

DISS. ETH No. 19559

**Machine Learning Approaches for Neuron
Geometry Extraction and Synapse Detection
in Electron Microscopy Images**

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
VERENA SABINE KAYNIG-FITTKAU
Dipl. Inf., University of Hamburg
born 29 September 1979
citizen of Germany

accepted on the recommendation of
Prof. Dr. Joachim M. Buhmann, examiner
Prof. Dr. Pascal Fua, co-examiner
Prof. Dr. Kevan Martin, co-examiner

2011

Acknowledgments

I am very grateful to the following persons for their support and effort during my studies. First of all, I want to thank my advisor Joachim M. Buhmann, who taught me how to do research and guided my efforts with critical questions and advise. I am also grateful to Roger Wepf, who introduced me to electron microscopy and provided me with the idea for my project and financial support. I want to thank Kevan Martin, Nuno M. da Costa and Albert Cardona for the great collaboration and very valuable discussions about neuroanatomy. Furthermore, I want to thank Bernd Fischer who supported me with fruitful discussions and ideas regarding image registration and the distortion correction. I also thank Thomas Fuchs for motivating and valuable discussions and for improving my writing skills. I am grateful to Elisabeth Müller for her support with the electron microscopy experiments and the evaluation of the distortion correction. I am also grateful to Shona Cameron for proofreading parts of this thesis and to the Fiji development team for their support in implementing open source plugins. In addition I want to thank all the people who spend time and effort for data acquisition and annotation. The cat brain electron microscopy images used throughout this thesis were recorded by Rita Bopp and German Köstinger and annotated by Nuno M. da Costa and John Anderson. The drosophila larva images were provided and annotated by Albert Cardona. Last not least, my deepest thanks go to my colleagues at EMEZ and the ML-group and to my husband and my family for their loving support throughout the last years.

Abstract

In this thesis we investigate the whole image processing pipeline for neuronal geometry extraction and synapse detection in electron microscopy images. Advancements in automated sample preparation and image acquisition for electron microscopy enable recording of large data sets. This process is especially important for the field of computational neuroanatomy and connectomics, as the analysis of neuronal connections requires imaging of large volumes with a resolution sufficient for synapse detection. Manual processing of electron microscopy data is time-consuming and becoming the main bottleneck in gaining new insights into the functional structure of the brain. Automated processing of biological electron microscopy images is challenging due to the rich texture, low signal to noise ratio and the great variability of image characteristics depending on sample preparation and animal type. To enable quantitative evaluation of the data, the images are corrected against lens distortions, stitched, and aligned. Structures of interest are then segmented and grouped across serial sections to extract the 3d geometry. The proposed registration methods employ unsupervised approaches to identify artifact signals like non-linear distortions, cracks, or staining blurs. We demonstrate that identification of these signals leads to superior registration results compared to state-of-the-art methods. The distortion correction enables structure preserving mosaicing with sub-pixel precision. The non-linear distortion field is estimated from overlapping image areas and does not require special calibration samples.

To solve for correspondences between adjacent images we propose a non-linear warping with anomaly detection. Correspondences are not fixed beforehand, but estimated as latent variables in an expectation maximization framework, which assigns image anomalies to an outlier class. Our approaches to membrane segmentation and geometry extraction combine the output of a trained random forest classifier with biologically inspired smoothness constraints. We demonstrate that the proposed feature sets capture the relevant statistics with very little training data. Thus, the framework can be adapted to data sets of varying animal types and staining protocols with minimal effort. For the segmentation of thin elongated structure we propose a sub-modular binary energy term which enforces gap completion by good continuation. An additional term incorporates membrane detections from adjacent sections. The resulting energy function can be globally optimized using graph cut and significantly outperforms state-of-the-art segmentation using gradient flux. Geometry extraction requires grouping of the segmented regions to 3d objects. The output of a classifier trained for region similarity based on shape and texture features is combined with geometrical consistency constraints to obtain a weighted connectivity matrix. Agglomerative clustering is employed to find 3d groupings of correspondent regions. We demonstrate that our method yields fully automatic reconstructions of drosophila larva neurites over 30 sections. With respect to dense reconstruction and synapse classification the proposed framework can be employed for semi automatic processing, enabling the analysis of large data sets.

Zusammenfassung

Diese Dissertation behandelt den vollständigen Bildverarbeitungsprozess für die Geometriekonstruktion von Neuronen und die Identifikation von Synapsen in Elektronenmikroskopiebildern. Durch die Fortschritte in der Elektronenmikroskopie im Bereich der automatischen Probenpräparation sowie der Bildaufnahme ist es möglich, große Datensätze zu erstellen. Diese Entwicklung ist besonders wichtig für den Bereich der rechnergestützten Neuroanatomie und der Analyse von Neuronennetzwerken. Die Analyse neuronaler Kontakte benötigt Bilddatensätze über große Volumina, deren Auflösung die Identifikation von Synapsen zulässt. Manuelle Verarbeitung von Elektronenmikroskopiedaten ist zeitaufwendig und stellt ein bedeutendes Hindernis für neue Einsichten in die funktionale Struktur des Gehirns dar. Durch die starke Texturierung der Bilder und ein geringes Signal-Rauschverhältnis sowie eine große Varianz in Bezug auf Bildcharakteristiken, abhängig von der Probenpräparation und der Tierart, ist die automatische Verarbeitung dieser Daten eine große Herausforderung. Die vorgeschlagenen Registrierungsmethoden verwenden unüberwachte Methoden, um Artefaktsignale wie nichtlineare Verzerrungen, Risse oder Färbungsflecken zu identifizieren. Wir zeigen, dass die Erkennung dieser Signale bessere Ergebnisse liefert, als aktuelle Vergleichsmethoden. Die Korrektur von Linsenverzerrungen ermöglicht die struktur-erhaltende Montage von Bildern mit Subpixelpräzision. Hierzu wird das nichtlineare Verzerrungsfeld aus überlappenden Bildbereichen geschätzt, ohne auf speziellen Kalibrierungsproben zurückzugreifen. Zur

Lösung des Korrespondenzproblems zwischen benachbarten Bildern verwenden wir eine nichtlineare Registrierung, welche enthaltene Anomalien erkennt. Korrespondenzen werden nicht zuvor festgelegt, sondern während der Optimierung geschätzt. Unsere Methode verwendet Erwartungswertmaximierung, wobei Bildanomalien einer Ausreißerklasse zugewiesen werden. Für die Segmentierung von Membranen und die Extraktion von Geometrien kombinieren wir die Ausgabe eines trainierten Klassifikators mit biologisch inspirierten Glattheitsbeschränkungen. Wir zeigen, dass die vorgeschlagenen charakteristischen Merkmalsmengen die relevanten Statistiken mit einer geringen Menge von Trainingsdaten erfassen. Folglich kann das System für unterschiedliche Datensätze mit geringem Aufwand angepasst werden. Für die Segmentierung von dünnen, gerichteten Strukturen stellen wir einen submodularen, binären Energieterm vor, der Lücken nach dem Prinzip der guten Fortsetzung schließt. Ein weiterer Energieterm berücksichtigt Membranendetektionen von benachbarten Schichten. Die resultierende Energiefunktion kann durch Grafenzerlegung global optimiert werden und erzielt bessere Ergebnisse als aktuelle Segmentierungsmethoden, die das Gradientenfeld berücksichtigen. Für die Extraktion von Geometrien ist es notwendig segmentierter Regionen zu dreidimensionalen Objekten zu gruppieren. Hierzu kombinieren wir die Ausgabe eines Klassifikators, der auf der Grundlage von Form und Textur für die Erkennung von ähnlichen Regionen trainiert wurde, mit geometrischen Konsistenzbeschränkungen, um eine gewichtete Konnektivitätsmatrix zu erhalten. Hierarchische Gruppierung wird verwendet, um dreidimensionale Objekte bestehend aus korrespondierenden Regionen zu erhalten. Wir zeigen, dass unsere Methode die vollständig automatische Rekonstruktion von neuronalen Zellfortsätzen der *Drosophila* Larve über 30 Probenschnitte ermöglicht. In Bezug auf dichte Rekonstruktionen und die Klassifikation von Synapsen kann das vorgeschlagene Rahmenwerk dazu verwendet werden, Bilder halb automatisch zu verarbeiten und so die Annotierung großer Datensätze ermöglichen.

Contents

1	Introduction	3
1.1	Computational Neuroanatomy and Connectomics	4
1.2	Image Processing Workflow for Connectomics	5
1.3	Original Contributions	7
1.4	Organization	9
1.5	Publications	10
2	Electron Microscopy Techniques	11
2.1	Transmission Electron Microscopy	11
2.1.1	Sample Preparation	16
2.2	High Throughput Techniques	21
3	Registration	27
3.1	Automatic Landmark Extraction with SIFT	28
3.2	Stitching	31
3.2.1	Robust Stitching of Overlapping Regions	33
3.2.2	Multiple Resolutions	36
3.3	Distortion Correction	38
3.3.1	Related Work	41
3.3.2	Origins of Distortions	43
3.3.3	Estimation Framework	45
3.3.4	Experimental Validation	49
3.3.5	Simulated Ground Truth Data	50
3.3.6	Electron Microscopy Data Experiments	51

3.4	Z-Alignment: Affine vs. Non-linear Alignment	61
3.5	Affine Registration	62
3.6	Non-linear Warping with Anomaly Detection	64
3.6.1	Expectation Maximization	66
3.6.2	Experiments	70
3.6.3	Conclusion	76
4	Geometry Extraction	79
4.1	Segmentation	79
4.1.1	Random Forest Classifier for Membrane Detection	81
4.1.2	Segmentation with Graph Cut	82
4.1.3	Graph Cut with Gradient Flux	85
4.1.4	Closing Gaps: Submodular Energy Term for Good Continuation	89
4.1.5	Experiments and Results	91
4.1.6	Conclusion	96
4.2	Region Grouping	100
4.2.1	Similarity Measures between Regions	102
4.2.2	Global Geometrical Consistency	105
4.2.3	Evaluation	107
4.2.4	Conclusion	110
5	Functional Structure Annotation	111
5.1	Basic Structure Annotation	112
5.1.1	Vesicle detection	113
5.2	Synapse Detection	116
6	Conclusion	125
A	Derivation of Closed Form Solutions	127
A.1	Distortion Correction Transform	127
A.2	Non-linear Warping	132
	Curriculum Vitae	145



Chapter 1

Introduction

Biological electron microscopy images pose a great challenge for automated analysis with pattern recognition methods. Due to electron beam damage, the electron dose used for image acquisition is limited and leads to low signal to noise ratio. In addition, biological samples are densely packed with structures, which lead to rich image textures and non uniform background. Different data sets typically express a large variability not only by shape and size of biological structures, but also in terms of image characteristics dependent on the staining protocol and the microscopy technique. The acquired images typically consist of a combination of structural and artifact signals, which need to be identified. In this thesis we propose a range of automated image processing methods for electron microscopy data, in the context of computational neuroanatomy and connectomics. The main idea behind the proposed techniques is to discriminate relevant signal in data for geometry reconstruction from irrelevant signal related to sample preparation and to processing artifacts. For the developed image registration methods, this discrimination is learned in an unsupervised fashion. For image segmentation and structure annotation we employ supervised learning and combine the output of a trained classifier with constraints which are biologically given and thus invariant to animal types or staining protocols.

1.1 Computational Neuroanatomy and Connectomics

The field of computational neuroanatomy investigates neuronal structures employing computational techniques, such as visualization, modeling, and analysis. An important aspect of computational neuroanatomy is to identify relations between biological structure and function. To gain insight in the functional structure of the brain, neuroanatomists analyze the structure of individual neurons, the density of neuronal elements within specific brain areas and their connectivity. Imaging methods employed for data acquisition range from high resolution electron tomography of single synapses over light microscope images of neurons to MIR imaging of whole brains. While light microscopy is sufficient in resolution for the analysis of single neurons, only electron microscopy can provide a resolution which renders the identification of synapses possible. Thus, the field of neuroanatomy faces the challenge to acquire and analyze data volumes which cover a brain tissue volume large enough to allow meaningful analysis of circuits and detailed enough to detect the synapses and thus the connectivity structure of the circuit. White et al. (1986) reconstructed the connectivity matrix for *C. elegans* from serial section transmission electron microscopy images. The *C. elegans* worm consists of 959 cells of which 302 are neurons. In comparison, a cubic millimeter of cerebral cortex contains roughly 50,000 neurons. While recent advances in the sample preparation and image acquisition workflow of electron microscopy make significant progress towards automatic acquisition of large data sets, manual segmentation and annotation of these images is a serious bottleneck for the evaluation of the acquired data. Identification of synapses requires a resolution of 5 nm per pixel. Imaging a one cubic millimeter of brain tissue at this resolution and with a section thickness of 50 nm leads to 20,000 sections of 20 gigapixels each. Thus, automated image processing in this context is not only a spoonful of sugar to lighten the work of the neuroanatomist. Instead, automated image processing is a necessity to gain new insights

in the functional structure of the brain. While computational neuroanatomy deals with all functional structures, e.g. vesicle formations, mitochondria, etc. connectomics as a sub-field of computational neuroanatomy, concentrates on the wiring diagram of the brain. Thus, the two key tasks with respect to image processing are neuron geometry extraction and synapse detection in electron microscopy images. To create connectivity matrices on the cell level, dense reconstruction of all neurons is required. To analyze the connectivity between different brain areas, sparse reconstruction of randomly sampled structures is often employed. The results can be extrapolated to whole brain regions using stereological methods. With respect to the image processing workflow, dense and sparse geometry extraction are very similar. The main difference is that dense reconstruction typically involves processing of larger data sets than sparse reconstruction. Thus, dense reconstruction leads to high demands regarding memory management and parallel processing techniques. But, sparse reconstruction often involves estimation of detailed statistics like volume and shape measurements and hence requires a detailed analysis of the data.

1.2 Image Processing Workflow for Connectomics

The image processing workflow for connectomics, as depicted in Figure 1.1 consists of the following steps:

Distortion correction: The electromagnetic optics of the electron microscope induces a non-linear distortion in the images. Distortion Correction is necessary to preserve structural information in the images and to obtain seamless stitchings in the next step of the processing workflow.

Stitching: To acquire an image larger than the field of view of the camera, the microscopist takes several translated images. These images have to be assembled into one panorama image covering the

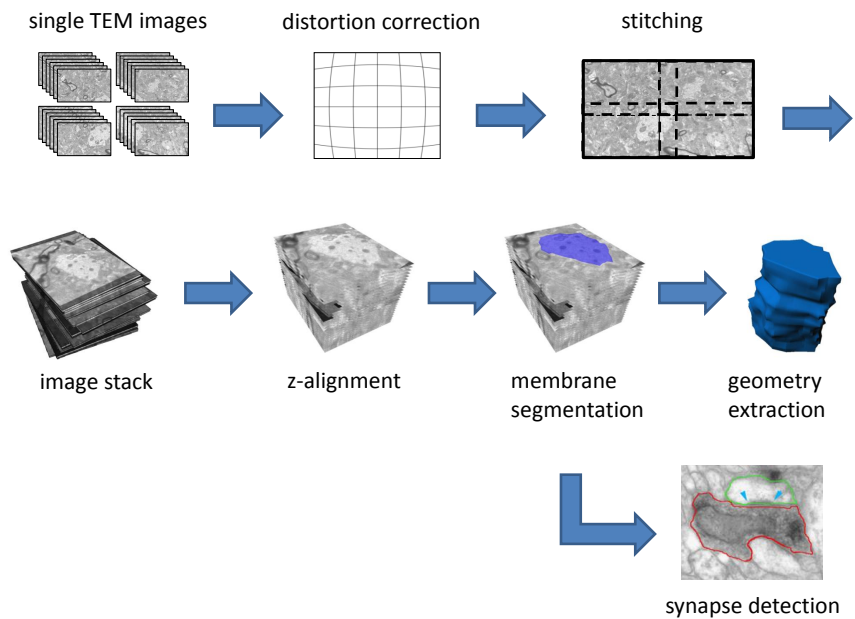


Figure 1.1: Image processing workflow for geometry extraction and synapse detection

area of interest on the sample. This process is called stitching or mosaicing.

Z-alignment: To render recording of an image possible with a transmission electron microscope, the sample has to be sliced into ultra thin sections of 40-60 nm thickness. Afterwards, the images of the single sections have to be aligned into one image stack.

Membrane segmentation: While the texture of neuronal regions can express different characteristics based on intracellular structures and cutting angle, all neuronal regions are surrounded by a cell membrane. Thus, segmentation of these membranes permits identification of single neuronal regions by solving a binary classification task.

Geometry extraction: Due to the anisotropic resolution of transmission electron microscopy images, segmentation of membranes is performed on the two dimensional images. To extract the neuron geometry from the segmentation, corresponding regions have to be grouped across sections to three dimensional objects.

Synapse detection: Detection of synapses identifies connections between neurons. Thus, synapse detection is a crucial task for connectomics. Figure 1.1 depicts the synapse detection as dependent on membrane segmentation. In principle, it is possible to perform synapse detection without this pre-segmentation. However, the presynaptic and postsynaptic membranes are necessary components of a synapse and therefore a pre-segmentation of neuronal membranes significantly reduces possible synapse locations.

