6.35 and 6.36. We were able to successfully segment two Purkinje cells from the entire stack. In comparison, the same MLP training when applied on the same training stack but using intensity-only 3D neighbourhood patches, yielded less satisfactory segmentation results (in blue) as only cell bodies od larger scales were retained while dendrites and parallel fibres were mostly omitted (see Figure 6.37). Additionally, we also applied the same MLP training on the same training stack but using GST tensor-based 3D neighbourhood feature patches. Again, tensor approach

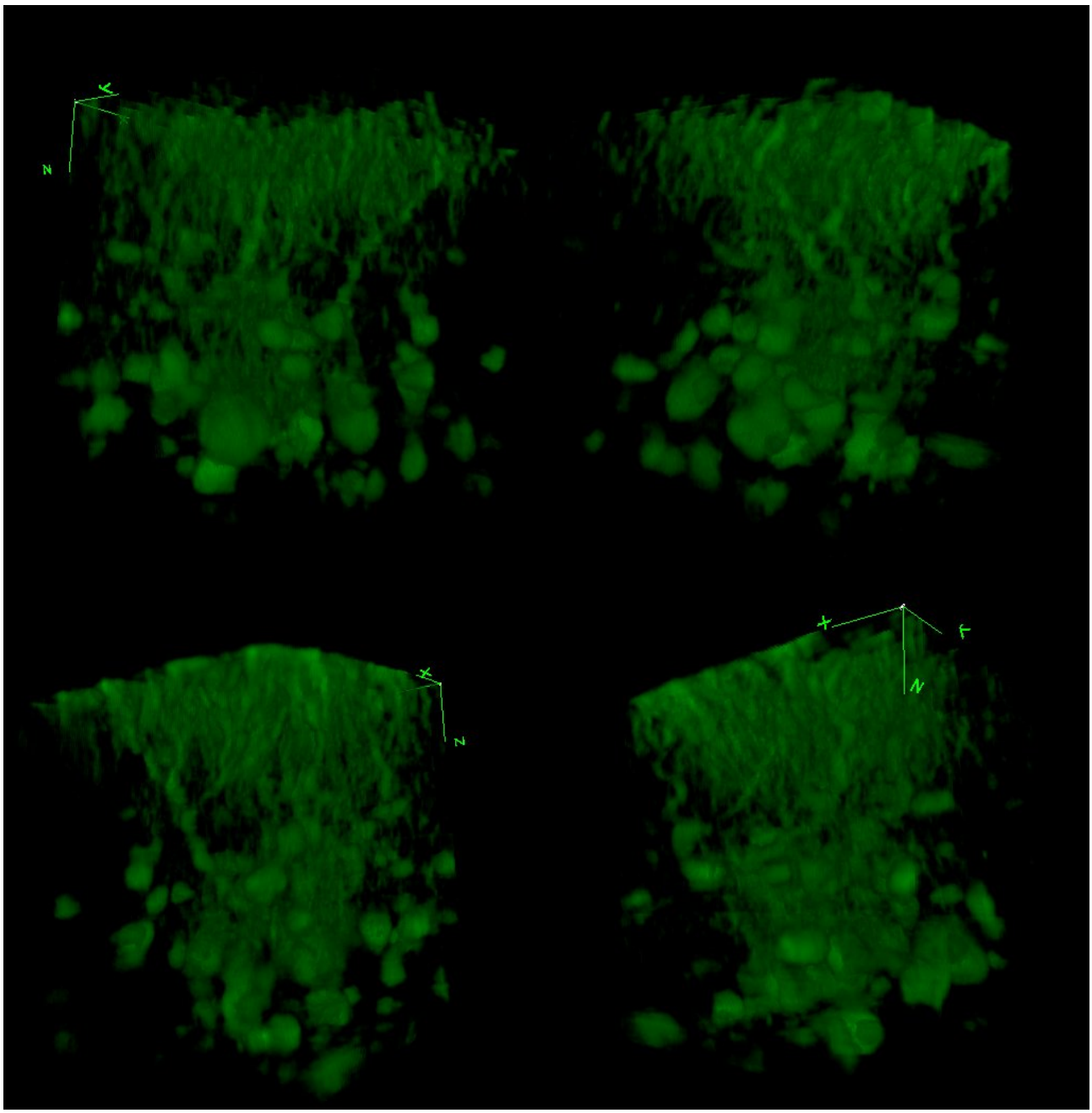


Figure 6.30: Four rotated 3D view of segmented neural cellular structures (in green) from the experimental rat cerebellar cortex microscopic tissue stack using our "PP-MLP" method.

yielded less satisfactory segmentation results (in aqua) as even less dendrites and parallel fibres were retained from classification (see Figure 6.38). This comparison illustrates that in lab environment, the presence of noise is highly distorting to complex decision surface when only using voxel intensity or eigenvalue information in a structural patch, whereas orientation information are far more effective and indicative in characteristic structural differentiation.

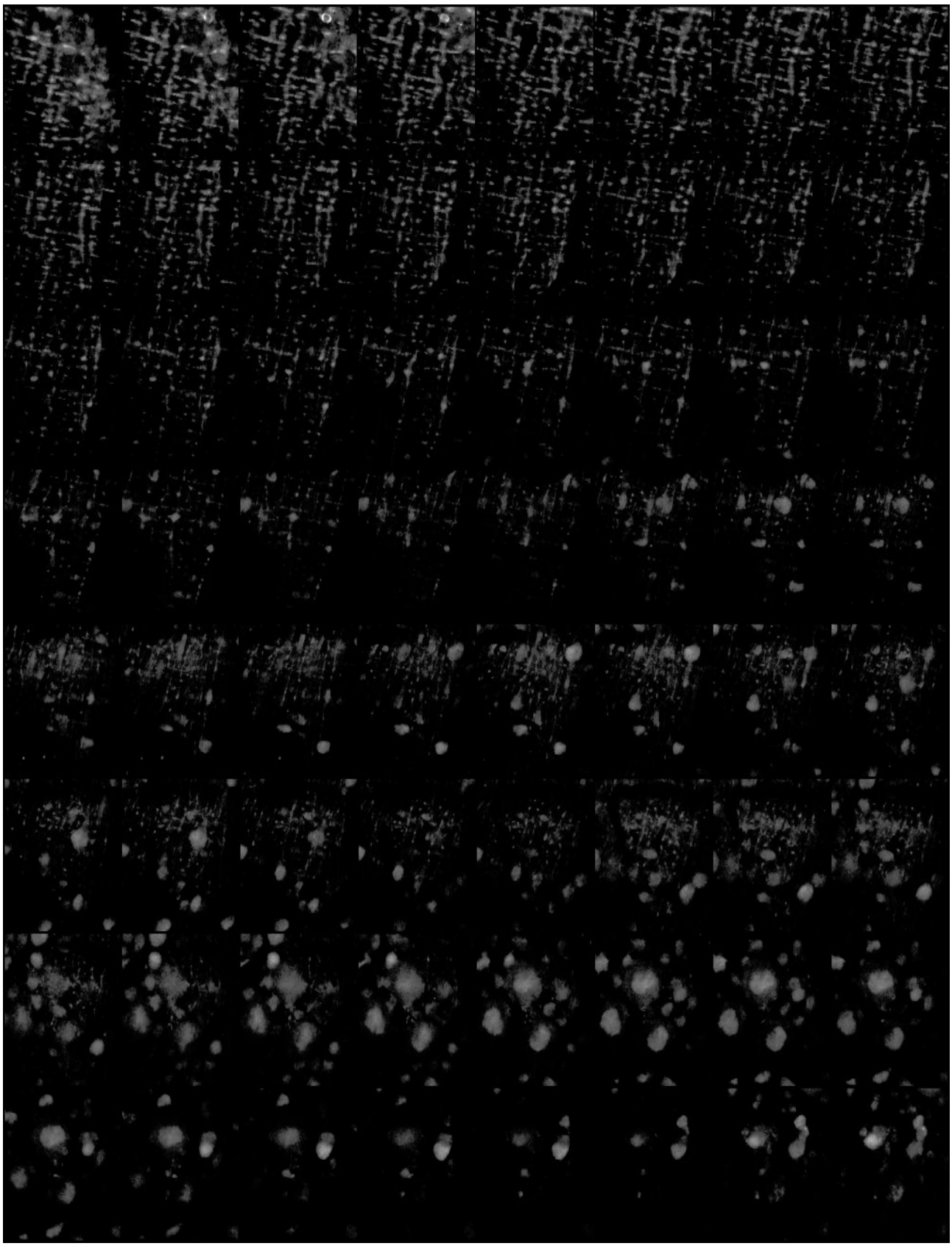


Figure 6.31: A montage view of the segmented neural cellular structures from the experimental rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method, only displaying every other slice along the Z axis.