1.3 Original Contributions

The main contributions of this thesis are the following:

- **Multi scale stitching:** We developed a framework for multi scale stitching, employing automatic landmark extraction with SIFT features and robust estimation of the stitching transforms with Huber loss. The framework efficiently solves the mosaicing problem without constraints on the specific ordering of tiles (Kaynig et al., 2007).
- **Distortion correction for electromagnetic lenses:** Lens distortions do not only prevent seamless stitchings of images, but also alter the size and shape of structures, leading to biased measurements. We developed a novel auto distortion correction based on overlapping image areas, which corrects for non-linear deformations in the images. The method does not require special calibration samples, preserves structural information and reduces stitching errors to sub-pixel precision. The distortion correction enabled us to evaluate a series of experiments, identifying the projective lens system as the main source for lens distortions in transmission electron microscopes (Kaynig et al., 2010a).
- **Non-linear warping with anomaly detection:** Image artifacts such as cracks or folds in an image, make non-linear warping of adjacent sections challenging. We demonstrate that treating the warping problem in a Bayesian framework with correspondences as latent variables and an additional outlier term, leads to significant improvements in the warping, compared to a warping based on fixed correspondences and robust Huber loss (Kaynig et al., 2008).
- **Gap completion for graph cut:** For the segmentation of thin and elongated structures, graph cut is well known to have problems with shrinking bias. We propose a novel sub-modular energy term, which enhances gap completion in segmented structures. The proposed approach outperforms state-of-the-art segmentation using smoothness and gradient flux on electron mi-

croscopy images as well as street satellite imagery (Kaynig et al., 2010c).

- **Geometrical consistent region grouping:** Geometry extraction of neuronal processes from segmented regions requires grouping these regions across sections. We propose a novel framework for geometry extraction which takes the geometrical consistency of whole sections into account (Kaynig et al., 2010b).
- **Synapse detection:** We present a new feature set for the detection of synapses which takes characteristics of the pre- and post-synaptic side into account. Our experiments demonstrate significant improvement over state-of-the-art methods, which concentrate on the post-synaptic density.
- **Open source software:** The lens distortion correction has been implemented as open source plugin for the `ImageJ` distribution `Fiji` (Fiji, 2010). It also has been integrated into `TrakEM` (Cardona, 2006) by Stephan Saalfeld. Furthermore, we wrote an additional open source plugin for `Fiji`, which allows us to interactively train a random forest classifier for membrane detection. The plugin has been extended to multiclass classification and multithreaded processing by Ignacio Arganda-Carreras and Albert Cardona.

1.4 Organization

The structure of this thesis follows the image processing workflow for connectomics. Chapter two discusses different high throughput electron microscopy techniques and their special challenges with respect to image processing. Chapter three contains the different registration methods, from multi resolution stitching and distortion correction over affine z-alignment and non-linear warping. Chapter four

describes neuron geometry extraction including membrane segmentation with gap enhancement and geometrical consistent region grouping. Chapter five contains methods for functional structure annotation like vesicles and synapses.

1.5 Publications

Parts of this thesis have been published in the following papers:

- Verena Kaynig and Bernd Fischer and Roger Wepf and Joachim M. Buhmann. Fully automatic registration of electron microscopy images with high and low resolution. *Microscopy and Microanalysis*, 13: 198-199, 2007.
- Verena Kaynig, Bernd Fischer, Joachim M. Buhmann. Probabilistic Image Registration and Anomaly Detection by Nonlinear Warping. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- Verena Kaynig, Thomas Fuchs, Joachim M. Buhmann. Neuron Geometry Extraction by Perceptual Grouping in ssTEM Images. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- Verena Kaynig, Bernd Fischer, Elisabeth Müller, Joachim M. Buhmann. Fully Automatic Stitching and Distortion Correction of Transmission Electron Microscope Images. *Journal of Structural Biology*, 171(2):163-173, 2010.
- Verena Kaynig, Thomas Fuchs, Joachim M. Buhmann. Geometrical Consistent 3D Tracing of Neuronal Processes in ssTEM Data, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 209-216, 2010.

Chapter 2

Electron Microscopy Techniques

Electron microscopy is the only image acquisition technique which can provide a resolution of below 5 nm per pixel. Such a fine resolution is necessary to identify neuroanatomical structures like vesicles and synapses. The following chapter describes the different electron microscopy techniques and their implications for automated image processing. The focus of the chapter is on transmission electron microscopy, since this technique is widely used and all data sets used for training and evaluation in this thesis are generated by transmission electron microscopy.

2.1 Transmission Electron Microscopy

Traditionally, Transmission electron microscopy (TEM) is the method of choice for the reconstruction of neuronal circuits. The whole nervous system of *C. elegans* has been manually reconstructed from TEM images (White et al., 1986).

Figure 2.1 depicts the typical design of a basic TEM. At the top of the microscope, an electron gun emits electrons which are accelerated by a high tension (e.g. 100 kV) into the electromagnetic col-

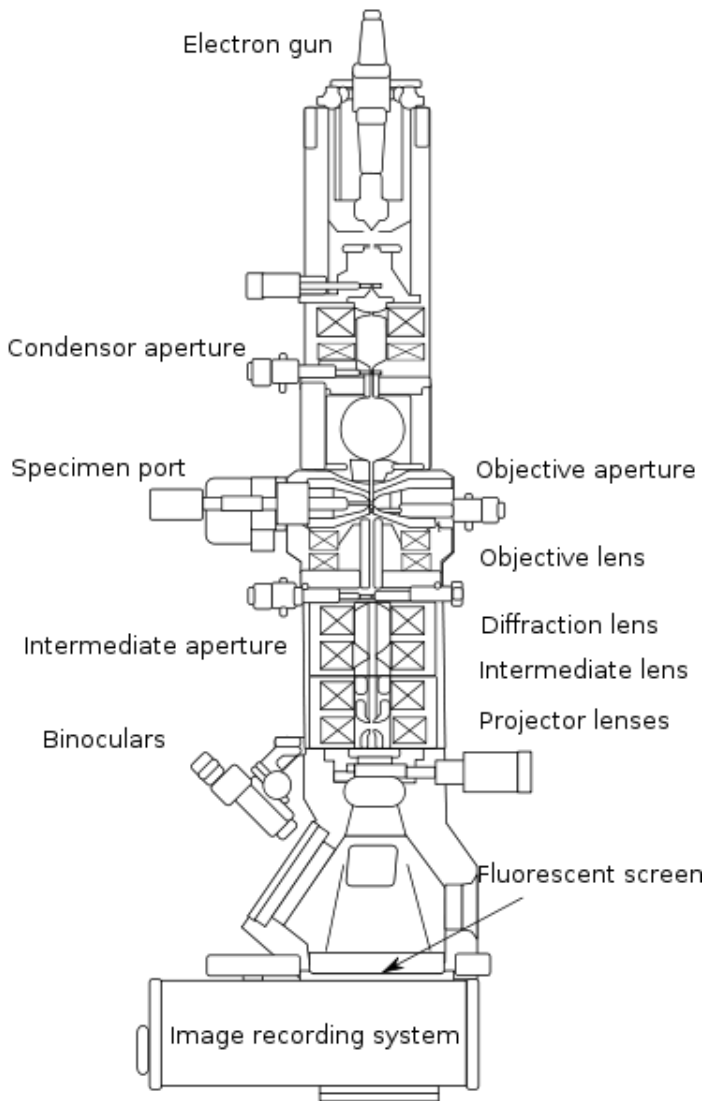


Figure 2.1: Layout of optical components in a basic TEM (Eccles, 2009)

umn. Inside the column the electron beam is first collimated by the condenser lens and aperture and then focused by the objective lens. When the electrons hit the inserted sample, one part of the electrons is absorbed and another is scattered according to the atomic number of the local atoms. For heavy metal stained samples the signal is formed by inserting an objective aperture, which absorbs scattered electrons. Thus, electrons hitting the specimen at thick regions or in regions with high atomic number are excluded from the imaging system, revealing scattering absorption contrast.

After interaction with the sample, the electron beam is focused and magnified by a set of electromagnetic lenses. The objective lens focuses the beam and the projective lens system magnifies the image and expands the beam on the imaging device, i.e. a CCD Camera or a phosphor screen.

The nature of the image formation process in a TEM leads to several important consequences for further processing of the captured images. An obvious consequence is the limitation to gray value images. Unlike light or fluorescence microscopy, where staining of samples can lead to valuable color information, electron microscopy is limited to image formation by electron counts. Stainings can be employed to enhance the electron density of structures, e.g. membranes, and thus improve the image contrast. But, identification of structures like vesicles, mitochondria, boutons, etc., often cannot be based on the gray value of single pixels. Instead, local context like shape information and texture is crucial for structure identification.

Another important aspect of TEM is that the image is in principle a two dimensional projection of the sample. Thus, the section thickness has significant impact on the image quality. Figure 2.2 demonstrates the influence of the section thickness. The upper part of the figure shows a TEM image of a 60 nm section of cat brain tissue and the lower part of the figure an image of a 40 nm section from the same sample. The thick section leads to overlaid projections and multiple electron scattering, reducing the image contrast and rendering the detection of single structures, like vesicles, difficult. Furthermore,

membranes surrounding neuronal processes appear more blurred or even merged in the 60 nm section and textures are not as well defined as in the image of the 40 nm section. Thus, automated structure detection, segmentation and classification becomes more difficult with increased section thickness. Very recent work by Veeraraghavan et al. (2010) virtually reduces the section thickness by reconstruction the section volume from different tomographic views of the TEM section. While this approach leads to much better image quality, the required tilt views lead to considerably more overhead in the image acquisition process.

While the before mentioned aspects of TEM image acquisition mainly influence the detection and segmentation of neuronal structures, there are two additional points which are important for image alignment. These points are the lens distortions induced by the electromagnetic field lenses and the mass loss of the sample due to beam damage. Both aspects lead to non-linear deformations in the images, which render the alignment of TEM images challenging. Electron lenses in principle work similar to optical light microscopy lenses and also lead to respective distortions. TEM lenses typically consist of electromagnetic coils which generate a convex lens to focus the electron beam at a constant focal length. If the produced field is not rotationally symmetric, the lens induces severe aberrations like astigmatism and spherical aberration. Unlike optical lenses, which are solid and not deformable, electromagnetic lenses can be influenced by electrostatic interferences, leading to more variability in possible distortions. In section 3.3 we describe a novel approach for automated distortion correction of TEM images.

Biological samples are sensitive to beam damage, leading to a mass loss during image acquisition. The interaction with the electron beam causes non-linear distortions by physically deforming the sample. As this process is unavoidable and the major part of the damage is caused at first exposure, the sample is often exposed twice to the electron beam. First, in the pre-baking stage, the sample is deformed by the beam. Afterward the actual images are acquired. This technique leads

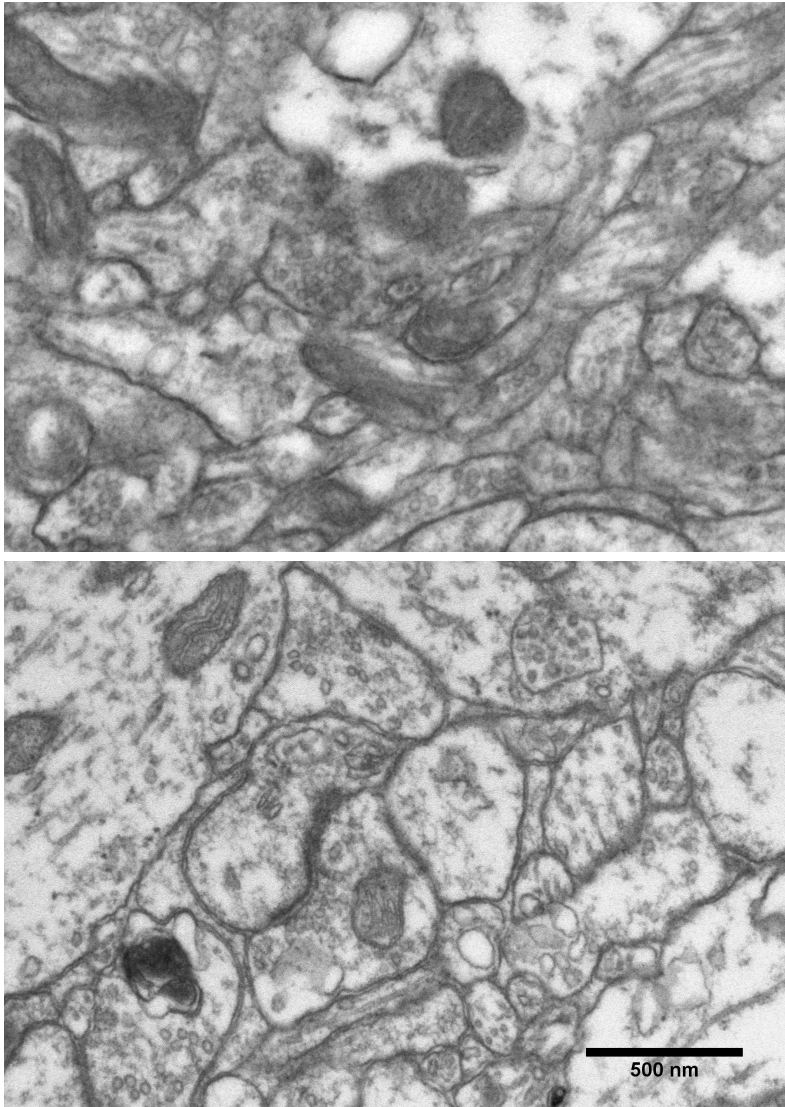


Figure 2.2: Influence of the section thickness on image quality. Top: 60 nm section, bottom: 40 nm section

to stable mosaicing of images in 2d. However, for z-alignment the deformation of the sample leads to a non-linear registration problem (see Section 3.6).

2.1.1 Sample Preparation

Most biological samples have a minimum dimension of more than 100 nm and thus need to be cut into ultra thin sections to be viewed in TEM. Several steps are necessary to enable the cutting of the sample into ultra thin sections of about 40-60 nm and make structures of interest visible in the TEM. According to Flegler et al. (1995) these steps include specimen isolation, fixing, preembedded staining, dehydration, resin infiltration and embedding, ultra thin sectioning, and on section staining. Figure 2.3 provides an overview of the whole sample preparation process.

The specimen isolation and fixation aim at stabilizing the sample protecting it as good as possible from disruption, dimensional change and loss of material in the following dehydration and embedding process. Typically, fixation consists of two steps. In the first step the sample is treated with a mixture of different aldehydes. In the second step, a complementary secondary fixative, i.e. osmium tetroxide, is applied. Regardless of the fixation method, dehydration of the sample is necessary for later embedding. Dehydration is usually accomplished by serially exchanging the water in the sample with a solvent that is soluble with the embedding medium and water. Common agents are ethanol and acetone. To phase the water out of the specimen, the sample is transferred to a graded series of solutions containing the solvent and water. When all water is replaced by the solvent, the tissue is infiltrated by resin monomers and finally polymerized to render the sample solid. Embedding provides support during the sectioning and retains the spatial organization of the section on the TEM grid. A good embedding medium must provide satisfactory cutting properties as well as low viscosity to infuse the sample and compatibility with the staining method applied after cutting the sample. Epoxy resins are widely used for TEM sample embedding. After the speci-

2.1. TRANSMISSION ELECTRON MICROSCOPY

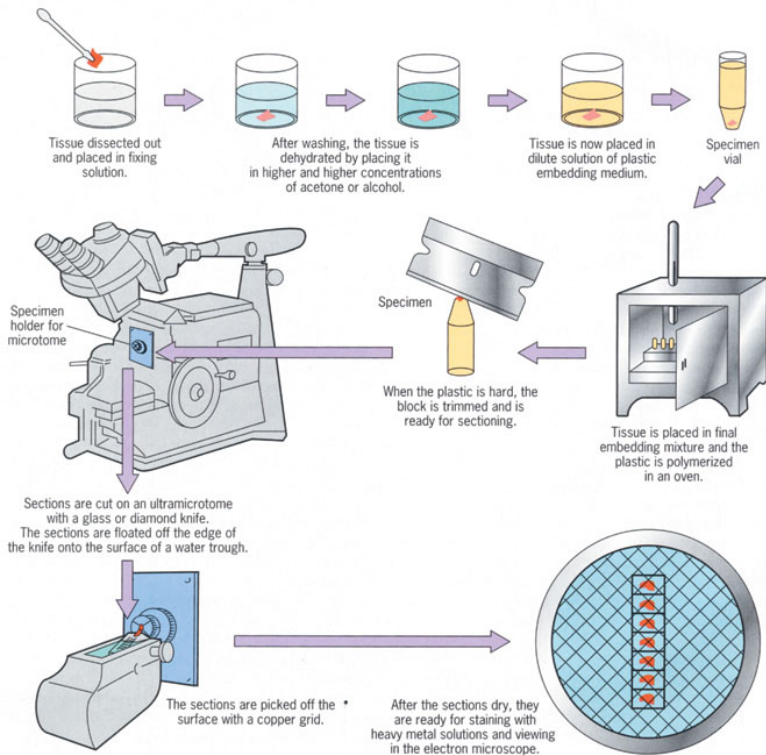


Figure 2.3: Specimen preparation process for transmission electron microscopy (Preparation, 2010)

men has been cast into a block of resin, the next step is to cut ultra thin sections. Ideally the sections should be cut as thin as possible, while still providing uniform thickness. The lower left corner of Figure 2.3 shows a diamond knife with a boat for conventional ultra thin sectioning. The sections cut from the sample float onto the water, forming a ribbon of subsequent sections as the cutting continues.