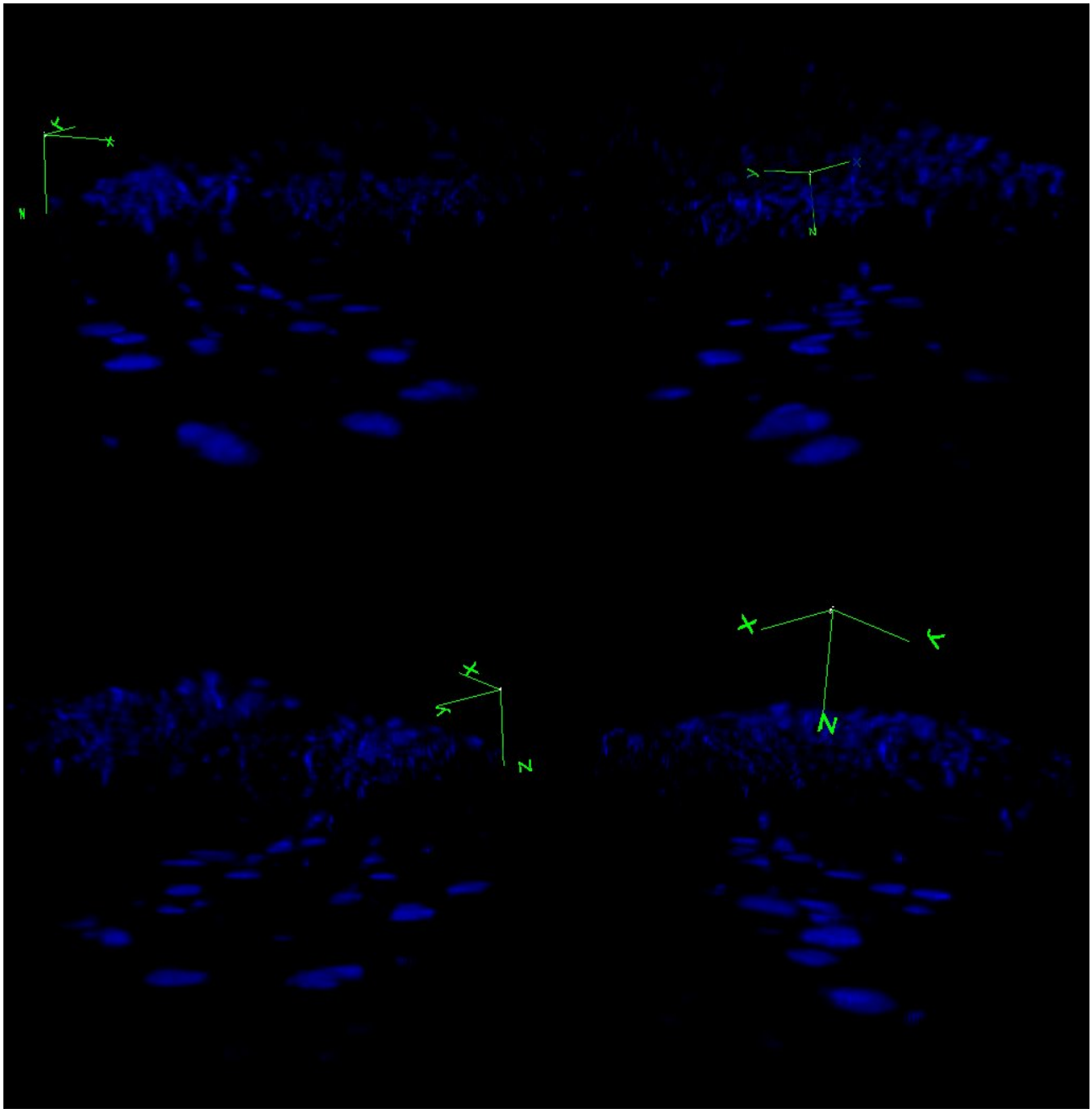


Figure 6.32: Four rotated 3D view of segmented neural cellular structures (in blue) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch.

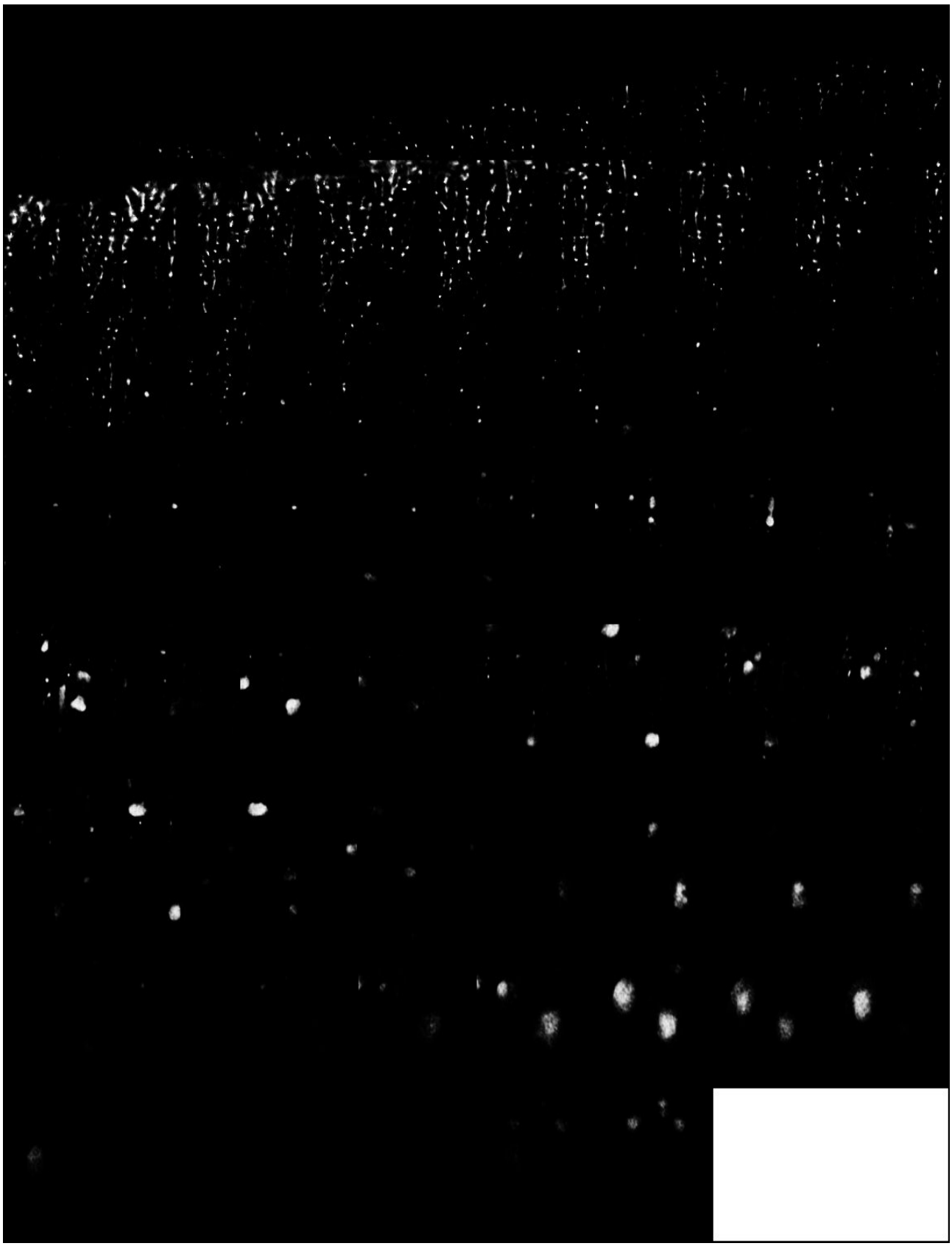


Figure 6.33: A montage view of the segmented neural cellular structures from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch, only displaying every other slice along the Z axis.