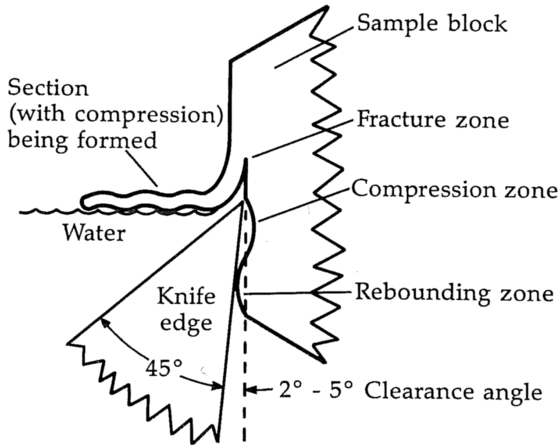


Figure 2.4: Cutting of ultra thin sections with a diamond knife (Flegler et al., 1995).

As can be seen in Figure 2.4 the cutting can cause severe vertical compression of the sections. To reduce compression artifacts, the sections can be temporarily softened by either chemical or thermal treatment. The surface tension of the water then acts on the hydrophobic section and pulls the compression out. Once the sections have been decompressed, they are picked up on a small copper supporting grid, which can be placed into a TEM specimen holder for imaging. Before the grids are transferred to the microscope, the sections are usually stained with an electro-dense stain to increase electron contrast. To work effectively, the stain must be able to penetrate the resin sur-

face of the section. Common EM stains for biological samples are uranylacetate and lead citrate.

In summary, the sample preparation tries to find a compromise between preserving the original sample structure and enabling recording of the specimen in the unsustainable TEM environment. The different stages of the preparation process can cause a variety of artifacts to the structure and the final image reflects the sum of all these changes. Staining can cause dark blobs or stripes in the images, which occlude the structural signal. The final staining procedure takes place after the plastic sections have been cut and adsorbed on a TEM grid. Thus, the staining can vary between sections and grids leading to very different image contrast. During cutting and placement on the grid, plastic sections can suffer folds or cracks, which lead to highly non-linear deformations. In addition, entire sections can be lost during the cutting. Figure 2.5 provides examples for different artifacts caused by the preparation process.

The preparation pipeline also compromises structural geometry information. Shrinkage during chemical fixation can be reduced by high pressure freezing followed by freeze substitution or cryo TEM. But, these techniques are very difficult to apply, expensive and time consuming and thus not applicable for large neuroanatomy projects. With respect to geometry extraction, the most interesting part of the sample preparation is the actual cutting of sections. Slicing the sample and recording images of the single sections can be seen as a discrete sampling of the neuron geometry in the direction orthogonal to the cutting plane. According to the sampling theorem, the section thickness limits the geometry reconstruction to frequencies lower than half the sampling frequency. In general this limitation is not severe for neurites, as the geometry is smooth enough to be reconstructed from 40 nm sections. However, small structures like spine necks or thin parts of dendrites can be lost due to the coarse sampling. Cutting of serial sections not only leads to a discrete sampling of the neurite geometry, it also causes non-linear deformations of the single sections. As depicted in Figure 2.4, a lot of stress is put on the sam-

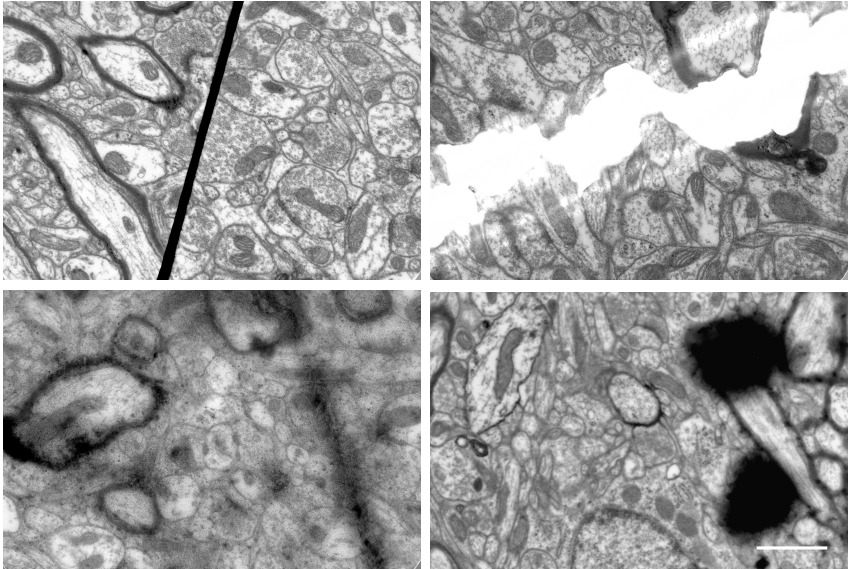


Figure 2.5: Image artifact examples. Upper left: plastic fold, upper right: crack in plastic section, lower left: badly stained section, lower right: staining blobs. The scale bar corresponds to $1\mu\text{m}$

ple during sectioning, leading to non-linear transformations. Thus, changes in the geometry signal between adjacent sections are caused by two non-linear transformations. One transformation consists of the changes in structural geometry of the neurites between sections. The other transformations represents artifact deformations caused by cutting of the sample and beam damage during image acquisition. An ideal image alignment would correct for artifact transformations, while preserving the structural changes between sections. As such an alignment method does not exist up to today, images are typically registered with restriction to rigid or affine transformations. The linear image alignment preserves all structural transformations, at the cost of also keeping all artifact transformations in the reconstructed signal. Saalfeld et al. (2010) recently proposed as-rigid-as-possible image registration, which follows the rigid alignment approach, but allows for non-linear adjustments to gain smooth 3d reconstructions. However, no distinction is made between structural and artifact transformations. From the experimental side, it is possible to reconstruct the sample geometry with correlative microscopy. Before serial sectioning of the sample, a light microscopy image is taken. The geometry reconstructed from TEM data, can then be matched to the light microscopy image to verify the reconstruction (see Figure 2.6).

2.2 High Throughput Techniques

Until recently high throughput techniques were limited to light microscopy. Photons are easier to detect than electrons. In addition the sample preparation is not as complicated as for electron microscopy (EM) and can be automated with robots. But, the resolution of a light microscope is limited by the wavelength of light and cannot resolve important neuroanatomical structures like synapses or mitochondria. Thus, the field of neuroanatomy is facing the challenge of producing image data, which on one hand provides sufficient resolution to identify all important structures and on the other hand captures a field of view, which is large enough to contain a meaningful circuit of connec-

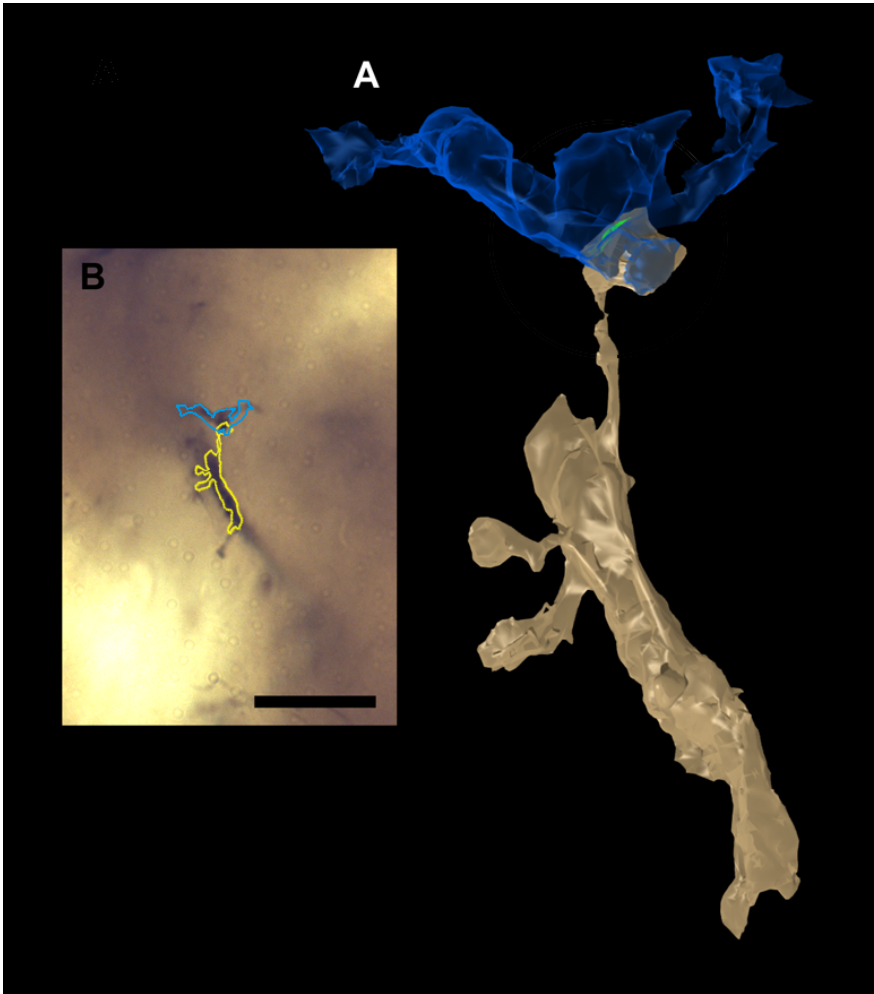


Figure 2.6: A: geometry reconstruction from TEM images, B: the same reconstruction superimposed on a light microscopy photograph. The scale bar corresponds to $10\mu m$ (Costa and Martin, 2010)

tions. Recent developments like automated microtomes, dual beam and block face microscopes, aim at transforming EM from qualitative hypothesis testing to a high throughput technique which renders quantitative evaluation of structures on a nanometer scale possible. The following section describes the most prominent high throughput techniques in the field of EM for neuroanatomy and discusses the advantages and challenges of the single techniques with respect to image processing.

High throughput sectioning: Cutting of ultra thin sections under 100 nm thickness is not only labor intensive, but also the main source for heavy artifacts like compression, knife marks, or cracks in the images. In addition, continued sectioning and collection of section ribbons on TEM grids causes problems such as folds, wrinkles or even loss of individual sections. The Automatic Tape-collecting Lathe ultramicrotome (ATLUM) automatizes this process by cutting a continuous tape from a cylindrical resin block. The process is similar to continuously peeling an apple. The tissue sample is embedded in the resin block such that each rotation of the microtome cuts away one section of the tissue block. A typical block may be 2mm wide, 20mm in circumference and can contain four $1mm^3$ tissue samples (Hayworth et al., 2006).

High throughput TEM: Image acquisition with the microscope can be automated by motorized stages which acquire large image stitchings without user interaction (Suloway et al., 2005). Bock et al. used a custom build Transmission Electron Microscopy Camera Array (TEMCA) to speed up the image recording in a TEM and imaged a $450 \times 350 \times 50$ micron volume of neuropil from 1,215 serial sections (Bock et al., 2010). Up to today, no other microscope technique is able to record such a large tissue volume with a resolution of $4nm$ per pixel. The challenges of TEM images for image alignment and segmentation are discussed in detail in Section 2. With respect to high throughput TEM data volumes up to multiple terabytes pose an

additional challenge to adapt the developed methods for TEM image analysis.

Focused Ion Beam Scanning Electron Microscope (FIBSEM): Unlike TEM, Scanning Electron Microscopes (SEM) do not require ultra thin sections for imaging. The image in a SEM is formed from electrons which are back scattered from below the surface of a specimen block. Regions of high electron dense stain cause great back scattering and lead to high image intensity. In a FIBSEM a focused ion beam is directed parallel to a block face to mill away thin layers of embedded tissue. A data volume is captured by sequentially imaging the block surface and then milling away a layer of tissue. Knott et al. (2008) used a FIBSEM to record a sub volume of neuronal tissue. From a milled surface of $30 \times 30 \mu m$ a $8.2 \times 7.1 \mu m$ field of view was selected and recorded continuously through the volume with $4 nm$ per pixel. Surprisingly the image acquisition time and not the milling time is the main limiting factor for recording large volumes with FIBSEM. Images are generated by the electron beam scanning the block surface and a balance needs to be reached between fast acquisition time, high resolution and signal to noise ratio. For the described data set each image was acquired in four minutes with a dwell time of $100 \mu s$ per pixel. With respect to image processing, FIBSEM data has the advantage of possible isotropic voxel size of $5 \times 5 \times 5 nm$. This resolution renders structure segmentation and identification an easier task than for TEM images. Due to the projective nature of TEM images, structures like patches of vesicles or non-orthogonal membranes appear blurry. In FIBSEM these structures can be recorded with well defined boundaries. In addition the high z-resolution allows to distinguish small 3d textures like balls (vesicles) and tubes (microtubuli) which appear only as rings (vesicles) vs. filled circles (microtubuli cut orthogonally) in TEM images. As FIBSEM captures the block surface before it is milled away, image alignment is restricted to compensating a translational drift caused by charging effects. Currently the main disadvantage of FIBSEM is the very limited field of view.

Serial Block-Face Scanning Electron Microscopy (SBFSEM):

The SBFSEM also uses the SEM technique for image acquisition. In contrast to FIBSEM, the block surface is not milled away after image acquisition. Instead, a microtome is integrated in the vacuum chamber of the microscope (Denk and Horstmann, 2004). Using a microtome renders large surface areas accessible, which could not be milled entirely with a focused ion beam. In principle the SBFSEM captures fields of view larger than possible with current TEM techniques (Knott et al., 2008). But, imaging the surface with a resolution suitable to see synapses means that significant energy is delivered to the block face, resulting in changes in the resin structure (i.e. cross-linking). The increased hardness of the embedding medium hinders continuous sectioning with the knife. With respect to image processing, SBFSEM like FIBSEM has the advantage that image alignment is not necessary. For image segmentation and structure classification, SBFSEM data provides a lateral resolution of about $10nm$ per pixel, which renders vesicle and synapse identification very challenging. A couple of publications in the field of image processing have investigated geometry reconstruction of neurons from SBFSEM data with an isotropic resolution of $25 \times 25 \times 25nm$ per voxel (Andres et al., 2008; Jain et al., 2007; Yang and Choe, 2009). But, the staining used for this data set concentrates on intracellular space, eliminating all structures and texture from the inner cell areas. Thus, the images are basically chemically segmented and the image processing task shifts from structure identification to the challenge of reconstructing the neuron geometry from low resolution images.

Chapter 3

Registration

To enable geometry extraction of neuronal processes from transmission electron microscopy imagery, the images have to be distortion corrected, stitched, and aligned across sections. We employ feature based registration to estimate the respective transformations.

The following chapter first introduces automatic landmark extraction based on SIFT features for electron microscopy images and then addresses the individual registration methods from stitching and distortion correction to non-linear warping with anomaly detection. For image stitching, feature based registration has the advantage, that correspondences can be extracted from overlapping image areas, without prior knowledge about the image arrangement. Our experiments demonstrate, that the distribution of the found correspondences is dense enough to reliably estimate a non-linear transformation to correct for lens distortions. For the affine alignment of adjacent sections, the rotation invariance of SIFT features is beneficial. However, due to the changing image content across sections, the landmarks are not dense enough to estimate a non-linear warping. Thus, the proposed warping method refines the initial feature based affine alignment based on image intensities.

3.1 Automatic Landmark Extraction with SIFT

Alignment of images defines an important part of the image processing pipeline for neuron geometry extraction from ssTEM images. Montaging of several images from one serial section as well as alignment of images from different sections requires to extract correspondence points between images. Important aspects of automatic landmark extraction are: (i) scale invariance to enable montaging of high resolution images on a lower resolution overview image, (ii) robustness against noise in the images (iii) robustness against non-linear transformations between images. As our experiments demonstrate, Scale Invariant Feature Transform (SIFT) descriptors of local image properties introduced by Lowe (2004) yield robust correspondence points with good localization. Alternative approaches for automatic landmark extraction include SURF (Bay et al., 2008) and DAISY (Tola et al., 2010).

Lowe’s keypoint extraction, consists of the following steps: (i) First, points of interest are detected in both images. These landmarks preferentially cluster around regions with high contrast. (ii) Then scale and rotation invariant feature descriptors based on local gradient information are calculated for each point. (iii) In the next step, correspondences are detected according to Euclidean distance in feature space: For each point $x^{(i)}$ in image i , a correspondence point $x^{(j)}$ in image j is identified that is the nearest neighbor of $x^{(i)}$ in feature space. If this nearest neighbor point is significantly closer to $x^{(i)}$ than all other points of image j in feature space, then the points $x^{(i)}$ and $x^{(j)}$ are marked as correspondent.

Scale invariance of SIFT features enables to match correspondences between images taken with different magnification settings. Figure 3.1 shows an example of correspondences found automatically between an overview image, recorded with a magnification of 3400x and a high resolution image recorded at a magnification of 25000x.

Cryo TEM images are challenging for correspondence point extraction due to low contrast, noise, and preparation artifacts. Figure

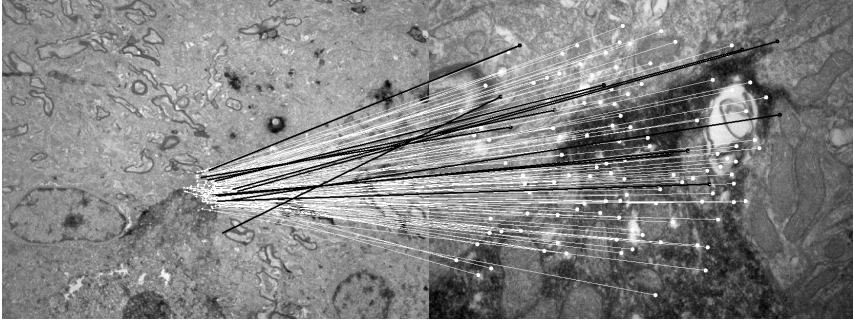


Figure 3.1: Automatically extracted landmarks between images taken at different magnifications. White: Matching correspondences, Black: Outlier correspondences

3.2 depicts correspondences automatically extracted from two cryo TEM images. The left image shows a cryo section containing a plasma membrane, vesicles and a mitochondrion in the lower right corner. The image on the right shows the same section with a translated field of view. Preparation artifacts in the form of compression waves and knife marks are clearly visible in both images. Despite the low contrast, the noise and the preparation artifacts, 14 correct correspondences are automatically found in the area with structural information. The two false positive matches can be filtered out automatically based on large affine alignment error.

In the context of lens distortion correction or alignment of consecutive sections, it is important, that the automatically extracted correspondences are robust with respect to non-linear transformations. In order to demonstrate the reliable performance of SIFT features in this context we applied a highly non-linear transform to a neuroanatomical TEM image. The results are shown in Figure 3.3. The original image (a) and the transformed image (b) were normalized to a gray value range from zero to one and white noise with a standard deviation of 0.1 was added to both images. Subfigure (c) depicts correspondences

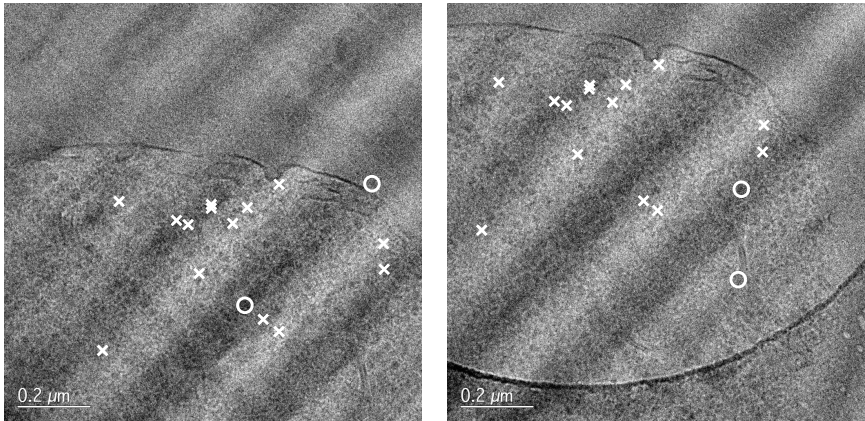


Figure 3.2: Correspondences extracted from two translated cryo TEM images with low contrast and preparation artifacts in form of compression waves and knife marks. The sample contains a plasma membrane with an ice area above and structural information like vesicles and a mitochondrion below. The landmarks are concentrated on the area with textural information and neglect the ice part in the upper part of the images. Out of the 16 correspondences found, only two are outliers (marked with a circle).

found in the images by a line for each point pair. The line connects the coordinates of the point from the original image with the coordinates of the point from the transformed image when plotted into the same coordinate system. As can be seen, the whole image range is densely covered with correspondences. Furthermore, the needle diagram of the estimated correspondence connections forms a smooth field which follows the ground truth transformation (d). Subfigure (e) and (f) provide a quantitative evaluation of the accuracy of the landmarks. The histograms show the distance between the landmarks found in the transformed image and the known ground truth position of the landmarks. Without noise present in the images the median distance is 0.54 pixels demonstrating that SIFT features allow for automatic correspondences with sub-pixel precision for large non-linear transformations. With noise added to the images, the correspondence points are found on a coarser scale level of the feature space than for the noise free case. As a consequence the number of correspondences decreases and the median distance to the ground truth precision increases to 1.46 pixels. As conventional TEM or SEM images typically contain less noise than the noise level chosen for this experiment, this result can be seen as a lower bound on the accuracy of the automatically detected landmarks.

As the examples listed above demonstrate, automatic correspondence point extraction via SIFT features yields robust results with respect to scale, noise and non-linear transformations.

3.2 Stitching

Frequently, microscopists are interested in analyzing regions of a sample which are too large to be captured by a single image at a sufficient magnification. Therefore, it is common experimental practice to capture several translated images of such a large region of interest and to assemble these images afterwards to cover the whole area under investigation. This process is commonly referred to as stitching, mosaicing or montaging. Montaging of images taken from one section is

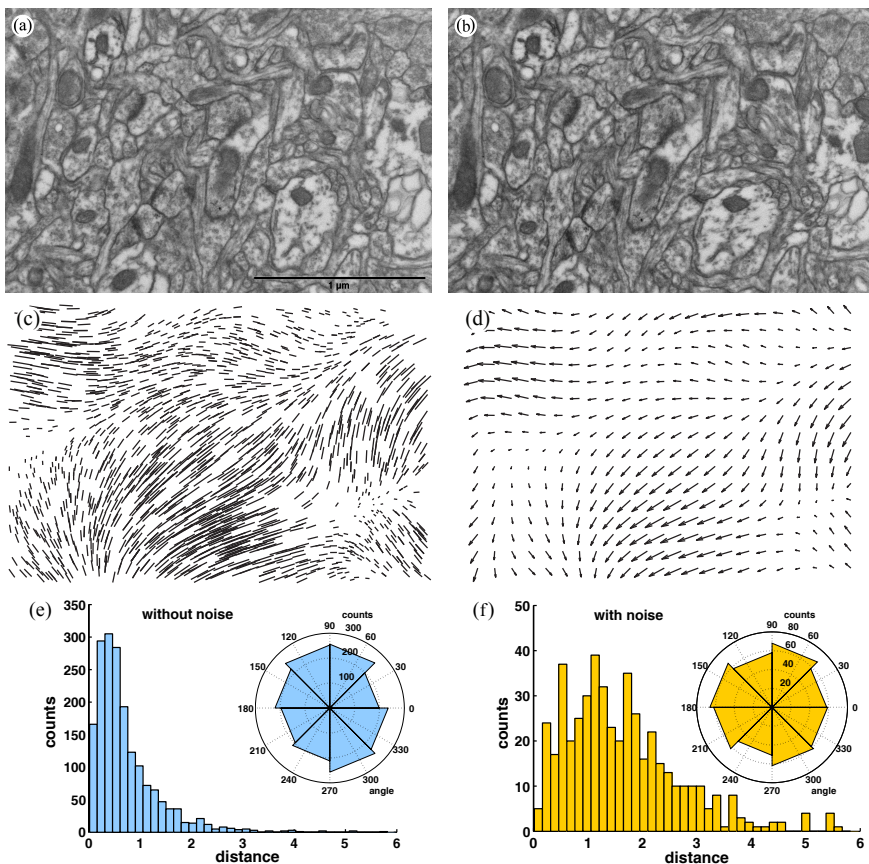


Figure 3.3: SIFT correspondences for a highly non-linear deformation and considerable noise. (a) original image with noise, (b) transformed image with noise, (c) connection between automatically obtain correspondences plotted in one coordinate system, (d) ground truth transformation, (e) and (f) histograms of distances and orientation error between found landmarks and ground truth without noise (e) and with noise (f).

widely done by maximizing the normalized cross correlation of overlapping image areas. Typically, images are assumed to be arranged in a regular grid with overlapping areas of 10% - 20% of the image size.

When stitching two images together, two conditions have to be fulfilled:

1. a number of reliably recognizable contrast patterns have to be identified and captured by both images;
2. the images have to be free of distortions, e.g. perspective distortions or vignetting.

In this chapter we introduce two approaches to the stitching problem. First, we solve the stitching based on overlapping areas by feature based registration, instead of intensity based cross correlation methods. In the next subsection, the problem is extended to stitching of images with multiple resolutions.

3.2.1 Robust Stitching of Overlapping Regions

The problem of image montaging based on overlapping areas is typically solved by finding the affine transformation that maximizes the cross correlation between the overlapping areas. Computation of cross correlation in the Fourier Domain is very fast, rendering this approach attractive for user interaction. However, in practice maximizing the cross correlation often yields undesirable matchings, forcing the user to correct the alignment manually. When montaging images it is important that the stitching border yields good matches for edges in the image. Thus, the optimization criterion should depend on the image gradient instead of intensity values. One way to solve this issue is to calculate the cross correlation of the gradient image instead of the intensity values. However, we chose to employ a feature based approach based on SIFT keypoint detection instead, for the following reasons:

- SIFT is based on the image gradients

- keypoints are detected at well localized positions with sub-pixel accuracy leading to precise landmarks
- correspondence point extraction based on SIFT features is very robust against non-linear distortions and noise (see Figure 3.2)

Stitching of images based on automatically extracted landmarks is also used in the field of photogrammetry. Lingua et al. (2009) evaluated the performance of SIFT correspondences in this field and demonstrate good matching results for large geometric and photometric distortions.

Assuming that the microscope does not cause non-linear distortions, all images can be stitched together by estimating an affine transformation $A^{(i)}$ for each image i . The transformation maps each image coordinate system to the common stitching coordinate system. The mapping is optimal, if the distances between correspondence points in the stitching coordinate system are minimal.

This translates to the following least squares problem:

$$\min_A \sum_{\substack{i,j=1, \\ j \neq i}}^B \left\| X^{(i,j)} A^{(i)} - X^{(j,i)} A^{(j)} \right\|^2 \quad (3.1)$$

The coordinates $X^{(i,j)}$ of the correspondence points between image i and image j are represented as homogeneous coordinates, transforming the 2D-coordinate (x, y) to a 3D-coordinate $(x, y, 1)$ by adding a one as the constant third dimension. The affine transformations $A^{(i)}$ are then represented as 3×3 transformation matrices. B denotes the total number of images to be stitched together.

Typically, one affine transformation is set to identity to define the coordinate system of the stitching image. To estimate the remaining affine transformations we define $X^{(i)}$ as the set of all correspondence points in image i , i.e., $X^{(i)} = \bigcup_{j=1}^B X^{(i,j)}$ and $Y^{(i)}$ as the set of all corresponding points in the stitching coordinate system, i.e., $Y^{(i)} = \bigcup_{j=1}^B X^{(j,i)} A^{(j)}$

The minimization problem (3.1) can now be solved by pseudo-inverse method

$$A^{(i)} = \left(X^{(i)T} \cdot X^{(i)} \right)^{-1} \cdot X^{(i)T} \cdot Y^{(i)}. \quad (3.2)$$

As $Y^{(i)}$ depends on the affine transformations estimated, the problem is solved iteratively, alternating between updating the affine transformations $A^{(i)}$ and the mapped correspondence coordinates $Y^{(i)}$.

The solution described so far weights the distance ξ between two correspondent points quadratically. But, SIFT features may lead to some false correspondence points, because points are only compared based on their feature vectors. Similarities in the structure or repeated patterns can cause incorrect correspondences. Thus, it is beneficial to use robust estimation methods, like Huber loss or RANSAC to estimate the affine transformations (Szeliski, 2004; Kaynig et al., 2007; Saalfeld and Tomančák, 2008).

Applying Huber-loss, one replaces the squared loss function $L(\xi) = \xi^2$ by the Huber-loss function which only gives linear weight to large errors:

$$L_c(\xi) = \begin{cases} c|\xi| - \frac{c^2}{2} & , \text{ for } |\xi| > c \\ \frac{\xi^2}{2} & , \text{ for } |\xi| \leq c. \end{cases} \quad (3.3)$$

In Equation (3.2) the Huber loss introduces a diagonal weight matrix Ω which reduces the influence of outliers on the estimated transformation accordingly.

$$A^{(i)} = \left(X^{(i)T} \cdot \Omega \cdot X^{(i)} \right)^{-1} \cdot X^{(i)T} \cdot \Omega \cdot Y^{(i)} \quad (3.4)$$

The assigned weights Ω depend on the residuals ξ and thus on the estimated transformation $A^{(i)}$. Therefore, the optimization iterates between estimating the transformation matrix and updating the weight matrix accordingly.

3.2.2 Multiple Resolutions

So far we have only considered stitching of images with the same resolution, based on overlapping image areas. The stitching gets more reliable, with increasing size of the overlapping image area, but large overlaps lead to a lot of redundant data acquisition. However, SIFT features are scale invariant and thus enable automatic landmark detection between images with different resolutions. Thus we propose to employ an overview image at a coarse resolution as a template for the stitching. This approach has several advantages:

- Reliability of the image montage does not depend on the size of overlapping areas between high resolution images.
- Acquisition of an overview image is less effort for the microscopist than providing sufficient overlap for all high resolution images.
- The context between regions of interest with high resolution is preserved in the overview image, e.g. allowing for measurements between the areas.

Figure 3.4 depicts an example stitching of high resolution images on an overview image. High resolution images are not required to be arranged in a grid like structure and the context between regions of interest is preserved in the low resolution overview image. In principle the approach can be easily extended to stitchings of multi resolution pyramids of images.

In addition, if non-linear transformations like lens distortions are present in the images, and no calibration data is available, the high resolution images can be stitched non-linearly to the overview image. This way, a seamless stitching can be achieved, with the smallest artifact deformation possible.

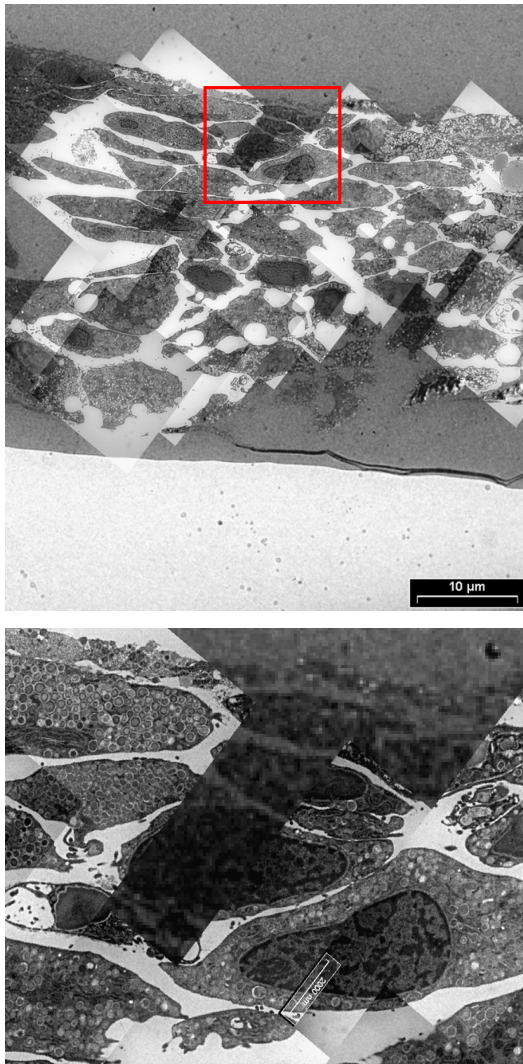


Figure 3.4: Upper image: Example stitching of high resolution images on a low resolution overview image. Lower image: region of interest marked by a rectangle in the upper image.

3.3 Distortion Correction

The goal of stitching is to assemble a set of images that are capturing a larger part of the sample and that have sufficient overlap (Figure 3.6). When the images are distorted by the electromagnetic lenses of the microscope, it is not possible to find an affine transformation that stitches the images together without major errors. Figure 3.5 shows a zoom in two different regions with large stitching errors by an affine transformation. The stitching boundary can easily be detected in both images. A non-linear correction is required to stitch the images neatly together. Here we introduce a new method to estimate this non-linear transformation that can be used to correct the images.

When image distortions occur in the imaging process, it is not only difficult to seamlessly stitch the images together in a reliable way. Distortions also alter the shape of structures in the images and thus are a possible source of measurement errors. Typically, distortion fields are growing with increasing distance to the image center. This observation suggests, that the distortion problem is less severe for CCD-cameras with a small number of pixels than for image capturing devices with a large field of view despite the increasing demand for stitching. Using films, film plates or the latest generation of 10 mega pixel large-size CCD-cameras, the image distortions can severely perturb reliable stitching of large field of view images. The following chapter introduces a fully automatic calibration and stitching approach, that can correct for non-linear distortions in the images caused by the electromagnetic lenses, while preserving structural information of the images.