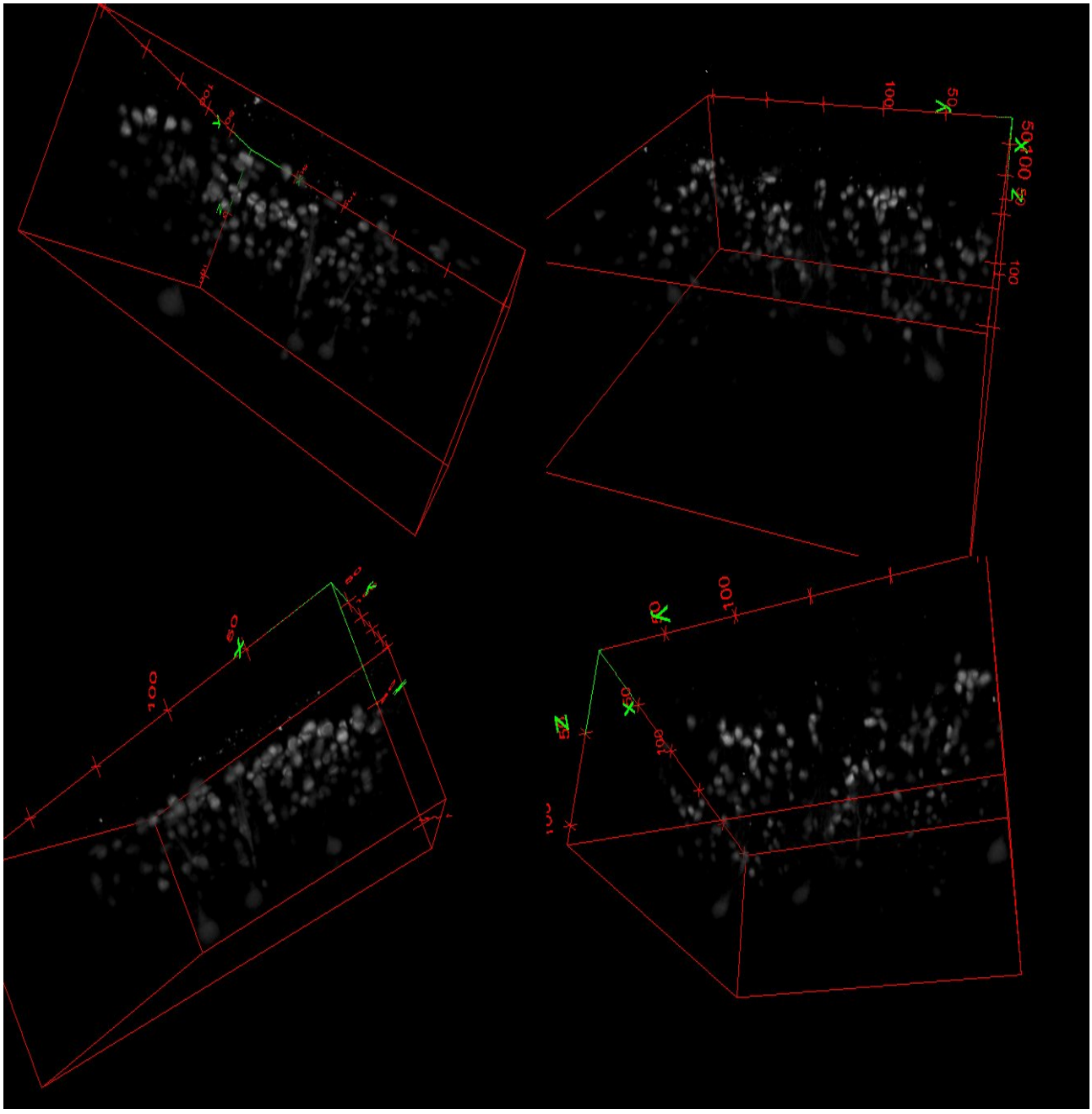


Figure 6.34: A 3D view of a sample rat cerebellar cortex microscopic tissue stack (512×512×223 at 8-bit).

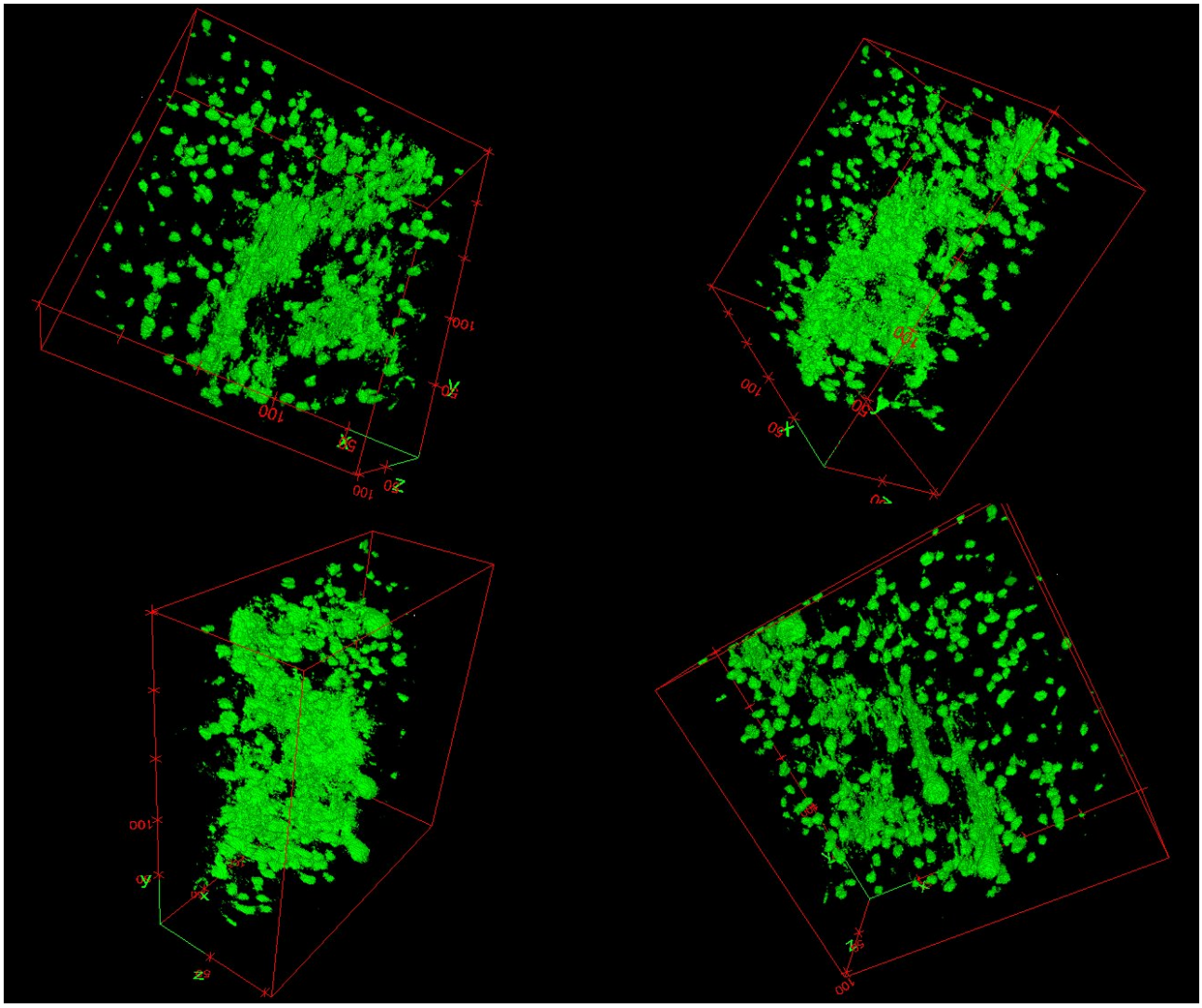


Figure 6.35: Four rotated 3D view of segmented neural cellular structures (in green) from the experimental rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method.

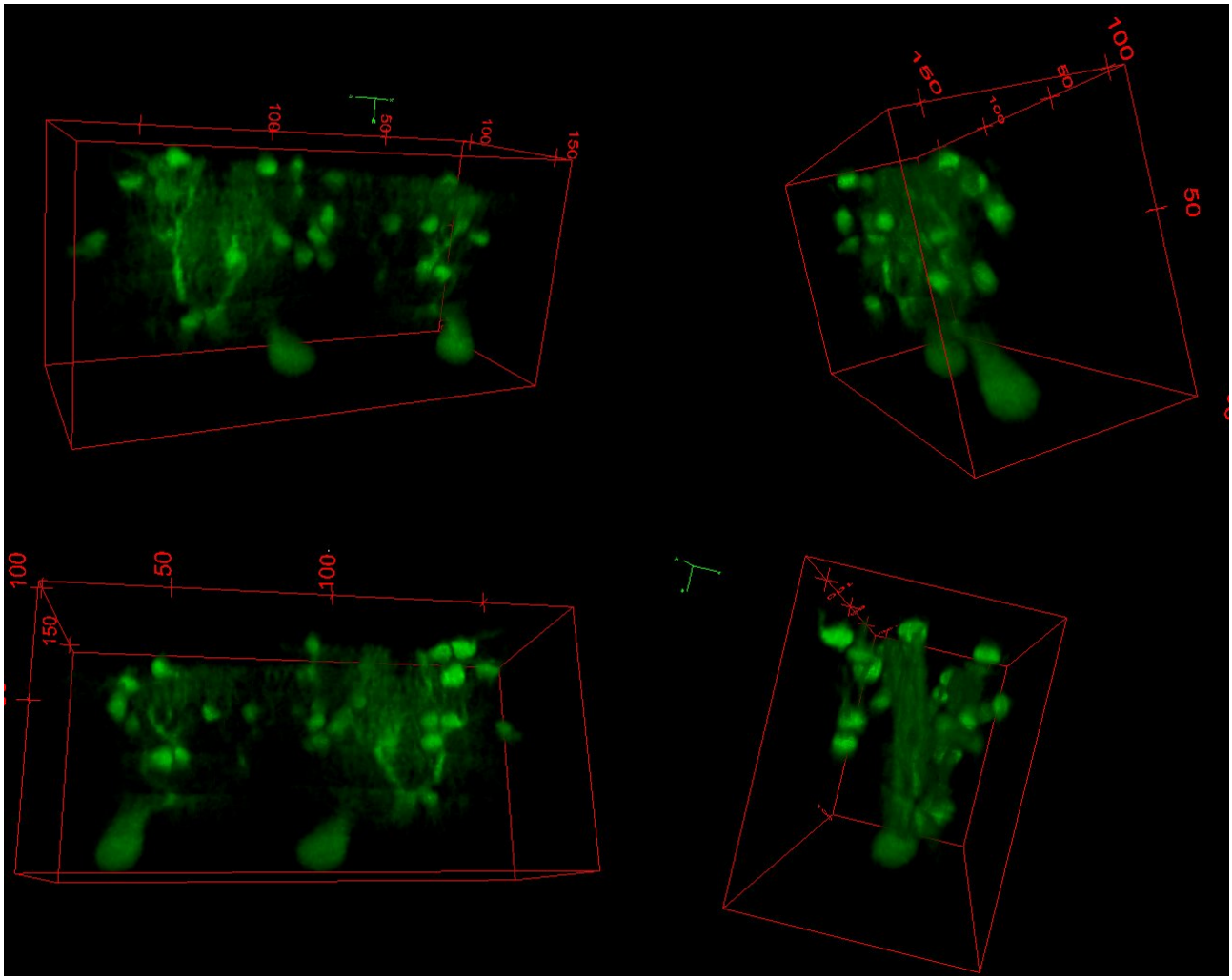


Figure 6.36: Four rotated 3D view of the segmented Pukinje cells (two in total) from rat cerebellar cortex microscopic tissue stack using our proposed "PP-MLP" method.