Figure 3.6 illustrates the stitching scenario with distortions. The same region of the specimen is captured by two different images. The microscope induces the same distortion to all images. But, the original structure of the specimen is recorded in different areas of the local image coordinate system and thus the structure is distorted differently in both images. Our approach uses the information of the overlapping image areas to estimate a non-linear distortion correction

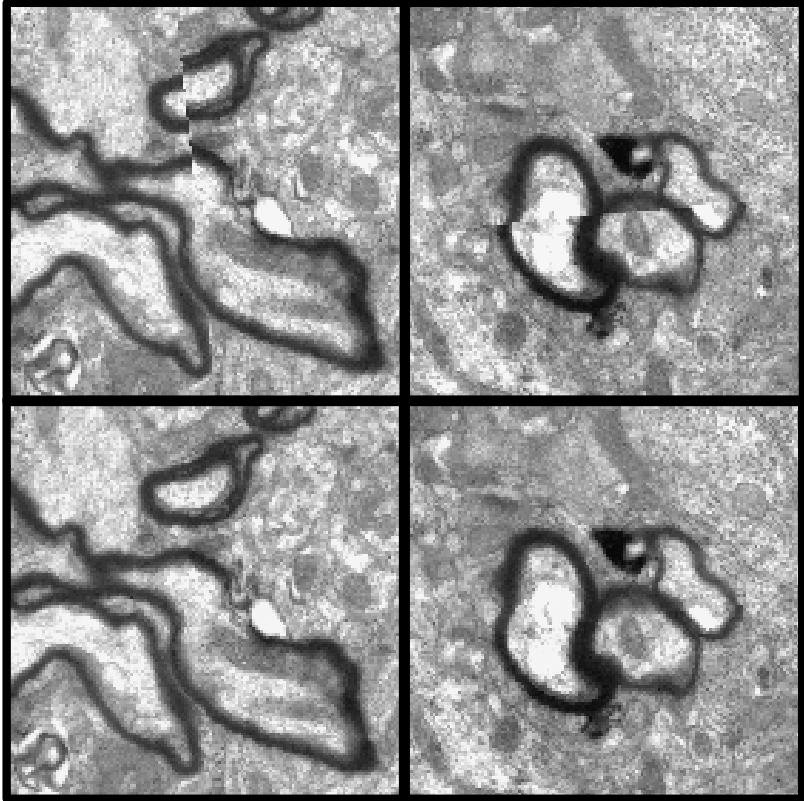


Figure 3.5: Two example regions of the stitching intersection ($2.7 \times 2.7 \mu m$). In the top line without distortion correction the image border is clearly visible. In the lower line the distortion correction produces a seamless stitching.

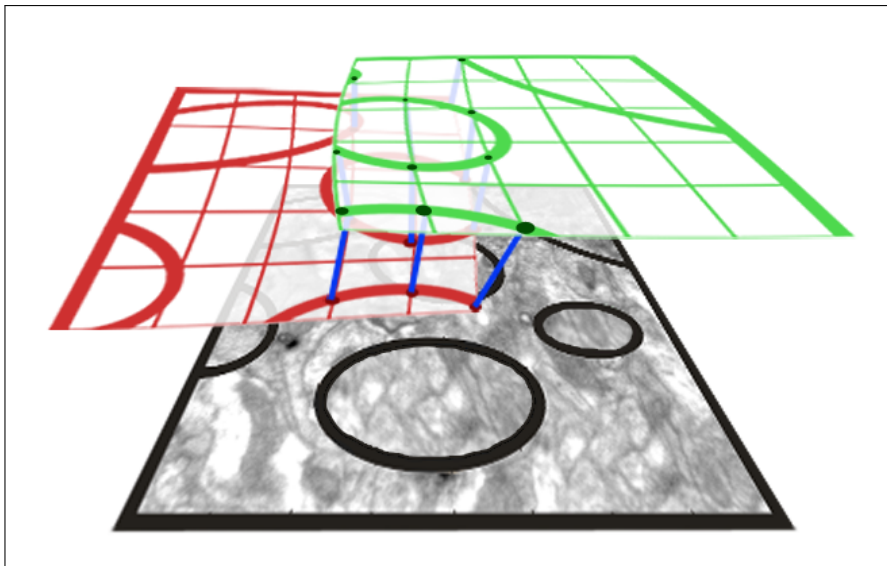


Figure 3.6: Illustration of the stitching and calibration scenario. The same region of the specimen is covered by the overlapping image areas. Due to non-linear distortions in the images, overlapping regions cannot be matched with an affine transformation alone. The proposed self calibration method uses the redundant information in the overlapping areas to estimate the non-linear transformation field that corrects the distortion.

that preserves the structural geometry of the images. For a series of images taken with the same parameter settings of the microscope, the distortion correction field has to be estimated only once in the beginning and can then afterwards be applied to all images. The method uses redundant information in overlapping images to estimate the non-linear transformation. Thus, no special calibration samples are required.

The approach has been implemented as an open source plugin for the **ImageJ** (Rasband, 2010) distribution **Fiji** (Fiji, 2010), is freely available and also has been integrated into **TrakEM2** (Cardona, 2006). More information about the plugin is on the website:

http://pacific.mpi-cbg.de/wiki/index.php/Distortion_Correction

3.3.1 Related Work

Lens calibration is commonly achieved by use of special calibration images, like pictures of checkerboards or other objects with straight lines (Devernay and Faugeras, 2001; Gremban et al., 1988; Zhang, 2000) in order to reproducibly correct the images after their acquisition. However, this method is not applicable to electron microscopy. While there are crystalline structures that might be used as calibration samples (i.e. a non-scale checker board), it is possible that the whole process of removing the calibration sample from the microscope, inserting the actual sample of interest and refocusing on that sample again changes the distortion field. This often uncontrollable variability is caused by the electromagnetic lenses, which do not correspond to fixed shaped glass lenses of camera objectives. Instead they change shape according to different configurations of the microscope parameters. Due to hysteresis effects, there is no clear correspondence between the parameter settings and the induced distortion.

Other methods often incorporate a specific distortion model, like radial symmetric “barrel” or “pincushion” distortions (Wang et al., 2009; Shih-Schon and Bajcsy, 2001; Hartley and Kang, 2007). While in principle, these distortions may affect imaging in electron microscopes, we prefer to not restrict our distortion model to specific shapes

as done in (Sawhney and Kumar, 1999; Stein, 1997). This generality enables superior correction in case of non-ideal electromagnetic fields in the microscope column. To our knowledge the only alternative approach to calibrate electron microscopes is described in (Koshevoy et al., 2006). Koshevoy et al. parametrize lens distortions by Legendre polynomials where intensity variance is used as a similarity measure. Our new proposed method distinguishes itself in three major points from the method by Koshevoy et al.: First, SIFT features are used to measure similarity between image patches/corners instead of intensity. This choice renders the approach more robust and therefore better applicable to handle noisy electron microscopy images than the alternative distortion correction. Second, our image assembly relies on a grid like matching, where we maximize the similarity between all overlapping pairs jointly. Instead, Koshevoy et al. only estimate a cascade of transformations over a predefined hierarchical order of the images. Due to the fact that errors are propagated over the image cascade, large errors can occur at the end nodes of such a cascade as a consequence of error propagation. Third, we only correct a non-linear distortion field caused by the electron microscope that is shared by all images. Koshevoy estimates an additional non-linear transformation independently for each image. This procedure renders the approach problematic, as the non-linear correction induces structural changes to the images. Our method instead corrects only for the distortion field that is the same in all images. Marsh (2007) automatically extracts correspondences in tomograms by cross correlation of sub-volumes stating that, manual interaction is necessary to exclude false positive matches even for a large support of $80 \times 80 \times 70$ voxels. SIFT features instead yield robust matches and are typically computed at an interpolated position with sub-pixel accuracy, which significantly outperforms localization based on cross correlation. Lawrence et al. (2006) describe non-linear lens distortions as an issue for three dimensional tomogram reconstructions and account for these distortions by extending the linear reconstruction method to a curvilinear model of cubic order. The distortion model described in this paper enables the

correction of non-linear transformations of higher than cubic order while simultaneously preserving structural information.

3.3.2 Origins of Distortions

The most faithful and efficient way to remove distortions is by hardware adjustments, e.g. by changes of the electron optical conditions. A hardware correction would restore the signal before analog to digital conversion and, therefore, it would recover the signal with the smallest possible loss. Such a strategy succeeds e.g. for light microscopes, where objective lens systems are built to minimize distortions. In electron microscopy, however, there exist several obstacles that render hardware controlled distortion corrections difficult. First, the distortions can sensitively depend on the electron optical configuration selected for image acquisition. This dependency implies that the correction would need to be adaptable to different parameter settings. Second, the distortion parameters are commonly specific for the respective microscope. Third, little information has been reported in the literature about typical characteristics of imaging distortions induced by electron microscopes. In Figure 3.7, a sketch of a transmission electron microscope (TEM) is shown. The main electron optical components of a TEM to be identified as possible sources for distortions are the condenser lens system, the objective lens and the projective lens system. The condenser lenses define the beam incident on the specimen, i.e. an almost parallel beam (cross-over far after the specimen), a focused beam (cross-over on the specimen) or an over-focused beam (beam cross-over before the specimen). The objective lens is located directly below the sample. This lens has short focal length and gives rise to a first magnified image. The first intermediate image plane is further magnified by the projective lens system, which consists of all lenses below the objective lens, e.g. diffraction, intermediate and projective lenses. Finally, the magnified image is captured by the camera or a viewing screen. The objective lens is well known for giving rise to distortions, because it suffers like every rotationally symmetric electromagnetic lens from spherical aberration. This effect

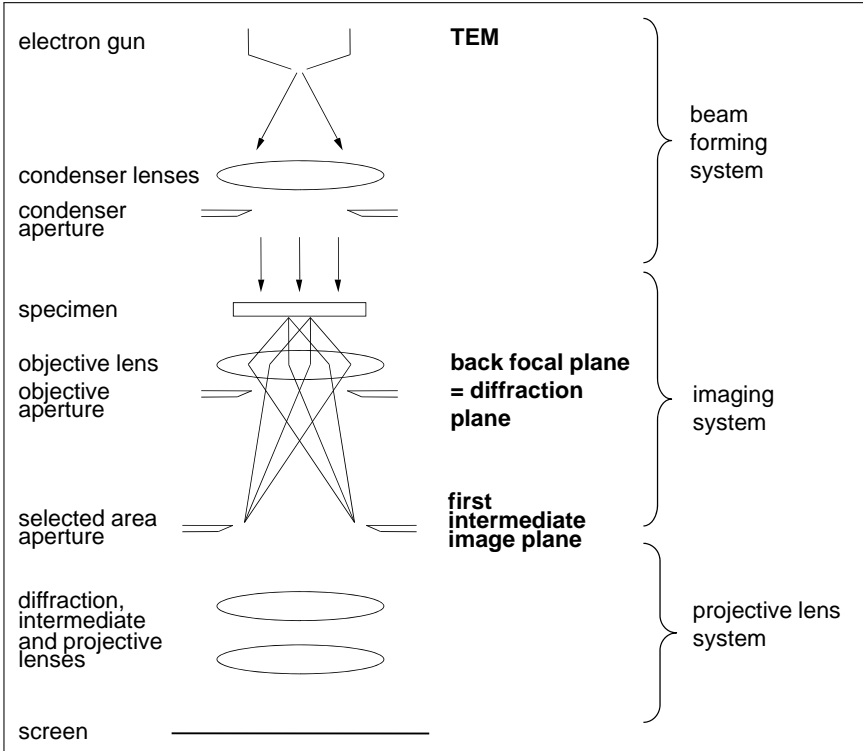


Figure 3.7: Sketch of the basic components of a transmission electron microscope. Components influencing the distortions are the condenser lens system, the objective lens and the projective lens system.

mainly leads to a blurring of the images being equivalent to a loss of resolution. The spherical aberration exhibits its main degrading impact in high-resolution imaging. The projective lens system so far has not been thoroughly analyzed in the literature with respect to distortions. The large variety of magnifications accessible by the projective lens system is achieved by a combination of two effects. The strength of the lens current can be controlled for each individual lens separately, resulting in a change of the focal length of the respective lens. The experimenter can also select different combinations of the set of lenses below the objective lens. As charged particles move on a helical trajectory in an electromagnetic field, a change of the focal length of a lens not only causes the desired change in magnification, but it also induces a comparably small image rotation, while a different combination of the lenses can generate a large jump in image rotation. The condenser lens, the objective lens and the projective lens systems may all suffer from instabilities caused by hysteresis effects when changing the focal length of electromagnetic lenses.

3.3.3 Estimation Framework

Stitching of images can be seen as an inversion of the coordinate transforms which are applied during the image acquisition process. The diagram in Figure 3.8 illustrates the two main coordinate transformations which we consider in the stitching approach presented in this paper. The sample itself defines the original object coordinate system. If an image is taken from the sample, a region of interest from the original object coordinate system is mapped to its own image coordinate system. The transformation between object coordinate system and image coordinate system is affine with a translation defined by the positioning of the field of view of the microscope, a rotation caused by the spiral movement of the electrons inside the microscope and a scaling according to the magnification settings. The lens distortion of the microscope now acts on the intermediate coordinate system and hence is the same for all images. Thus, for example the upper left corner of each image is distorted in the same way for all

images, regardless of the absolute positioning of the region shown in the image in the original object coordinate system of the sample. To stitch a set of single images back to a common coordinate system, the two transformations applied to the images beforehand have to be inverted. Hence, first a non-linear transformation is applied to correct for the lens distortion and then the images are mapped with an affine transformation to the coordinate system of the stitched image. The coordinate system of the stitched image and the original object coordinate system of the sample are only equivalent up to an affine transform because the relative positioning of the sample in the object coordinate system is lost when the sample signal is mapped to the intermediate coordinate system (see Figure 3.8). A common practice is to define one of the image coordinate systems as the stitching coordinate system and then map all other images to this coordinate frame.

A non-linear transformation is necessary to correct the images against distortions. To model the non-linear transformation we use an explicit polynomial kernel expansion to map the correspondence points into high dimensions and then estimate the transformation matrix α that projects the points back to the 2D image plane.

The polynomial expansion of degree d for a 2D point x with coordinates (u, v) yields

$$\phi_d(u, v) = \left(1, u, v, u^2, uv, v^2, \dots, v^d\right). \quad (3.5)$$

For a $n \times 2$ point matrix X the kernel expansion $\phi_d(X)$ denotes the $n \times \frac{(d+1)(d+2)}{2}$ -matrix where the kernel expansion is applied to all n rows separately. In all our experiments $d = 5$ provided sufficient degree of freedom to estimate the non-linear distortion correction. As little is known about the characteristics of distortions induced by electromagnetic lenses, the polynomial kernel used for our experiments does not restrict the estimated transformation to correct for particular models like barrel or pincushion distortions. However, the method described in this paper can easily be restricted to specific transformation models by choosing another kernel function ϕ , and lifting the

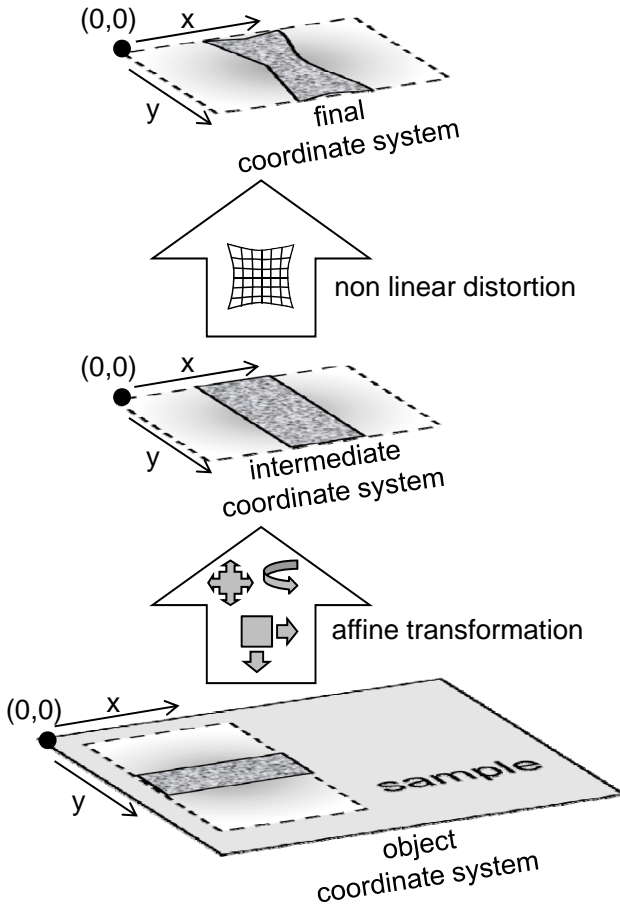


Figure 3.8: Illustration of the image acquisition process. First a region of interest from the object coordinate system is chosen and mapped with an affine transformation (translation, rotation, scaling) to its own intermediate coordinate system. The lens distortion then transforms this image further to the final image coordinate system.

points e.g. to a circular feature space (Geyer and Daniilidis, 2001; Claus and Fitzgibbon, 2005).

The non-linear transformation of a set of points $X^{(i)}$ is written as a matrix multiplication $\phi_d(X^{(i)})\alpha$ with a $\frac{(d+1)(d+2)}{2} \times 2$ transformation matrix α which projects the kernel expanded points back into 2D space. For homogeneous coordinates a third column is added to α which maps the third coordinate of the transformed points to a constant.

The joint calibration and stitching is formulated as a minimization problem which directly follows the inverse transformations of Figure 3.8: for each correspondence point the non-linear transformation is applied to its local image coordinates in image i and image j . After correcting the distortions, the correspondences are mapped by an affine transformation to a common coordinate system corresponding to the mosaic image. For the set of all images B , the squared Euclidean distance between correspondence points should be minimal in the coordinate system of the mosaic image. This goal induces the following optimization problem:

$$\min_{\alpha, A} \sum_{\substack{i, j=1, \\ j \neq i}}^B \left(\left\| \left(\phi_d(X^{(i,j)})\alpha A^{(i)} \right) - \left(\phi_d(X^{(j,i)})\alpha A^{(j)} \right) \right\|^2 + \lambda \left\| \phi_d(X^{(i,j)})\alpha - X^{(i,j)} \right\|^2 \right). \quad (3.6)$$

Using homogeneous coordinates, the 3×3 transformation matrix $A^{(i)}$ for each image is estimated according to Eq. (3.4). The regularization term in equation (3.6) penalizes transformations which map the transformed image points far apart from the original image points. Experiments clearly demonstrate that $\lambda = 0.01$ is a sufficient weight for the regularization term. By explicitly regularizing the In practice the objective can be optimized by iteratively obtaining the affine transformations $A^{(i)}$ and then the non-linear transformation matrix α . Our experiments demonstrate, that very few iterations are sufficient to obtain α and $A^{(i)}$.