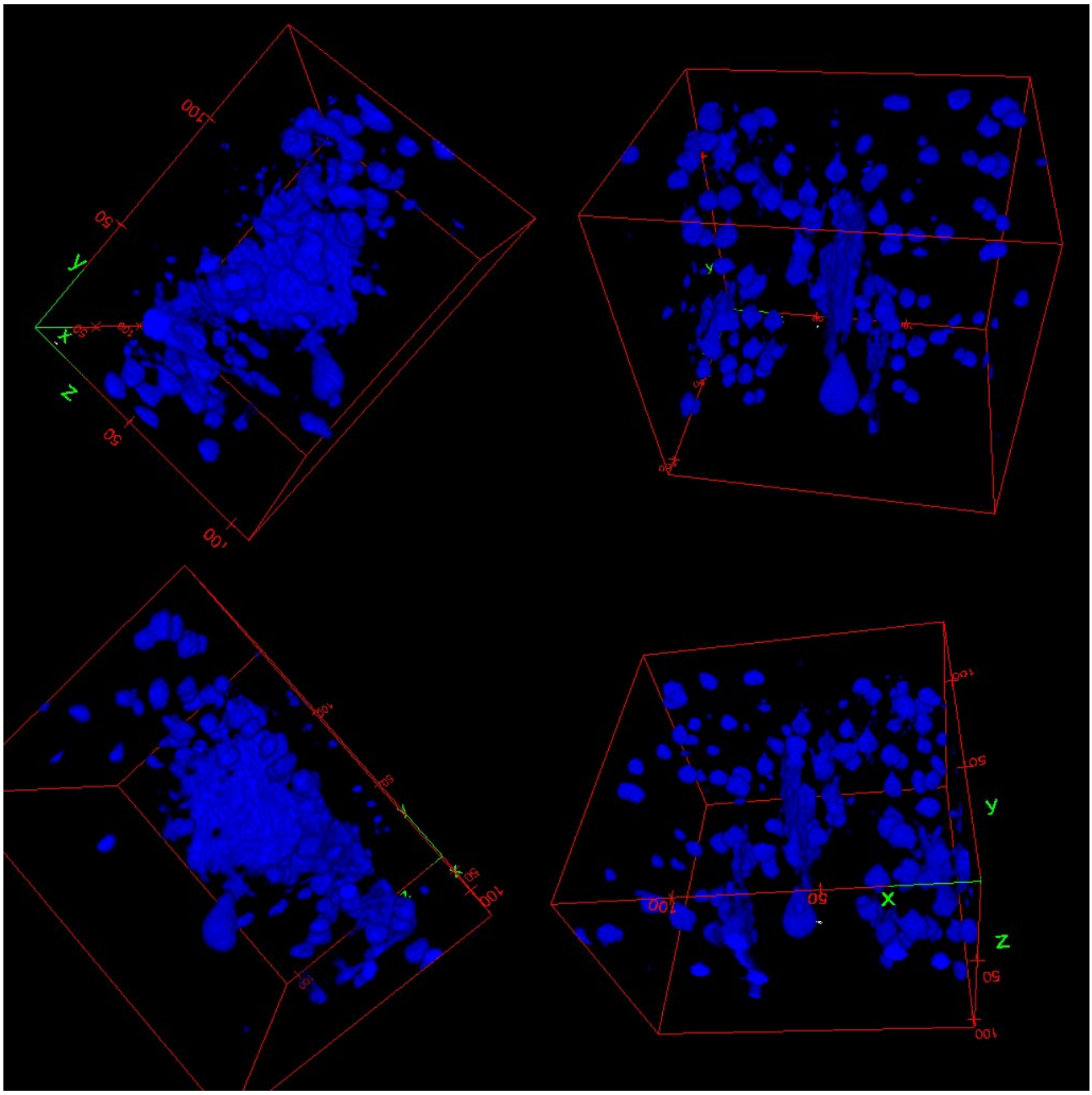


Figure 6.37: Four rotated 3D view of segmented neural cellular structures (in blue) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on intensity-only 3D neighbourhood patch.

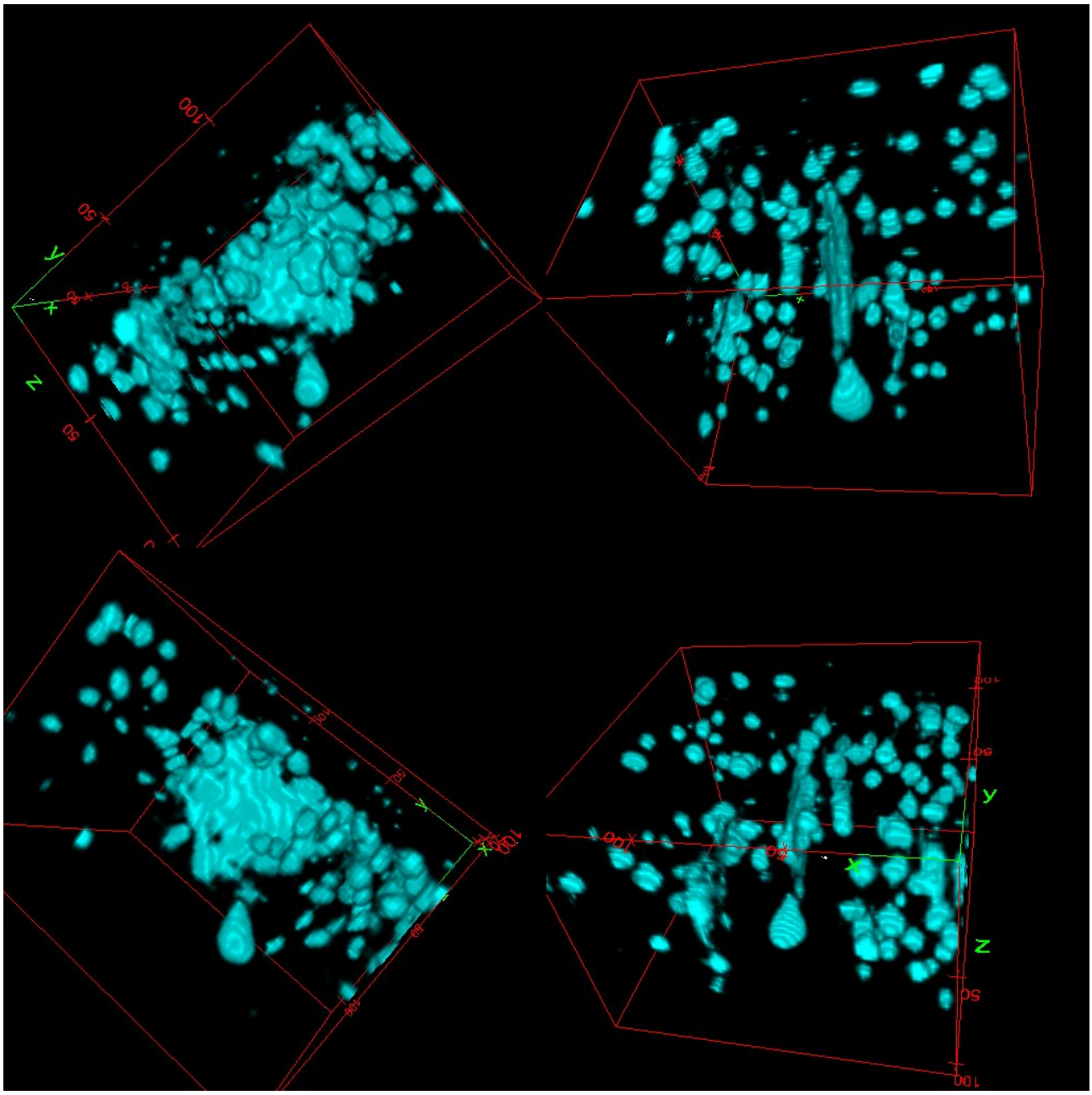


Figure 6.38: Four rotated 3D view of segmented neural cellular structures (in aqua) from the experimental rat cerebellar cortex microscopic tissue stack using MLP training on GST-based 3D neighbourhood feature patch.

6.5 Discussion and Conclusion

In this chapter, we first present the compute Receiver Operating Characteristic (ROC) curves on existing public retinal fundus image databases ("DRIVE" & "STARE") and conclude that our "Pixel-level Patch-based Segmentation with Multilayer Perceptron" (PP-MLP) method is highly favourable when compared with up to nine other other benchmark methods. We then introduce the "Centreline Error" that utilizes our proposed "link while extract" piece-wise centreline extraction and tracking method (detailed in Chapter 4) to assess microvasculature registration in fluorescein angiography. This is error metric is subsequently applied to benchmark our "Joint Global RANSAC" registration method against a Global Affine registration approach on 2D+t sequential retinal angiograms. We conclude our model has a more stable performance with less (sub)-pixel misalignment in all registered frames. We then, start by introducing our method to construct surrogate neural data stack in Matlab using existing public database ("NeuroMorpho"). This is followed by evaluating the benefit of our proposed 114-dimensional feature descriptor (detailed in Chapter 5) against intensity-only or GST-based 3D feature descriptors. We conclude that in the presence of noise, for both surrogate and experimental data stacks, our PP-MLP approach is far more effective for fine and elongated structural segmentation.

Chapter 7

Discussion and Conclusion

In Chapter 2, we present the problem of 2D retinal funduscopy segmentation on two public available databases, DRIVE and STARE. We first present a literature review on existing methods for vascular segmentation in retinal funduscopy. This is followed by the introduction of our proposed framework for structural segmentation with the explanation and justification of the feature sets that we use, as well as the the justification of phase-invariant patch representation to encode spatial structures. When combined with the appropriate classifier, the system can perform satisfactory vascular segmentation of the global vessel-bed yet with a pixel-level local approach. The principles behind construction of phase-invariant orientation field of the local spatial neighbourhood and the combination of intensity-based measures and local orientation feature maps to form local "feature patches" would unify through the thesis, but here we first introduce supervised machine learning (MLP or SVM) as an option for "soft" classification.

One key observation is that the performance of patches of features is more robust than a simple combination of selected features. This is not reflected in standard feature selection tools, as one cannot generalize the contribution of each individual element of a patch without prior information. In our experience, a powerful machine learning method, such as MLP or SVM, tends to interpret all the elements in the patch to reach a joint classification decision. Patches of non-linear feature spaces offer the potential for much better performance than

single features, even if such features might be produced from operators of wide spatial support.

With regards to MLP, at first, we have concerns as optimization methods such as steepest descent and conjugate gradient are susceptible to finding local minima if the search starts near a local minimum, hence the global minimum might not be guaranteed; while SVM guarantees to deliver unique solution (the global minima), it compromises on speed and computation load for training. However, our experimental results suggest that MLP-classifiers are more successful in segmenting global vasculature in funduscopy. We suspect the reasons are two-fold. First, the specific layout of MLP-classifier we apply was initially configured for handwriting recognition, as hand-written alphabets to some extent also exhibit elongation and intermittent gaps at times, the specific configuration may be more fine-tuned for our biomedical applications. Secondly, the high SNR and varying illumination, often observed in retinal imaging, may result in disadvantage in applying SVM in this particular application. However, this is an application-specific observation as we have successfully applied SVM for cell membrane segmentation in the past [68].

In Chapter 3, after extensive literature review on existing methods for retinal registration, we address a novel co-registration procedure ("Joint Global RANSAC") to align and utilize the temporal factor in 2D+t sequential fluorescein retinal angiograms. We present an explanation and justification of image processing techniques we employed in our 6-step **pairwise** registration that incorporates projective RANSAC with quadratic homography transforms. This is then followed by our **joint** hierarchical compound of transformation matrices that eventually allow us to align fluorescein angiogram sequences taken before and after operation with spatio- and temporal variation on top of the existing geometric distortion, radiometric degradation, and additive noise corruption between any two frames. We then explain in depth the feature sets that we use, and justify the use of a phase-invariant patch representation to capture spatial structure. When combined with the appropriate classifier, the system can perform satisfactory vascular segmentation of the global vessel-bed yet with a pixel-level local approach.

The purpose of 2D+t sequential retinal image registration is to spatio-temporally align set(s) of retinal images inter- and/or intra-sequentially to clinical assess disease-related ocular circulation. Even in the presence of substantial coordinate and field-of-view shift between pre- and post- operative sequences, our registration algorithm is able to perform a satisfactory results to map the coordinates without resorting to state-of-the-art local-to-global bootstrapping methods that, though allow higher degree of transformation require higher computation power.

Another point to note is that, our "Joint Global RANSAC" method manages to remove manual bias in its automated reference frame selection by choosing the frame where fluorescein concentration peaks within field-stop which is the imperative approach adopted from clinical practices.

In Chapter 4, we then take a non-training approach (our proposed "link while extract" piece-wise segmentation method) that searches in both the spatial neighbourhood of each pixel and the filter output across varying scales to locate and link microvascular centrelines to (sub-) pixel accuracy. In essence, our approach relying on phase information extracted after steering complex wavelets according to the phase-invariant orientation field to obtain a soft classification of the centreline (sub-) pixel locations. This is an alternative to our learning-based classification in Chapter 2. Each sub-steps in our "link while extract" piece-wise segmentation approach can be computed simultaneously and combined in one final step, and the direction of vessel centreline pixel extraction can be bi-directional (the nature of orientation description does not not differentiate directional vectors with 180 degrees angular difference in 2D). And more importantly, the entire algorithm is completely automated requiring no manual "seeding" of the start and end (vascular centreline) pixels. Thus we conclude that our approach is computational efficient due to the temporally independent and spatially correlational nature of the measures.

One potential application of our centreline extraction method is that when coupled with our proposed spatio-temporal registration framework, it allows comparison and non-invasive monitoring of fine-resolution microvasculature from fluorescein angiography. It provides the

potential for detecting temporal changes in the circulation (possibly caused by (micro)embolism) in real-time. This allows early preventative measures to be taken to reduce aggravated blood-clotting, thus improving the post-operative recovery of the patients.

In Chapter 5, we present our pixel-level patch-based segmentation with MLP training applied in 3D neural segmentation. This is a 3D extension of our 2D approach in Chapter 2, yet the complex nature of orientation representation in higher-dimensions requires us to implement a more complex 3D to 5D orientation mapping in order to preserve angular information and line curvature in a local neighbourhood. Our feature descriptor is then constructed in an 19-connected neighbourhood that combines the intensity measure with the 5D Knutsson-mapped orientation representation. Our 114-d feature space is then used for classification on a MLP Leaning Machine to differentiate among the complex and inter-connected network of neural structures. Our descriptor, when compared with intensity-only or GST/tensor approach, is able to perform better in a multi-directional complex neighbourhood as our 5D orientation representation offers more accurate encoding of the structural orientations.

In a local neighbourhood enriched with fine neurites, interneurons and parallel fibres, the neighbourhood orientation is often multi-directional. The presence of more than one dominant orientation and the discontinuous representation of eigenvector orientations in GST often result in difficulties applying structure tensor approach for neural segmentation. Moreover, intensity alone or GST-based eigenvalues may not encode sufficient information to differentiate among cellular structures, such as Purkinje cells, neural dendrites and interneurons, all display certain elongation along their medial axes, yet each class has a characteristic shape captured by an orientation field that distinguishes it from other structures. As expected, even though all three input feature space (27-d intensity-only, 81-d eigenvalue from GST/tensor, 114-d intensity measure with Knutsson orientation mapping) are successful in segmenting prominent blood vessels in the microscopic stack, our 114-d feature vector yields much more successful results in segmenting Purkinje cell bodies and associated dendrites that are much smaller in scale and more prone to the presence of noise.

In Chapter 6, we summarise performance evaluations and/or validation of each of the techniques presented in this thesis. we present the compute ROC curves on existing public retinal fundus image databases ("DRIVE" & "STARE") with that our PP-MLP method achieved $AUC = 0.9779$ and $Acc = 0.9270$ for DRIVE test images and $AUC = 0.9605$ and $Acc = 0.9490$ for STARE images. Our results are highly favourable when compared with up to nine other other binary and non-binary benchmark methods. Furthermore, we take precaution in preventing our model from over-training by varying patch size to illustrate the tight clustering of ROC curves with varying patch sizes. We notice some state-of-the-art approaches have the tendency to "under-segment" the likely vessel pixels in the pursuit of a steeper ROC curve (high true positive rate and low false positive rate, however this would in turn results in lower AUC and Acc values, not present in our results.