When keeping all $A^{(i)}$ fixed, the solution for α is unique and is obtained by setting the derivative of Eq. (3.6) with respect to α to zero and then solving for α . This yields

$$\begin{aligned} \text{vec}(\alpha) = & \left[\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(2 \cdot A^{(i)} A^{(i)T} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right. \right. \\ & \left. \left. - 2 \cdot A^{(j)} A^{(i)T} \otimes \tilde{x}_n^{(j,i)T} \tilde{x}_n^{(i,j)} + \lambda (I_{3 \times 3} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)}) \right) \right]^{-1} \\ & \text{vec} \left(\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(+\lambda \cdot x_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right) \right) \end{aligned} \quad (3.7)$$

where $\tilde{x}^{(i,j)} = \phi_d(x^{(i,j)})$ and $x_n^{(i,j)}$ denotes the coordinates of the n th correspondence point between image i and image j in the coordinate system of image i . The index n of the second sum runs over all $N(i, j)$ correspondence points between image i and image j .

The operator $\text{vec}(\alpha)$ applied to a matrix α concatenates the columns to a vector. The operation $Y \otimes Z$ denotes the Kronecker product of an $m \times n$ matrix Y and an $r \times q$ matrix Z that yields a $mr \times nq$ matrix, defined as:

$$Y \otimes Z = \begin{pmatrix} Y_{11}Z & Y_{12}Z & \dots & Y_{1n}Z \\ Y_{21}Z & Y_{22}Z & \dots & Y_{2n}Z \\ \vdots & & & \vdots \\ Y_{m1}Z & Y_{m2}Z & \dots & Y_{mn}Z \end{pmatrix} \quad (3.8)$$

3.3.4 Experimental Validation

In this section we experimentally evaluate our calibration approach. Simulated data are employed to test the quality of the correction against known ground truth. In addition we show on real electron

microscopy data, that our calibration method is able to reduce distortions with an average uncorrected stitching error larger than 10 pixels to sub-pixel precision. The experiments are designed to focus on different lens systems in order to determine the origin of the distortions in a TEM and to provide guidelines how often a new calibration has to be estimated.

The distortion correction estimation does not require images from special calibration samples. The only limitation is, that the images have enough contrast and texture to find correspondence points with SIFT features and that sufficient image overlap is provided. We propose to take one initial calibration set of nine images arranged in a 3x3 grid with a vertical and horizontal overlap of at least 50% per image in both directions. The large overlap region ensures that all areas of the image, including the center, contribute to the estimated distortion correction. Afterwards the obtained correction transform can be applied to any image taken under equivalent electron optical conditions.

3.3.5 Simulated Ground Truth Data

To test our calibration method with available ground truth, a set of 9 calibration images (1603x1069 pixels) has been warped by using the `SplineDeformationGenerator ImageJ` plugin (Arganda-Carreras et al., 2006). Figure 3.9 summarizes the results after the images have been unwarped by our method. Subfigure (a) shows one of the nine original images. The distortion applied to the images can be seen in Subfigure (b). To demonstrate the stitching quality we show an image overlay of the example image and another one with 50% overlap. The uncorrected stitching (c) appears blurred in the overlapping area, as the stitching could not seamlessly match image features together. After the correction (d) the images show a satisfactory correspondence. Subfigures (e) and (f) depict the same situation as inverted difference images instead of overlays. Here, a dark pixel in the overlapping area demonstrates an error in the stitching. The good quality of the

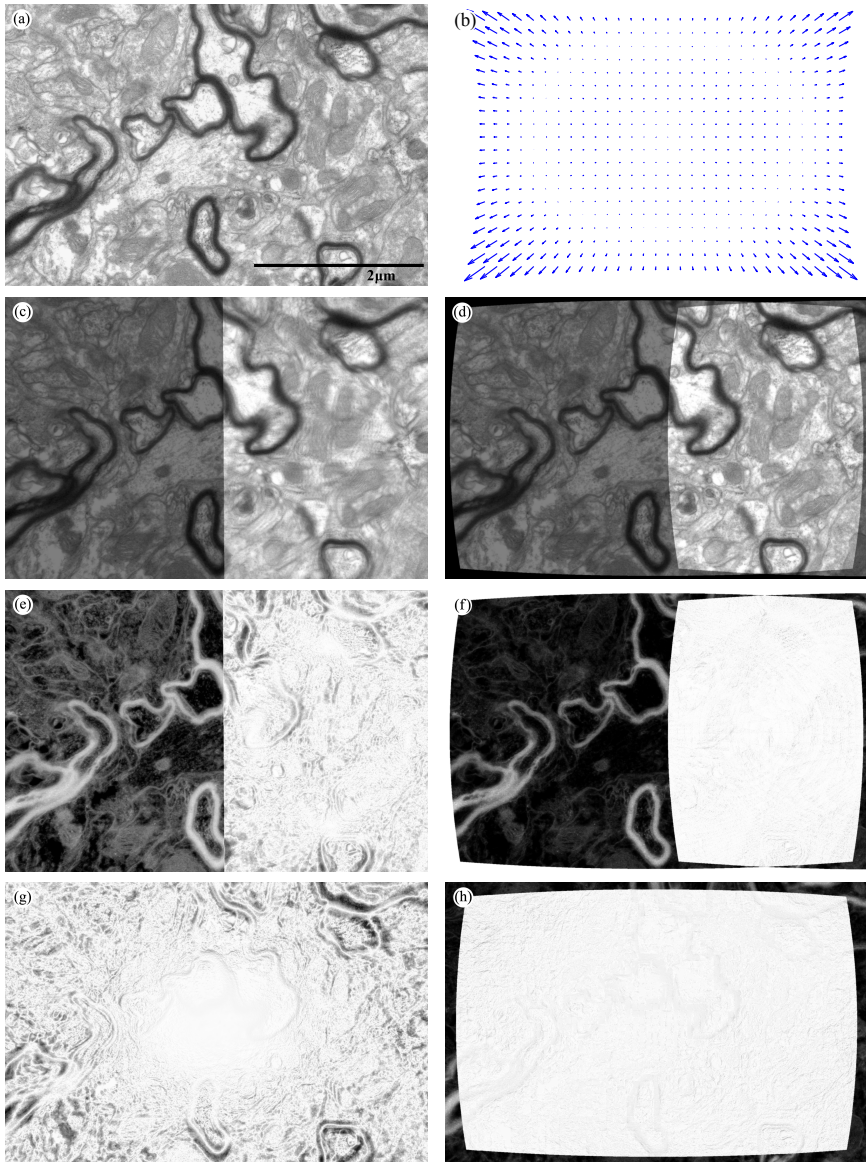
calibration can also be seen by the large reduction of differences between the original and the distorted image (g) when the correction is applied (h). The distortion is removed and the original structure of the image has been recovered. In the distorted images, the average matching error is 6.88 pixels. After the correction had been applied, the error was reduced to 0.41 pixels.

Our method is limited by isotropic scaling since it cannot correct for anisotropic scaling induced by the distortions without knowledge of the original size of image structures. Therefore the regularization term in Eq. (3.6) ensures that the applied correction changes the scaling factor as little as possible.

3.3.6 Electron Microscopy Data Experiments

To test our new method on real world data, we performed a series of experiments. All images (4008x2672 pixels) were acquired with a Philips CM100 equipped with a side-mounted Morada TEM camera from Olympus. The microscope was operated at 100 kV. We used a commercial cross-grating sample, which is a ruled diffraction grating with a periodicity of about 460 nm along two perpendicular axis. The grating is shadowed with gold which yields high contrast in a transmission electron microscope. The grating sample is stable under the electron beam and it is not expected to suffer from severe charging effects. As demonstrated before, our method does not require special

Figure 3.9 (*following page*): Evaluation on ground truth data: (a) original image, (b) applied barrel distortion; (c) stitching overlay of the distorted images, (d) stitching after distortion correction; (e) and (f) difference images for the stitchings shown in the row above; (g) difference image of the original and the distorted image, (h) difference image of the original and the corrected image. Difference images are shown with inverted contrast to enhance visibility and sub figure (f) and (h) are rescaled quadratically.



calibration samples, but the grid structure of the cross-grating sample enhances the visual impression of the distortion and the correction. The experiments have been designed to answer two questions: (i) The experiments are conducted to measure the reduction of the average stitching error by the distortion correction. (ii) The experiments should allow us to gain a first insight into possible causes and mechanisms of the distortions. In order to distinguish distortions caused by the different lens systems of the TEM, images were captured under the following experimental conditions:

1. condenser lens in over- and under-focus, i.e. crossover of the beam before and after the specimen
2. objective lens in focus and in strong under-focus ($20\mu m$), i.e. severely different focal lengths
3. sample not at correct height, compensation for the correspondingly wrong position of the image by the focal length of the objective lens (z-height of specimen too low, objective at $20\mu m$ over-focus)
4. change of magnification leaving all other parameters unchanged (3'400, 19'000, 25'000 and 64'000 times magnification)
5. specimen shifted sideways by some distance, leaving all electron optical parameters unchanged
6. all settings listed have been checked before and after a full alignment of the microscope.

The following information is expected from these settings: the first setup tests the influence of the condenser lens system; conditions two and three probe the dependency of distortions on the objective lens; test four identifies the influence of the projective system; the setting five is expected to yield information about effects due to the specimen. Procedure six, finally, is performed to measure how sensitively

iter	0	1	2	3	4
med	10.0708	1.2921	0.6438	0.5968	0.5941
stddev	1.9843	0.4875	0.1971	0.1637	0.1599

Table 3.1: The table contains the median stitching error (med) and its standard deviation (stddev) over all 28 calibration sets. The error is rapidly reduced with very few iteration steps (iter)

distortions depend on the electron optical parameters. This information is important to judge how often a new distortion correction has to be estimated.

A general analysis over all experiments enables us to measure the performance of the distortion correction. Then the influence of the single parameter settings on the distortion is discussed in relation to the experimental conditions. Figure 3.10 shows the results summarized over all experiments in a box plot. The first box, corresponding to iteration step 0, depicts the original uncorrected error in the stitching. The large variance of this entry is due to the different parameter settings that cause different distortions. But, already the second iteration step corrects the distortion sufficient enough to reach sub-pixel accuracy. The small variance in the corrected result demonstrates, that the method performs reliably for all parameter settings. The numerical values of the median stitching error and the standard deviation for the single iteration steps are summarized in Table 3.1.

A visual impression of a distortion and the correction result is depicted in Figure 3.11, showing the lower left region of a cross-grating sample image. When the grid structure is compared to the straight lines drawn in blue, it is clearly visible, that the upper image contains distortions. In the lower left corner the distortions exceed the grid distance of 460 nm. In the bottom image, the correction has been applied and the geometry of the sample is restored.

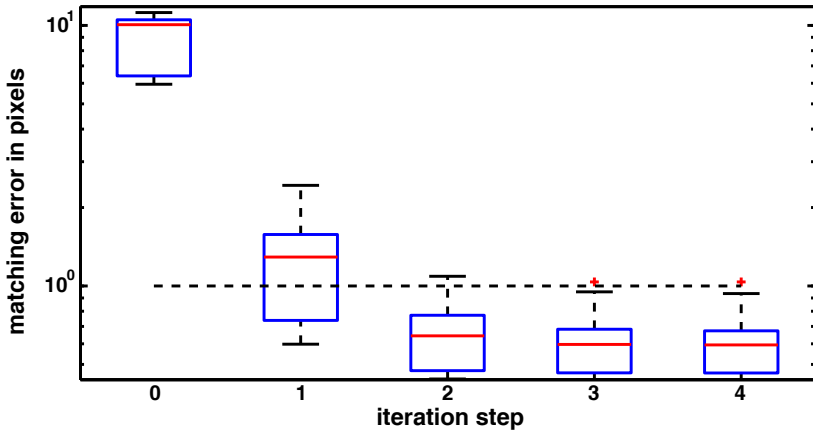


Figure 3.10: Box plot showing the convergence of the algorithm for 28 different sets of calibration images. On each box, the red line marks the median stitching error in pixels (log scale), the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers. The dashed black line marks the 1 pixel error boundary.

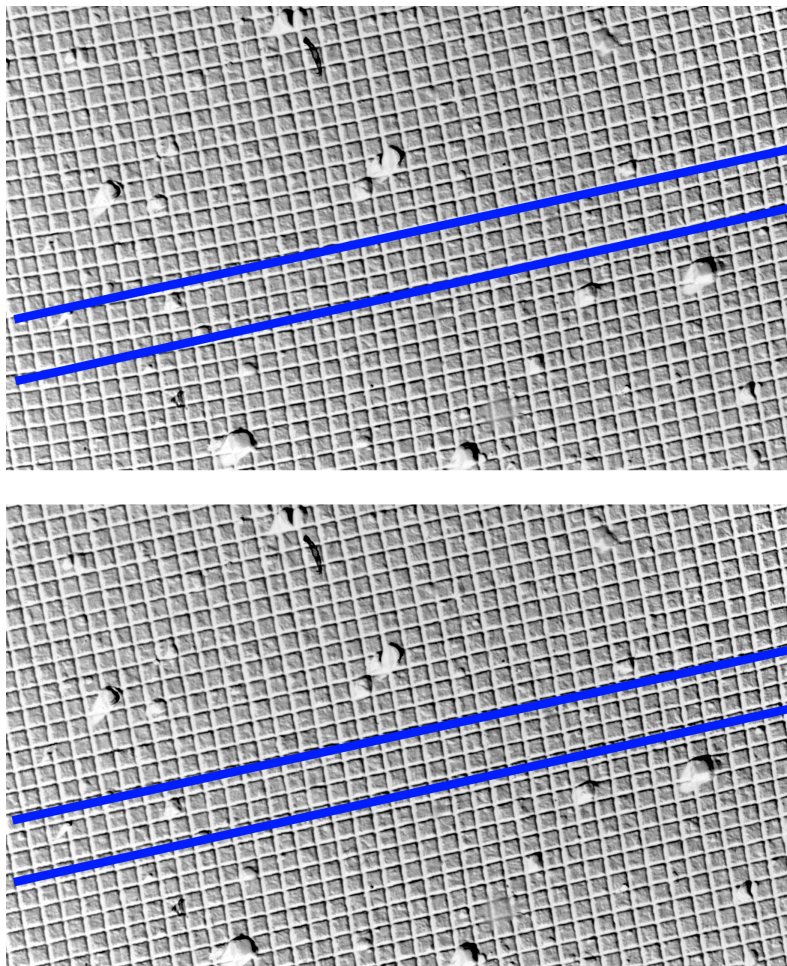


Figure 3.11: A region of a cross-grating sample image. Top: before correction, bottom: after correction. The distortion correction visibly restores the geometry of the grid structure in the image. In the lower left corner, the distortion is about as large as the grid distance which is 460 nm.

3.3. DISTORTION CORRECTION

mag	3400	3400	19000	25000	25000	64000
before alignment						
A	6.24		11.20			
B	6.32	6.92	9.76	10.07	10.65	10.12
C	6.43		11.14	8.43	8.95	
D	6.10	5.95	10.48	10.93	10.30	
after alignment						
A						
B	6.22	6.64		10.69	10.61	9.16
C	6.32			10.47	10.71	
D	6.13			10.22	10.18	

Table 3.2: The table contains the median stitching error in pixels for all 28 calibration sets. A: condenser lens in over-focus, B: condenser lens in under-focus, C: condenser lens in under-focus and objective in under-focus ($20\mu m$), D: condenser lens in under-focus and objective in over-focus ($20\mu m$) correcting for specimen lower than optimal eucentric height. For magnification settings 3400 and 25000 experiments were repeated on a second region of the sample.