Also in Chapter 6, we introduce the concept of our "Centreline Error" that utilizes our proposed "link while extract" piece-wise centreline extraction and tracking method (Chapter 4) to assess microvasculature registration in fluorescein angiography. We believe that the microvascular centrelines are more structurally stable and more patient specific, hence a more objective representation of the "ground-truth" microvasculature for fluorescein angiography. However, the exact estimate for the centreline location of the vessels requires subpixel resolution, high accuracy and intensive manual labour, and thus is not adopted in manual ground-truth segmentation. Our automated ground-truth, on the other hand, is constructed on the temporal average of the registered sequence(s) as a more perceptually accurate representation of the retinal microvasculature and contains a positive weighting factor that has the flexibility to characterize neighbourhood vessel tortuosity in the presence of retinopathy. We believe our metric is a fairer assessment of registration quality as the clinically interesting fine-scale microvasculatures contribute more strongly to the error metric than a spatially-averaged global measure. This is error metric is subsequently applied to benchmark our "Joint Global RANSAC" registration method against a Global Affine registration approach on 2D+t sequential retinal angiograms. We conclude our model has a more stable performance at 0.11 mean (distance) pixel misalignment and a much tighter range of 0.29 pixel in all registered frames, compare to on average 3.2 pixel misalignments and

a much larger range of nearly 13 pixels in Global Affine registration approach.

Finally in Chapter 6, we introduce our method to construct surrogate neural data stack in Matlab using existing public database ("NeuroMorpho"). This is followed by evaluating the benefit of our proposed 114-dimensional feature descriptor (detailed in Chapter 5) against intensity-only or GST-based 3D feature descriptors. We conclude that in the presence of noise, for both surrogate and experimental data stacks, our PP-MLP approach is far more effective for fine and elongated structural segmentation. This is because in lab environment, the presence of noise is highly distorting to complex decision surface when only using voxel intensity or eigenvalue information in a structural patch, whereas orientation information are far more effective and indicative in characteristic structural differentiation.

To conclude this thesis, we generalize our pixel-level segmentation scheme as consists of three essential components: i) constructing a phase-invariant orientation field of the local spatial neighbourhood; ii) combining local feature maps with intensity-based measures in a structural patch context; iii) using a complex supervised learning process to interpret the combination of all the elements in the patch in order to reach a classification decision. We have demonstrated in this thesis the flexibility and transferability of our segmentation framework from retinal blood vessels in 2D to microvascular centreline in 2D+t to microscopic neural structures in 3D.

Future Work From the aspect of our proposed pixel-level segmentation scheme, we aim to further investigate in feature selection methods that explicitly group fields of features, as well as various non-linear methods to encode such fields in a patch-based approach.

From the aspect of clinical applications of our PP-MLP retinal vasculature segmentation method and our "link-while-extract" piece-wise retinal microvascular centreline extraction and tracking method, future development will include more extensive validation of the system on larger amounts of data and real-time performance evaluation. We are optimistic in its potential of incorporation in existing medical softwares for real-time or follow-up retinal vascular tree assessment and diagnostics.

From the aspect of neural structural segmentation, our approach for surrogate data stack construction is efficient and open-source, which has potential to benchmark existing state-of-the-art segmentation approaches in a relative less data-biased fashion. Our 3D PP-MLP neural segmentation method would also benefit from more extensive validation of the system, training on more neural cell types, and implementation of neurite centreline extraction and tracking based on our 2D approach.

Bibliography

[1] I. Topcon Medical Systems, "http://www.topconmedical.com/."

[2] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501–509, 2004.

[3] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[4] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. Garland Science, 5th ed., 2007.

[5] M. Carter and J. C. Shieh, *Guide to Research Techniques in Neuroscience*. Academic Press, 1st ed., 2010.

[6] M. R. Sullivan, A. Nimmerjahn, D. V. Sarkisov, F. Helmchen, and S. S.-H. Wang, "In vivo calcium imaging of circuit activity in cerebellar cortex," *Journal of Neurophysiology*, vol. 94, no. 2, pp. 1636–1644, 2005.

[7] G. M. University, "Neuromorpho.org," 2006-2012.

[8] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. PWS Publishing, 2nd ed., 1999.

[9] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2nd ed., 2010.

[10] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[11] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *SPIE Medical Imaging*, vol. 5370, pp. 648–656, 2004.

[12] T. Acharya and A. Ray, *Image Processing: Principles and Applications*. Wiley-Interscience, 2005.

[13] H. Knutsson, "Producing a continuous and distance preserving 5-d vector representation of 3-d orientation," in *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pp. 175–182, 1985.

[14] H. Knutsson, "Representing local structure using tensors," in *the Sixth Scandinavian Conference in Image Analysis*, pp. 244–251, 1989.

[15] R. N. Weinreb and A. Harris, eds., *Ocular Blood Flow in Glaucoma*, ch. Clinical Measurement of Ocular Blood Flow, pp. 19–58. Kugler Publications, Amsterdam, The Netherlands, 2009.

[16] E. Rechtman, A. Harris, R. Kumar, L. B. Cantor, S. Ventrapragada, M. Desai, S. Friedman, L. Kagemann, and H. J. Garzozi, "An update on retinal circulation assessment technologies," *Current Eye Research*, vol. 27, no. 6, pp. 329–343, 2003.

[17] A. Harris, L. Kagemann, and G. A. Cioffi, "Assessment of human ocular hemodynamics," *Survey of Ophthalmology*, vol. 42, no. 6, pp. 509–533, 1998.

[18] R. Bernardes, P. Serranho, and C. Lobo, "Digital ocular fundus imaging: A review.," *Ophthalmologica*, vol. 226, pp. 161–181, 2011.

[19] M. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169 –208, 2010.

[20] A. D. Hughes, "The clinical assessment of retinal microvascular structure and therapeutic implications," *Current Treatment Options in Cardiovascular Medicine*, vol. 9, pp. 236–241, 2007.

[21] J. F. Arevalo, *Retinal Angiography and Optical Coherence Tomography*. Springer, 2009.

[22] S. T. Venkataraman, J. G. Flanagan, and C. Hudson, "Vascular reactivity of optic nerve head and retinal blood vessels in glaucoma - a review," *Microcirculation*, vol. 17, no. 7, pp. 568–581, 2010.

[23] A. V. Stanton, B. Wasan, A. Cerutti, S. Ford, R. Marsh, and P. S. Sever, "Vascular network changes in the retina with age and hypertension," *Journal of Hypertension*, vol. 13, no. 12, pp. 1724–1728, 1995.

[24] L. A. King, A. V. Stanton, P. S. Sever, S. Thom, and A. D. Hughes, "Arteriolar length/diameter (L:D) ratio: A geometric parameter of the retinal vasculature diagnostic of hypertension," *Journal of Human Hypertension*, vol. 10, pp. 417–418, 1996.

[25] J. E. Grunwald, C. E. Riva, S. H. Sinclair, A. J. Brucker, and B. L. Petrig, "Laser doppler velocimetry study of retinal circulation in diabetes mellitus.," *Archives of Ophthalmology*, vol. 104, pp. 991–996, 1986.

[26] S. H. Sinclair, "Macular retinal capillary hemodynamics in diabetic patients," *Ophthalmology*, vol. 98, pp. 1580–1586, 1991.

[27] O. Arend, S. Wolf, F. Jung, B. Bertram, H. Pstgens, H. Toonen, and M. Reim, "Retinal microcirculation in patients with diabetes mellitus: dynamic and morphological analysis of perifoveal capillary network," *British Journal of Ophthalmology*, vol. 75, pp. 514–518, 1991.

[28] M. E. Langham, R. Farrell, T. Krakau, and D. Silver, "Ocular pulsatile blood flow, hypotensive drugs, and differential light sensitivity in glaucoma," in *Glaucoma Update IV* (G. K. Krieglstein, ed.), pp. 162–172, Berlin, Germany: Springer-Verlag, 1991.

[29] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon, , and R. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, p. 673689, 1996.

[30] H. R. Novotny and D. L. Alvis, "A method of photographing fluorescence in circulating blood in the human retina," *Circulation*, vol. 28, pp. 82–86, 1961.

[31] J. H. Hipwell, A. Manivannan, P. Vieira, P. F. Sharp, and J. V. Forrester, "Quantifying changes in retinal circulation: the generation of parametric images from fluorescein angiograms," *Physiological Measurement*, vol. 19, no. 2, pp. 165–180, 1998.

[32] K. R. Brookenthal, K. B. Freedman, and P. A. Lotke, "A meta-analysis of thromboembolic prophylaxis in total knee arthroplasty," *The Journal of Arthroplasty*, vol. 16, pp. 293–300, 2001.

[33] C. W. Colwell, S. D. Berkowitz, B. L. Davidson, P. A. Lotke, J. S. Ginsberg, J. R. Lieberman, J. Neubauer, J. L. Mcelhattan, G. R. Peters, and C. W. Francis, "Comparison of ximelagatran, an oral direct thrombin inhibitor, with enoxaparin for the prevention of venous thromboembolism following total hip replacement. a randomized, double-blind study.," *Journal of Thrombosis and Haemostasis*, vol. 1, no. 10, pp. 2119–2130, 2003.

[34] P. E. P. P. trial Collaborative Group, "Prevention of pulmonary embolism and deep vein thrombosis with low dose aspirin: Pulmonary embolism prevention (PEP) trial," *The Lancet*, vol. 355, no. 9212, pp. 1295–1302, 2000.