A comparison of the distortion correction fields, that have been determined by the described procedures, reveals that they are very similar for all settings. The central part of the images shows almost no distortions at all, while severe distortions are observed mainly near the corners of the images (see Figure 3.12 for some examples). The stitching error for all single settings is given in Table 3.2. The most prominent change in the transformation fields can be detected when the magnification changes from 3'400x to 19'000x. At magnification 3'400x the stitching error amounts to six pixels. Varying other microscope parameters or even a full realignment of the electron microscope column has only a marginal effect on the distortion. Increasing the magnification to 19'000x, causes the error to grow up to ten pixels for all different parameter settings. The images as well as the estimated

mag	3400	19000	25000	64000
C1	391	391	391	391
C2	2601	2601	2599	2601
OBJ	2259	2259	2260	2259
DIF	1098	1619	1613	1579
INT	354	608	657	1007
P1	775	1544	1262	1364
P2	1972	1972	1952	1929

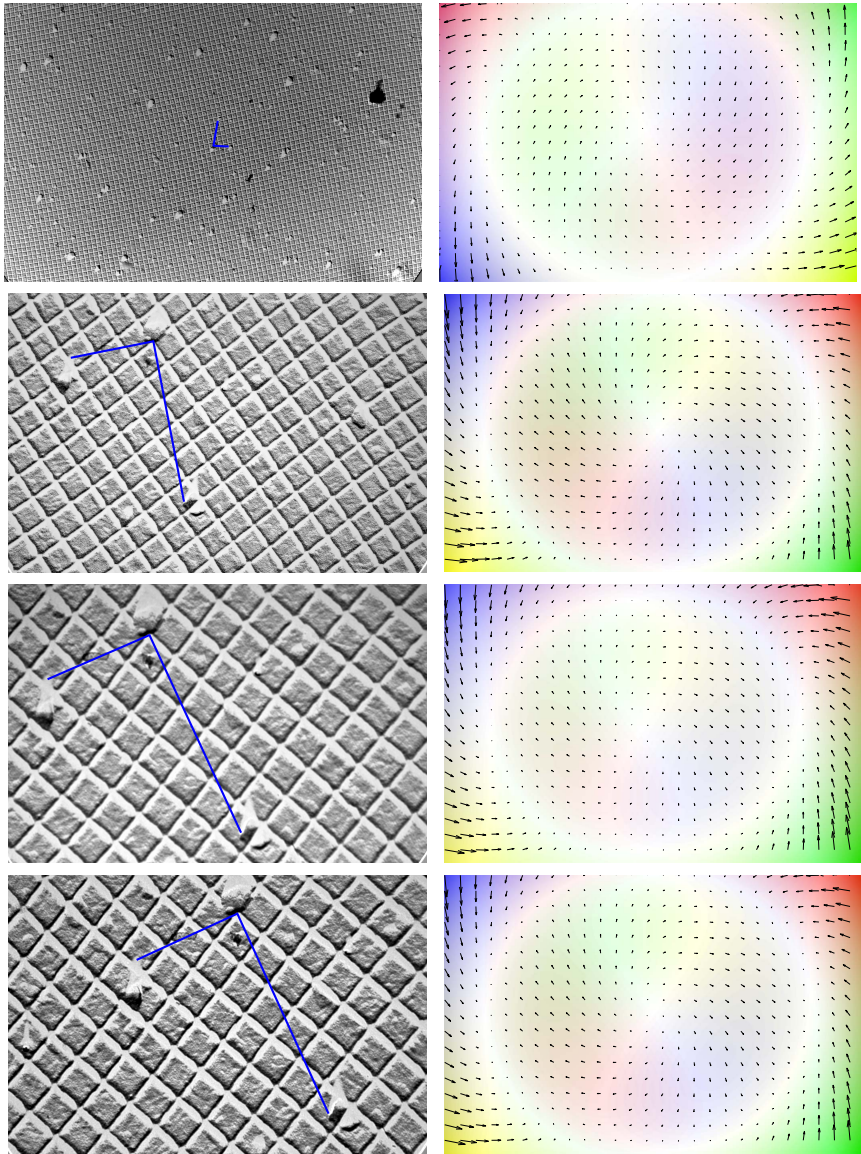
Table 3.3: The table contains the lens currents for the different magnification settings in [mA]. The abbreviations have the following meaning: C1 and C2 are the condenser lenses, OBJ is the objective lens, DIF is the diffraction lens, INT the intermediate lens, P1 and P2 the projective lenses of the Philips CM100

transformation fields exhibit a significant jump in rotation (see Figure 3.12). This result identifies the projective lens system as the main influencing factor for the distortions.

For different magnifications, the lens currents of the condenser lenses as well as the current of the objective lens remain constant (see Table 3.3) and thus, the first intermediate image is not changed. A region of this image plane is magnified by the projective lens system. With increasing magnification, the zoom area selected from the center of the intermediate image plane is shrinking. Hence, only the central, little distorted image area is further magnified and imaged by the CCD camera, and a decrease in the distortions for high magnifications would be expected if the condenser and objective lens system would cause the distortions. This analysis identifies the projective lens system as the most likely cause for the distortion change in our experiments. In addition, the increase in the distortions is correlated with a rotation jump, clearly visible in the images shown in Figure 3.12. Between the first two magnifications a clear rotation jump is visible. This is caused by a different setting for the projective lens

system. From 19'000x to 25'000x the rotation angle is small. The rotation jump is caused by a different lens configuration for the new magnification setting. As a consequence of this observation we suggest to estimate transformations for distortion correction for each major projective lens configuration. These transformations can then be used as default correction for single images, in cases where no sub-pixel accuracy is required. Table 3.2 shows, that a careful alignment of the microscope stabilizes the distortions with respect to different focus settings. Hence a good alignment of the microscope increases the benefit of an estimated default distortion correction field. The largest impact was observed on the under focus setting, which is widely used to enhance contrast for biological samples.

Figure 3.12 (*following page*): Original images (left) and the estimated distortion correction fields (right) for different magnifications. Arrows are scaled by a factor of three to enhance visibility. In the transformation field images, hue indicates the direction and saturation the magnitude of the transformation. Primary magnification from top to bottom: 3'400x, 19'000x, 25'000x, again 25'000x after alignment of the microscope. All images are 4008x2672 pixels.



3.4 Z-Alignment: Affine vs. Non-linear Alignment

During sample preparation, the specimen needs to be cut into ultra thin sections of approximately 50 nm. After the data acquisition, the images of single sections need to be aligned into one data volume. Registration of serial section images is not trivial, as the whole image content changes non-linearly between sections. In contrast to other registration problems, there is no separation between object and background, but the image contains densely packed objects, which all transform non-linearly between sections. In addition staining artifacts, tissue fractures, and neuronal processes starting and ending between sections, lead to images regions with no correspondence in the adjacent image.

The non-linear transformation between images of different sections is a combination of transformations originating from different sources. The individual transformations can be divided into two groups. One part are the structural transformations, which are non-linear and are caused by processes being cut at different positions. Structural transformations are larger for neuronal processes which run longitudinal to the cutting plane, than for processes which run orthogonal to the cutting plane. The other part are artifact transformations, which origin from the sample preparation and image acquisition process. The rotation of the section on the grid holder, the region of interest chosen by the microscopist and the spiral trajectory of the electrons inside the microscope cause an affine artifact transformation. In addition non-linear distortions origin in the cutting process, the placing of the section on the grid holder and the mass loss caused by the electron beam during image acquisition. An ideal registration of the images would correct for all artifact transformations, while preserving structural changes between the images. However, the correct transformation is not known, as the original shape information from the neuronal processes is lost in the cutting process. Therefore, there is a trade off between aligning images perfectly and preserving the

structural geometry in the data. Non-linear warping of the images to the best fit allows to get correspondences between the images for single pixels. If registration is restricted to affine transformations then structural geometry is preserved to a high degree, but non-linear artifact transformations are not corrected. As the methods described in this thesis are designed to be applicable to a wide range of data, we propose affine or rigid alignment methods to extract the 3D geometry of neuronal processes, and non-linear warping of images to solve for correspondences. If prior information about the correlation of the geometry of structures in the sample exists, this information can be employed to guide the registration. In this case the non-linear image alignment can be regularized to correct for artifact transformations which correlate over an area larger than the correlation area of the geometrical transformations.

3.5 Affine Registration

Affine registration is employed to assemble images of multiple sections into one data volume. Manual alignment of images is typically achieved with a color overlay interface. For translations the color overlay interface is sufficient, but rotations often already are hard to optimize. The alignment of sections is difficult, as the user has to find the rigid transformation that best corrects for the non-linear deformations between sections. Instead of finding the best compromise for the whole image, a human is prone to biasing the alignment in favor of a prominent structure in the images. Automatic alignment methods estimate an affine transform either by optimizing intensity based similarity measures between images, or extracting correspondent landmarks between images and estimating a transformation which maps correspondent landmarks with minimal error. For the registration between sections, the same principles apply, as discussed for the alignment of overlapping image areas for montaging (see Section 3.2.1). Hence we follow the same approach as for image stitching, employing SIFT keypoint detection for automatic landmark extraction and then

obtaining the optimal affine transformation by minimizing the Huber loss between correspondent points. The example given in Figure 3.13 shows a 3d reconstruction of a dendrite, on the left side with manually aligned images, on the right side with automatically aligned images. The segmentation was performed manually and is the same for both reconstructions, only the alignment differs. As can be seen, the automatic alignment leads to a more meandering reconstruction than the manual registration.

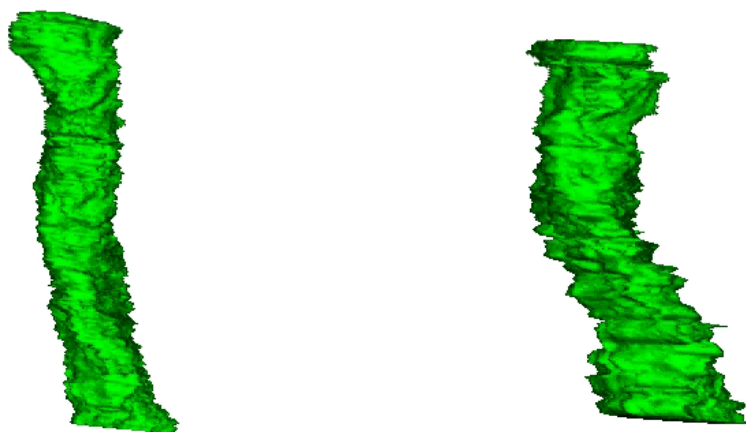


Figure 3.13: 3d reconstruction of dendrite from cat brain. Left: manual alignment, right: automatic alignment.

Figure 3.14 shows an analysis of the cross correlation coefficient between the aligned images. We compare the correlation between the image background and a region around the dendrite of interest for the reconstruction. The manual alignment focuses on enhancing the alignment around the region of interest leading to better correlations in this region. The automatic alignment instead uses the whole image context to find the optimal transformation. The increase in the cross correlation of the background demonstrates, that the automatic method corrects for a global linear transformation, which was not corrected in the manual alignment. As this transformation is linear over

the whole image, it is an artifact transformation and not caused by structural geometry. Therefore, this transformation needs to be corrected and thus the meandering reconstruction is closer to the original geometry of the dendrite, than the smooth manual reconstruction.

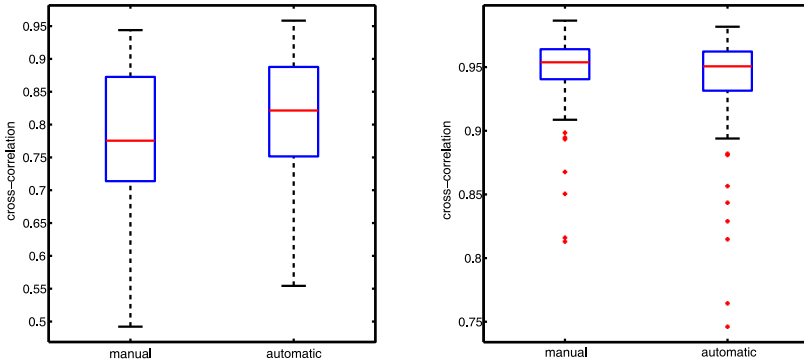


Figure 3.14: Comparison of the cross correlation coefficient between adjacent images for the dendrite reconstruction shown in Figure 3.13. Left: background area, right: region around the dendrite.

3.6 Non-linear Warping with Anomaly Detection

The target of non-linear warping methods is to estimate a non-linear transformation which minimizes the difference between two images. In contrast to affine registration, which aims at preserving structural geometry between images, warping solves the correspondence problem between images, by correcting for all transformations present. An ideal warping estimates correspondences for all pixels, thus the proposed warping method does not rely on SIFT landmarks, as described for the stitching and the affine alignment of adjacent sections.

Instead, we propose a direct alignment and estimate the optimal non-linear warping, given two images. Non-linear warping methods have a wide range of applications in medical image processing, as comparisons of images from different time points or across patients require to establish correspondences between images. In computer vision registration is required as a preprocessing step for motion tracking and 3D reconstruction. In the context of neuroanatomy, non-linear warping enables tracking of neuronal processes between sections and for propagating membrane detections across sections.

An overview of image registration is given by Zitova and Flusser (2003) in general and by Hill and Batchelor (2001) for medical image analysis. In (Glocker et al., 2007; Johnson and Christensen, 2002; Li and Leung, 2004; Perperidis et al., 2005) image registration methods are introduced that incorporate both spatial distance of correspondence points as well as intensity values. Gay-Bellile et al. (2006) registered images by using thin-plate splines. Thin-plate splines were also used for point matching (Chui and Rangarajan, 2000) which is highly related to image registration (Stewart et al., 2003). A robust framework to estimate optical flow was proposed by Black and Anandan (1996) and Brox et al. (2004), which is also related to image registration. An iterative approach to register TEM images of neuronal structures based on Gabor features is presented in (König et al., 2001). Luther et al. (1988) documented damages of the specimen caused by the electron beam of the TEM that leads to non-linear transformations of the acquired image. In our work we extend the previous work by introducing additional “visibility” variables that detect image anomalies. This concept is related to identifying regions that are visible from both images in stereo reconstruction. E.g., Strecha et al. (2004) used hidden visibility variables to detect visible and non-visible regions.

Transmission electron microscopy with its special preparation of biological samples causes problems that are not solved by existing image registration methods. In addition to strong changes of the image content between slices, the preparation process causes artifacts in the

images. Examples of such artifacts are fractures in the serial sections or staining errors which result in darker blobs in the image. Changing image content, as well as staining blobs lead to image regions, which do not have a correspondent region in adjacent images. In addition, fractures as well as folds, do cause strong non-linearly deformations of the section and thus the image to align.

3.6.1 Expectation Maximization

We solve the warping problem by way of an expectation maximization algorithm that calculates a non-linear warping which is parametrized by a polynomial kernel expansion of reference points. The correspondence points are not a priori fixed but selected during the registration process. Anomalies, which are caused by the biological sample preparation, are estimated in the image.

In a Bayesian framework the optimal transformation matrix β maximizes the posterior probability

$$p(\beta|X, Y) = \sum_{M \in \mathcal{M}} \frac{p(X, Y|\beta, M) \cdot p(\beta) \cdot p(M)}{p(X, Y)}, \quad (3.9)$$

where X and Y correspond to the warp image and the target image. The variable M denotes a binary correspondence matrix. M_{ij} is one, if point x_i in the warp image corresponds to point y_j in the target image and zero otherwise. In addition a point x_i can be assigned to an outlier class, denoted as M_{i0} . Thus, the whole matrix M is of size $n_1 \times (n_2 + 1)$ where n_1 is the number of chosen points in the warp image and n_2 is the number of possible correspondences for each of these points in the target image.

Application specific constraints on the assignments can be modeled by an appropriate definition of \mathcal{M} . As each point x_i should be assigned to only one correspondence point y_j the set $\mathcal{M} = \{0, 1\}^{n_1 \cdot (n_2 + 1)}$ denotes all admissible assignment matrices M . We define $p(M)$ to be zero for configurations that assign more than one correspondence

point to x_i and assume all valid configurations to be uniformly distributed:

$$p(M) = \begin{cases} \frac{1}{n_1 \cdot (n_2 + 1)} & \text{if } \sum_{j=0}^{n_2+1} M_{ij} = 1 \text{ for all } i = 1, \dots, n_1 \\ 0 & \text{else} \end{cases} \quad (3.10)$$

This definition of $p(M)$ ensures that each point x_i is assigned exactly to one correspondence point or is marked as not relevant.

To define $p(\beta)$ the components of the solution vector β are assumed to be normally distributed. Therefore, we introduce a ridge penalty which is described by the normal distribution $\varphi_{\mu, \sigma}$ with location parameter μ and variance σ^2 as prior distribution

$$p(\beta) = \prod_{i=1}^{n_\beta} \varphi_{0, 1/\sqrt{\lambda}}(\beta_i) \quad (3.11)$$

for β . The parameter λ is the ridge penalty that controls the complexity of the regression function.

The distribution of $p(X, Y | \beta, M)$ should depend on the similarity of the correspondence points based on gray values, as well as on the quality of the geometric fit. Furthermore, we need to take care of image anomalies that are assigned to the outlier class. This outlier class is modeled as a uniform distribution. The complete data likelihood is distributed as

$$\begin{aligned} p(X, Y | \beta, M) &= \frac{1}{N} \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} \left(\varphi_{0, \sigma_1}(v(x_i) - v(y_j)) \cdot \right. \\ &\quad \left. \varphi_{0, \sigma_2}(\phi(x_i)\beta - y_j) \right)^{M_{ij}} \\ &\quad \cdot \prod_{i=1}^{n_1} (\varphi_{0, \sigma_1}(c\sigma_1))^{M_{i0}} \end{aligned} \quad (3.12)$$

Here $v(x_i)$ is a vector of the gray values of a small patch centered at x_i . The difference of two such patches serves as a dissimilarity measure that is easy to compute and takes context information about a small area around the points into account. The non-linear transformation is modeled again as a polynomial kernel expansion (Eq. 3.5) followed by a multiplication with the matrix β to project the points back to two dimensions.

The second factor provides a penalty for points that are marked as not relevant. The constant c is given as the 0.98 quantile of the cumulative chi square distribution, where the degrees of freedom in principle correspond to the number of pixels in $v(x_i) - v(y_j)$. To reduce the influence of noise in the similarity measure, we perform a principal component analysis and project the high dimensional difference vectors down to the eigenvectors that correspond to the largest 98% of the eigenvalues. This also reduces the degree of freedom of the cumulative chi square distribution. Thus, the outlier factor transforms the assumed normal distribution of the gray value similarity into a heavy tailed distribution, providing a robust solution for outliers caused by non relevant elements.