[35] D. Hill, M. J. Aguilar, A. Baranco, P. de Lanerolle, and F. Gerbode, "Neuropathological manifestation of cardiac surgery," *The Annals of Thoracic Surgery*, vol. 1, pp. 409–419, 1969.

[36] K. M. Taylor, "Brain damage during open-heart surgery," *Thorax*, vol. 37, no. 12, pp. 873–876, 1982.

[37] C. Blauth, J. Arnold, E. M. Kohner, and K. M. Taylor, "Retinal microembolism during cardiopulmonary bypass demonstrated by fluorescein angiography," *Lancet*, vol. 328, no. 8511, pp. 837–839, 1986.

[38] M. P. Spencer, G. I. Thomas, S. C. Nicholls, and L. R. Sauvage, "Detection of middle cerebral artery emboli during carotid endarterectomy using transcranial doppler ultrasonography," *Stroke*, vol. 21, no. 3, pp. 415–423, 1990.

[39] R. Yuste, "Fluorescence microscopy today," *Nature Methods*, vol. 2, no. 12, pp. 902–904, 2005.

[40] E. D. Schutter, *Computational Modeling Methods for Neuroscientists*. The MIT Press, 1st ed., 2009.

[41] D. Jaeger, *Accurate reconstruction of neuronal morphology*, pp. 159–178. 2001.

[42] Z. Emri, K. Antal, A. I. Gulyás, M. Megías, and T. F. Freund, "Electrotonic profile and passive propagation of synaptic potentials in three subpopulations of hippocampal ca1 interneurons," *Neuroscience*, vol. 104, no. 4, pp. 1013–1026, 2001.

[43] C. Stosiek, O. Garaschuk, K. Holthoff, and A. Konnerth, "In vivo two-photon calcium imaging of neuronal networks," *Proceedings of the National Academy of Sciences*, vol. 100, no. 12, pp. 7319–7324, 2003.

[44] A. P. Davison, J. Feng, and D. Brown, "A reduced compartmental model of the mitral cell for use in network models of the olfactory bulb," *Brain Research Bulletin*, vol. 51, no. 5, pp. 393–399, 2000.

[45] T. Lee and L. Luo, "Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis," *Neuron*, vol. 22, no. 3, pp. 451–461, 1999.

[46] R. J. Harvey and R. M. A. Napper, "Quantitatives studies on the mammalian cerebellum," *Progress in Neurobiology*, vol. 36, pp. 437–463, 1991.

[47] S. R. Schultz, K. Kitamura, A. Post-Uiterweer, J. Krupic, and M. Hausser, "Spatial pattern coding of sensory information by climbing fiber-evoked calcium signals in

networks of neighboring cerebellar Purkinje cells.," *The Journal of Neuroscience*, vol. 29, no. 25, pp. 8005–8015, 2009.

[48] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 263–269, 1989.

[49] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 113–125, 1998.

[50] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 131–137, 2003.

[51] A. F. Frangi, W. J. Niessen, K. L. Vincken, M. A. Viergever, W. William, C. Alan, and D. Scott, "Live-vessel: extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 1496, p. 130, 1998.

[52] M. E. Martinez-Perez, A. D. Hughes, A. V. Stanton, S. A. Thom, A. A. Bharath, and K. H. Parker, "Retinal blood vessel segmentation by means of scale-space analysis and region growing," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 1679, pp. 90–97, 1999.

[53] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical Image Analysis*, vol. 11, pp. 47–61, 2007.

[54] J. Ng, S. T. Clay, S. A. Barman, A. R. Fielder, M. J. Moseley, K. H. Parker, and C. Paterson, "Maximum liklihood estimatation of vessel parameters from scale space analysis," *Image Vision Computing*, vol. 28, no. 1, pp. 55–63, 2010.

[55] T. Mouroutis, S. J. Roberts, and A. A. Bharath, "Robust cell nuclei segmentation using statistical modelling," *Bioimaging*, vol. 6, p. 7991, 2001.

[56] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, pp. 146–165, 2004.

[57] F. Zana and J. C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and hough transform," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 419–428, 1999.

[58] C. Ali, S. Hong, 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*, vol. 3, p. 125138, 1998.

[59] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images.," *British Journal of Ophthalmology*, vol. 83, pp. 902–910, 1999.

[60] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Transactions on Medical Imaging*, pp. 1214–1222, 2006.

[61] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1357–1365, 2007.

[62] Y. A. Tolias and S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 263–273, 1998.

[63] F. Villalobos-Castaldi, E. Felipe-Riverón, and L. Sánchez-Fernández, "A fast, efficient and automated method to extract vessels from fundus images," *Journal of Visualization*, p. 263270, 2010.

[64] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature based, robust, hierarchical algorithm for registering pairs of images of the curved human retina," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, pp. 347–364, 2002.

[65] A. A. Bharath, "Steerable filters from erlang functions," in *British Machine Vision Conference (BMVC)*, pp. 144–153, 1998.

[66] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 948–959, 2005.

[67] J. Ng and A. A. Bharath, "Steering in scale space to optimally detect image structures," in *Proceedings of 8th European Conference on Computer Vision (ECCV)*, vol. 3021 of *LNCS*, pp. 482–494, 2004.

[68] S. Cao, S. Iftikhar, and A. A. Bharath, "Patch-based feature maps for pixel-level image segmentation," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 2263–2267, 2012.

[69] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[70] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671–682, 2006.

[71] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[72] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (D. E. Rumelhart and J. L. McClelland, eds.), ch. 8, Cambridge, MA and London, England: MIT Press, 1986.

[73] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, p. 6, 1994.

[74] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[75] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001.

[76] W. R. Crum, T. Hartkens, and D. L. G. Hill, "Non-rigid image registration: theory and practice," *British Journal of Radiology*, vol. 77, pp. S140–S153, 2004.

[77] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey,"

[78] H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern Recognition*, vol. 32, no. 1, pp. 129–149, 1999.

[79] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–16, 1998.

[80] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[81] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, vol. 19, pp. 459–466, 1986.

[82] T. Tuytelaars and L. V. Gool, "Matching widely separated views based on affinely invariant neighbourhoods," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[83] J. W. Hsieh, H.-Y. M. Liao, K.-C. Fan, and M.-T. Ko, "A fast algorithm for image registration without predetermining correspondence," in *Proceedings of the 13th International Conference on Pattern Recognition (ICPR))*, vol. 1, pp. 765–769, 1996.

[84] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 320–334, 1995.

[85] C. Wang, H. Sun, S. Yadas, and A. Rosenfeld, "Some experiments in relaxation image matching using corner features," *Pattern Recognition*, vol. 16, no. 2, pp. 167–182, 1983.

[86] G. Stockman, S. Kopstein, and S. Benett, "Matching images to models for registration and object detection via clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, no. 3, pp. 229–241, 1982.

[87] A. Goshtasby, *A symbolically-assisted approach to digital image registration with application in computer vision.* PhD thesis, 1983.

[88] M. Ehlers, "Region-based matching for image registration in remote sensing databases," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, no. 4, pp. 2231–2234, 1991.

[89] W. S. Ali and F. S. Cohen, "Registering coronal histological 2-D sections of a rat brain with coronal sections of a 3-D brain atlas using geometric curve invariants and B-spline representation," *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 957–966, 1998.

[90] B. Likar and F. Pernus, "Automatic extraction of corresponding points for the registration of medical images," *Medical Physics*, vol. 26, no. 8, pp. 1678–1686, 1999.

[91] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 659–663, 1977.

[92] A. V. Cideciyan, "Registration of high resolution images of the retina," in *Proceedings of SPIE Medical Imaging VI*, vol. 1652, pp. 310–322, 1992.

[93] A. Goshtasby, "Piecewise cubic mapping functions for image registration," *Pattern Recognition*, vol. 20, pp. 525–533, 1987.

[94] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast mr images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[95] J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of interpolating methods for image resampling," *IEEE Transactions on Medical Imaging*, vol. 2, no. 1, pp. 31–39, 1983.

[96] W. K. Pratt, "Correlation techniques of image registration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, no. 3, pp. 353–358, 1974.

[97] T. Buzug, J. Weese, C. Fassnacht, and C. Lorenz, "Image registration: Convex weighting functions for histogram-based similarity measures," in *First Joint Conference Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assicated Surgery*, LNCS, pp. 203–212, 1997.

[98] R. N. Bracewell, *The Fourier Transform and Its Applications*. McGraw-Hill series in electrical and computer engineering, McGraw-Hill, 3rd ed., 2000.

[99] L. M. G. Brown and T. E. Boult., "Registration of planar film radiographs with computed tomography.," in *Proceedings of the Mathematical Methods in Biomedical Image Analysis (MMBIA) 1996*, pp. 42–51, 1996.

[100] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, , and D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 586–595, 1999.

[101] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[102] W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, "A robust graph transformation matching for non-rigid registration," *Image Vision Computing*, vol. 27, pp. 897–910, 2009.

[103] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1973–1989, 2007.

[104] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[105] P. C. Cattin, H. Bay, L. V. Gool, and G. Szkely, "Retina mosaicing using local features," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 4191, pp. 185–192, 2006.

[106] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 11, pp. 1379–1394, 2003.

[107] A. Almhdie, C. Léger, M. Deriche, and R. Lédée, "3D registration using a new implementation of the ICP algorithm based on a comprehensive lookup matrix: Application to medical imaging," *Pattern Recognition Letters*, vol. 28, no. 12, pp. 1523–1533, 2007.

[108] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[109] S. Lee, J. M. Reinhardt, P. C. Cattin, and M. D. Abrmoff, "Objective and expert-independent validation of retinal image registration algorithms by a projective imaging distortion model," *Medical Image Analysis*, vol. 14, no. 4, pp. 539–549, 2010.

[110] J. C. Nunes, Y. Bouaoune, E. Delechelle, and P. Bunel, "A multiscale elastic registration scheme for retinal angiograms," *Computer Vision and Image Understanding*, vol. 95, no. 2, pp. 129–149, 2004.

[111] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th ALVEY Vision Conference*, pp. 147–151, 1988.

[112] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, pp. 11–15, 1972.

[113] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[114] J. P. Lewis, "Fast normalized cross-correlation," in *Vision Interface*, pp. 120–123, 1995.

[115] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed., 2004.

[116] B. Fischer and J. Modersitzki, "Ill-posed medicine an introduction to image registration," *Inverse Problems*, vol. 24, no. 3, 2008.

[117] C. Marino, G. Penedo, M. Penas, J. Carreira, and F. Gonzalez, "Personal authentication using digital retinal images," *Pattern Analysis & Applications*, vol. 9, no. 1, pp. 21–33, 2006.

[118] B. Dhillonb, R. H. Eikelboomf, K. Yogesana, and I. J. Constablea, "Retinal image analysis: Concepts, applications and potential," *Progress in Retinal and Eye Research*, vol. 25, pp. 99–127, 2006.

[119] A. M. Mendonca and A. Campilho, "Improved automatic midline tracing of neurites with neuromantic," *Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction*, vol. 25, pp. 1200–1213, 2006.

[120] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-D modeling," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 1196–1204, 2004.

[121] O. Wink, W. J. Niessen, and M. A. Viergever, "Multiscale vessel tracking," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 130–133, 2004.

[122] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1531–1546, 2006.

[123] O. Chutatape, Z. Liu, and S. M. Krishnan, "Retinal blood vessel detection and tracking by matched gaussian and kalman filters," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1998.

[124] I. Liu and Y. Sun, "Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme," *IEEE Transactions on Medical Imaging*, vol. 12, pp. 334–341, 1993.

[125] M. A. Sutton, J.-J. Orteu, and H. Schreier, *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*. Springer, 1st ed., 2009.

[126] M. Halavi, K. A. Hamilton, R. Parekh, and G. A. Ascoli, "Digital reconstructions of neuronal morphology: Three decades of research trends," *Frontiers in Neuroscience*, vol. 6, no. 49, 2012.

[127] E. M. Glaser and H. V. D. Loos, "A semi-automatic computer-microscope for the analysis of neuronal morphology," *IEEE Transactions on Biomedical Engineering*, vol. BME-12, no. 1, pp. 22–31, 1965.

[128] J. J. Capowski, "An automatic neuron reconstruction system," *Journal of Neuroscience Methods*, vol. 8, no. 4, pp. 353–364, 1983.

[129] Q. Wu, F. A. Merchant, and K. R. Castleman, *Microscope Image Processing*. Academic Press, 1st ed., 2008.

[130] P. Sarder and A. Nehorai, "Deconvolution methods for 3-d fluorescence microscopy images," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 32–45, 2006.

[131] J. B. Pawley, *Handbook of Biological Confocal Microscopy*. Springer, 3rd ed., 2006.

[132] P. J. Broser, S. E., V. G., P. Osten, B. Sakmann, and D. J. Wallace, "Automated axon length quantification for populations of labelled neurons," *Journal of Neuroscience Methods*, vol. 169, no. 1, pp. 43–54, 2008.

[133] M. Masseroli, A. Bollea, and G. Forloni, "Quantitative morphology and shape classification of neurons by computerized image analysis," *Computer Methods and Programs in Biomedicine*, vol. 41, no. 2, pp. 89–99, 1993.

[134] W. Yu, H. K. Lee, S. Hariharan, W. Bu, and S. Ahmed, "Quantitative neurite outgrowth measurement based on image segmentation with topological dependence," *Cytometry Part A*, vol. 75A, no. 4, pp. 289–297, 2009.

[135] A. Shah, C. Fischer, C. F. Knapp, and B. F. Sisken, "Quantitation of neurite growth parameters in explant cultures using a new image processing program," *Journal of Neuroscience Methods*, vol. 136, no. 2, pp. 123–131, 2004.

[136] J. J. G. Leandro, R. M. Cesar-Jr, and L. da F. Costa, "Automatic contour extraction from 2D neuron images," *Journal of Neuroscience Methods*, vol. 177, no. 2, pp. 497–509, 2009.

[137] P. Vallotton, R. Lagerstrom, C. Sun, M. Buckley, D. Wang, M. D. Silva, S.-S. Tan, and J. M. Gunnersen, "Automated analysis of neurite branching in cultured cortical neurons using hca-vision," *Cytometry Part A*, vol. 71A, no. 10, pp. 889–895, 2007.

[138] M. Pool, J. Thiemann, A. Bar-Or, and A. E. Fournier, "Neuritetracer: A novel imagej plugin for automated quantification of neurite outgrowth," *Journal of Neuroscience Methods*, vol. 168, no. 1, pp. 134–139, 2008.

[139] F. Janoos, K. Mosaliganti, X. Xu, R. Machiraju, K. Huang, and S. T. C. Wong, "Robust 3D reconstruction and identification of dendritic spines from optical microscopy imaging," *Medical Image Analysis*, vol. 13, no. 1, pp. 167–179, 2009.

[140] J. Fan, X. Zhou, J. G. Dy, Y. Zhang, and S. T. C. Wong, "An automated pipeline for dendrite spine detection and tracking of 3d optical microscopy neuron images of in vivo mouse models," *Neuroinformatics*, vol. 7, pp. 113–130, 2009.

[141] X. Yuan, J. T. Trachtenberg, S. M. Potter, and B. Roysam, "MDL constrained 3-D grayscale skeletonization algorithm for automated extraction of dendrites and spines from fluorescence confocal images," *Neuroinformatics*, vol. 7, pp. 213–232, 2009.

[142] C. Sun and P. Vallotton, "Fast linear feature detection using multiple directional non-maximum suppression," *Journal of Microscopy*, vol. 234, no. 2, pp. 147–157, 2009.

[143] S. L. Wearne, A. Rodriguez, D. B. Ehlenberger, A. B. Rocher, S. C. Henderson, and P. R. Hof, "New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales," *Neuroscience*, vol. 136, no. 3, pp. 661–680, 2005.

[144] J. C. Fiala, "Reconstruct: a free editor for serial section microscopy," *Journal of Microscopy*, vol. 218, no. 1, pp. 52–61, 2005.

[145] R. Srinivasan, X. Zhou, E. Miller, J. Lu, J. Litchman, and S. T. C. Wong, "Automated axon tracking of 3d confocal laser scanning microscopy images using guided probabilistic region merging.," *Neuroinformatics*, vol. 5, pp. 189–203, 2007.

[146] H. Peng, Z. Ruan, D. Atasoy, and S. Sternson, "Automatic reconstruction of 3d neuron structures using a graph-augmented deformable model," *Bioinformatics*, vol. 26, pp. i38–i46, 2010.

[147] D. Myatt and S. Nasuto, "Improved automatic midline tracing of neurites with neuromantic," *BMC Neuroscience*, vol. 9, no. Suppl 1, p. P81, 2008.

[148] J. Bigün and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proceedings of the IEEE First International Conference on Computer Vision*, p. 433438, 1987.

[149] B. Rieger, L. J. van Vliet, and P. W. Verbeek, "Continuous orientation representation for arbitrary dimensions - a generalized Knutsson mapping," in *SCIA*, pp. 774–783, 2011.

[150] C. E. Metz, "Basic principles of ROC analysis.," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, 1978.

[151] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letter*, vol. 27, no. 8, pp. 861–874, 2006.

[152] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1488–1497, 2009.

[153] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 2, pp. 1137–1143, 1995.

[154] B. Zitov and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[155] A. J. Watt, H. Cuntz, M. Mori, Z. Nusser, P. J. Sjostrom, and M. Häusser, "Traveling waves in developing cerebellar cortex mediated by asymmetrical Purkinje cell connectivity," *Nature Neuroscience*, vol. 12, no. 4, pp. 463–473, 2009.

[156] H. Cuntz, F. Forstner, A. Borst, and M. Häusser, "One rule to grow them all: A general theory of neuronal branching and its practical application," *PLoS Computational Biology*, vol. 6, no. 8, p. e1000877, 2010.