In order to maximize $p(\beta|X, Y)$ which yields an optimized transformation, we maximize the logarithm of $p(X, Y|\beta, M) \cdot p(\beta)$ under the constraint that M is a valid matrix, i.e. $\sum_{j=0}^{n_2} M_{ij} = 1$ for all $i = 1, \dots, n_1$. Since the assignment variables M are unobservable, we use the EM-algorithm to maximize the joint log-posterior. The algorithm iterates between estimating the expectation of the latent variables M_{ij} while keeping β fix and maximizing the joint log-posterior while keeping the expectation values of M_{ij} constant. The variances for the normal distributions are also calculated during the maximization step. The log-posterior is maximized with respect to the transformation β as well as the variances of the normal distributions σ_1 and σ_2 .

E-step: In each iteration the expectation values γ_{ij} for all possible assignments are updated using the currently optimized β :

$$\begin{aligned}\gamma_{ij} &= \mathbf{E}[M_{ij}|X, Y, \beta] \\ &= \frac{p(X, Y|\beta, M_{ij} = 1)}{\sum_{l=0}^{n_2} p(X, Y|\beta, M_{il} = 1)}\end{aligned}\quad (3.13)$$

M-step: The computed expectation values γ_{ij} are substituted for the unobserved correspondence assignments M_{ij} in the joint log-posterior Equation 3.12. The missing parameters $\beta, \sigma_1, \sigma_2$ are then estimated by maximization. For the transformation β this MAP approach yields a weighted ridge regression problem (Hastie and Tibshirani, 2001) with weights γ_{ij} . The transformation matrix β is maximized by

$$\beta \leftarrow (\phi(\tilde{\mathbf{X}})^T \Gamma \phi(\tilde{\mathbf{X}}) + \lambda \sigma_2^2 \mathbf{I})^{-1} \phi(\tilde{\mathbf{X}})^T \Gamma \mathbf{Y} \quad (3.14)$$

where Γ is a $(n_1 \cdot n_2) \times (n_1 \cdot n_2)$ -dimensional diagonal matrix of the weights γ_{ij} . The $(n_1 \cdot n_2) \times 2$ matrix $\tilde{\mathbf{X}}$ contains n_2 copies of each position vector x_i and the $(n_1 \cdot n_2) \times 2$ matrix \mathbf{Y} contains n_2 possible correspondence points for each position x_i . The parameter λ is the regularization parameter defined by the prior distribution $p(\beta)$ (Eq. 3.11). In our experiments $\lambda = 0.001$ sufficiently regularizes the assignments.

The standard deviations are updated by

$$\sigma_1 \leftarrow \sqrt{\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \gamma_{ij} \cdot \rho(x_i, y_j)^2}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \gamma_{ij}}}\quad (3.15)$$

$$\sigma_2 \leftarrow \sqrt{\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \gamma_{ij} \cdot \|\phi(x_i)\beta - y_j\|^2}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \gamma_{ij}}}\quad (3.16)$$

where σ_1 and σ_2 are invariant to outliers since $\gamma_{i,0} \approx 1$ for these points and therefore $\gamma_{i,j} \approx 0$ for $1 \leq j \leq n_2$.

Choice of initial points x_i : So far all warping points x_i are assumed to be arranged on a regular grid. While this design ensures that all interesting structures in the image are covered by a warp point nearby, interest points are often placed in background areas. To increase precision we would like to position each point directly in content rich parts of the image while still covering all biologically relevant structure in the image. For this purpose we calculate the entropy of the intensity value in a 13^2 neighborhood around each point in the image. The entropy is high for pixels along structures with a high contrast. Now we shift each warp point x_i of the regular grid to the position with the highest entropy value in its neighborhood. This local adaptation method preserves the coverage of the whole image while emphasizing areas with rich image content.

3.6.2 Experiments

For our experiments we have used images gathered in a computational neuroscience project. When imaging with a TEM it must be possible for single electrons to penetrate the probe. Therefore the specimen is first stained, then embedded into resin and cut into ultra thin sections of 40 – 50 *nm* thickness.

As described in detail in section 2.1.1, the three major sample preparation steps that may cause artifacts in the image are the staining, the cutting and the recording with the electron beam. Staining may produce additional dark areas in the image that do not correspond to original biological structures. During the cutting process the slice is exposed to significant stress and it may be non-linearly transformed or it even can encounter fractures. Finally, exposure to an electron beam causes a mass loss of the specimen and leads to additional transformations.

Figure 3.15 depicts example image from the data set used for our experiments. The top part shows an image of typical quality. The structure in the left bottom quarter of the image with the pike on the right hand side shows a dendrite. The smaller ellipse like structures over the image are myelinated axons. The four smaller images below

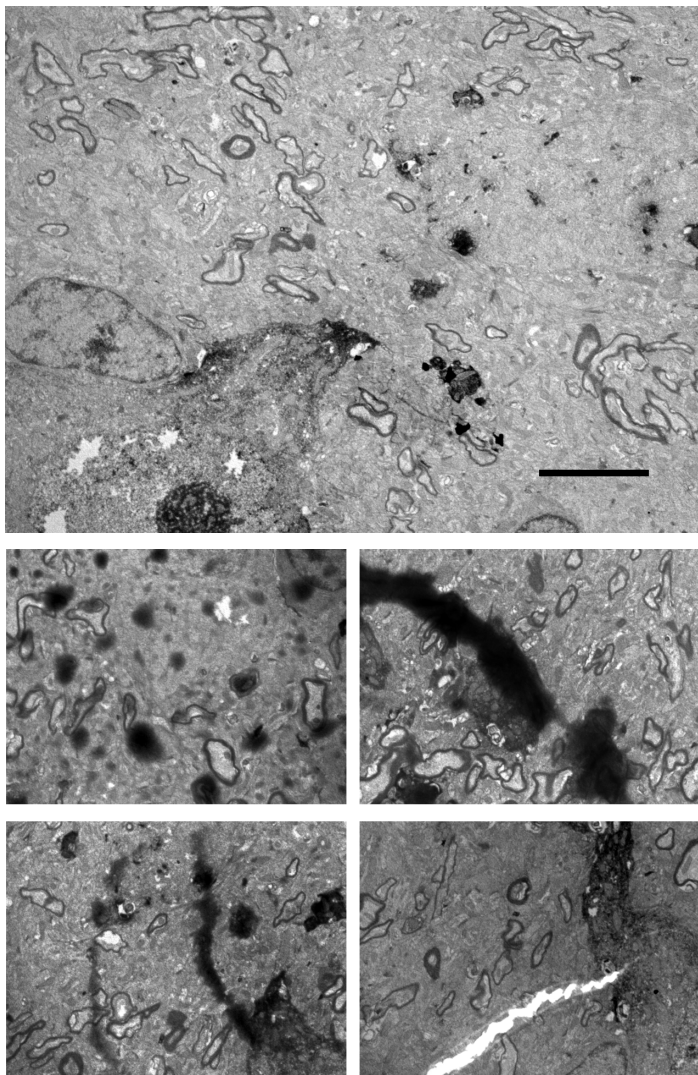


Figure 3.15: Examples of experimental data. Top: image of typical contrast, bottom: examples of staining blurs and a crack. The scale bar corresponds to $4\mu\text{m}$.

contain examples of image artifacts, that are caused by the sample preparation: In the lower right corner we see an example of a crack in the specimen caused by the cutting process. The dark spots in the upper left image and the dark stripes in the upper right and lower left image are artifacts of the staining procedure.

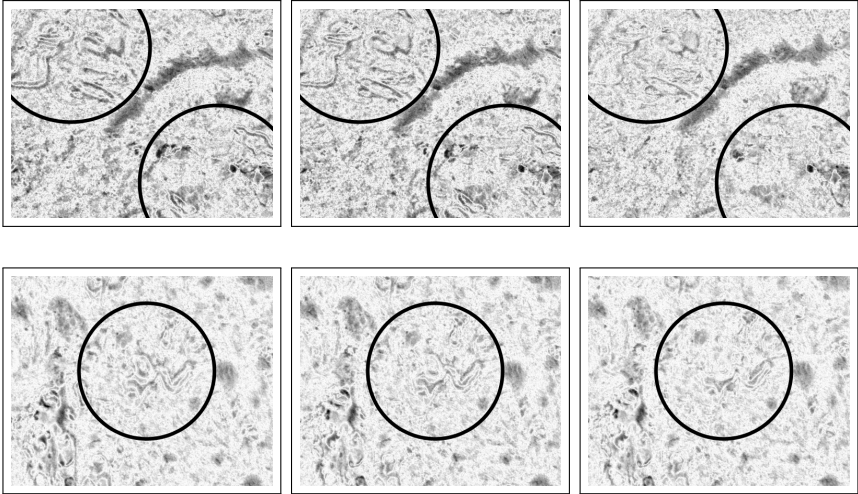


Figure 3.16: Difference images for (left) affine transformation, (middle) least squares matching with polynomial basis functions, (right) expectation-maximization including visibility estimation. The original images are shown in Figure 3.17

We tested our approach on two series of electron microscopy images. The first series contains 97 images that were taken at 3400x magnification with a resolution of 1032x1376 pixels, one image per section. The second series consists of 284 images taken at 13500x magnification that were distortion corrected and stitched into 71 section images. The resolution of the second stack is 2672x4008 pixels.

In a preprocessing step we correct the radial illumination gradient visible in the images by dividing each image with a smoothed version of itself ($\sigma = 30$ pixel). Then an initial affine transformation based

on SIFT features is estimated for each image. This transformation is then refined with our warping approach based on expectation maximization. For comparison we also provide the refinements for ridge regression with the ordinary least squares error

$$\min_{\beta} E(\beta) = \min_{\beta} \sum_{i=1}^n \|\phi(x_i)^T \beta - y_i\|^2 + \lambda \|\beta\|^2 \quad (3.17)$$

and the robust version, where the Huber loss function is used instead of the least squares error

$$L_c(\xi) = \begin{cases} c|\xi| - \frac{c^2}{2} & \text{for } |\xi| > c, \\ \frac{\xi^2}{2} & \text{for } |\xi| \leq c. \end{cases}$$

This robust method can be seen as a non-linear extension of the rigid approach described by Ourselin et al. (2001). As these methods require fixed corresponding points, we calculate the correlation coefficient of a patch around x_i and the appropriate patches of the target image. The result is then weighted with a Gaussian density centered at the position of x_i and a standard deviation of two times the standard error of the affine match. Each mapping point x_i is then assigned to its correspondence point y_i according to the maximum of the obtained function.

To provide a visual impression of the obtained warpings we first show difference images for the affine initialization and two warpings (Figure 3.16). The original images are shown in Figure 3.17. The darker the color in the difference image, the larger the absolute difference in gray values between the warped image and the target image. The images in the left column show the difference map for the robust affine transformation, the images in the middle the result for ridge regression with polynomial basis functions. The third column shows the difference maps for our new expectation-maximization method. If one focuses on the images in the upper row on the top left region as well as on the right and bottom border, one can observe dark stripes

in the affine transformation image that are getting thinner for the least squares and even more thinner for the expectation maximization method. This error measure shows clearly that important edges are not matched very well by the affine transformation, but for the expectation maximization solution, there are only very small differences left over. Edges in the image are fitting significantly better than for the standard technique. In addition we tried to compare our method to optical flow methods (Black and Anandan, 1996; Brox et al., 2004). We could not find parameter settings that were able to cope with the significant structural changes between sections.

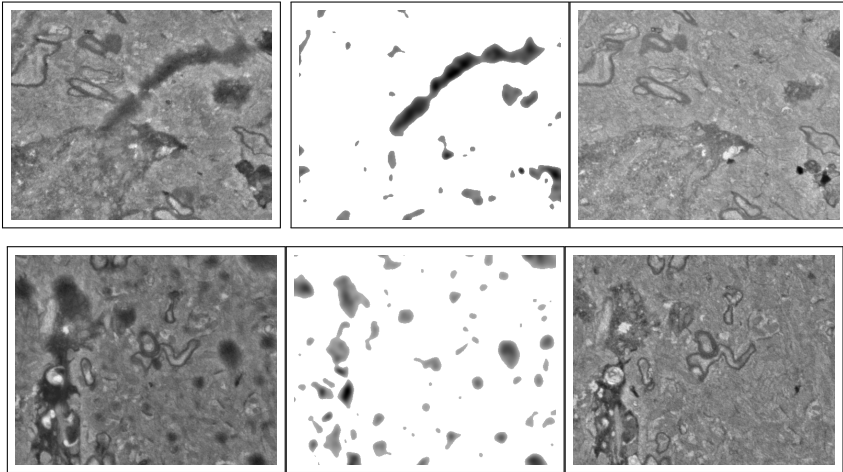


Figure 3.17: Left: The first image, Middle: The visibility map (Detected image anomalies). Right: The second image.

Since the detection and localization of image anomalies is important for our method, we show examples of the estimated image anomaly regions in Figure 3.17. The darker the color in this image, the more likely the region belongs to an artifact. The dark strip over the upper image as well as the dark blobs in the lower image are clearly detected as not relevant structures which is our desired goal. The information about these anomalies can now be used for further

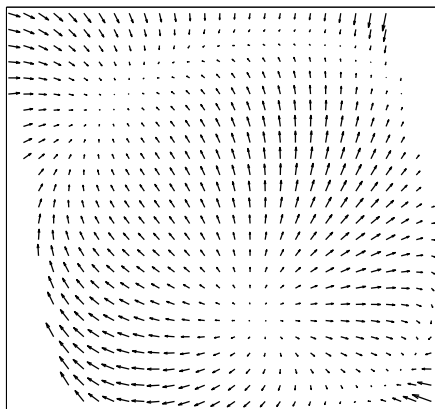


Figure 3.18: A needle diagram of non-linear image warping.

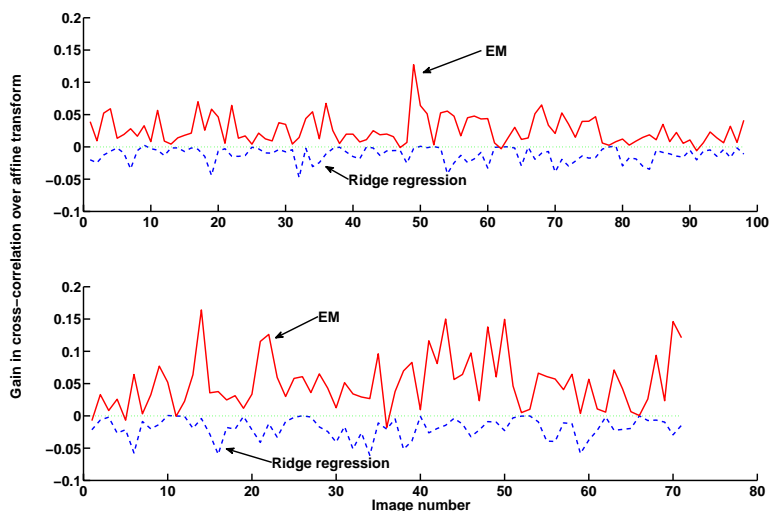


Figure 3.19: Gain or loss of EM warping (solid line) and ridge regression (dashed line) relative to robust Huber loss estimation. Top: the single image stack containing 97 images, bottom: distortion corrected and stitched image stack containing 71 images.

processing steps in computational neuroanatomy, e.g. in the 3D reconstruction of the neural connectivity structure. To visualize the estimated non-linear transformation of the image, we show a needle diagram of the transformation in Figure 3.18.

To demonstrate the improvement of the new method, we have registered both stacks of TEM images from the above described neuroanatomy project and measured the cross-correlation between the target image and the warped image. Figure 3.19 shows the results for the two image stacks. Our EM approach has been initialized with the robust ridge regression solution. Therefore, improvements in cross-correlation values over the robust ridge regression solution serves as a measure of success for our model of image registration, i.e., large differences in cross-correlations denote a significantly better registration of the TEM images than with the robust Huber loss. The EM method clearly outperforms the other techniques by up to 15 percent gain in cross-correlation.

3.6.3 Conclusion

Registration of images is an important step in the 3D reconstruction. Especially in Transmission Electron Microscopy of biological samples, image anomalies occur frequently caused by the sample preparation process and by the image acquisition process. In this paper we propose a novel method for image registration that jointly estimates image anomalies and an image matching in a Bayesian model. The mixture model enables us to estimate assignment probabilities as well as probabilities for damages. The method performs superior to standard methods like linear affine transformation, and non-linear transformation. Even state-of-the-art outlier detection methods are inferior in performance compared to our mixture model. The expectation-maximization algorithm optimizes the model efficiently and is straight forward to implement. The experiments convincingly demonstrate that the model not only improves the image registration process, but also detects image anomalies. So far we applied our method to solve the correspondence problem between two adjacent images. It is pos-

3.6. NON-LINEAR WARPING WITH ANOMALY DETECTION

sible to estimate the complete warping of a stack, by propagating the estimated non-linear transformations across sections. For artifacts like section raptures, estimating the warping independently for each pair of sections is beneficial, as the warping is not smooth across sections. However, if the non-linear transformations change smoothly across sections, the proposed approach can be employed to estimate a three dimensional warping, taking the lateral smoothness into account.

Chapter 4

Geometry Extraction

Due to the low lateral resolution of serial section TEM images, we divide the geometry extraction into two steps. First, cell membranes are segmented employing a trained random forest classifier combined with sub-modular binary potentials for gap closing. While the gap enhancement incorporates membrane detections in adjacent sections, the focus of the segmentation is on the two dimensional high resolution image. In the second step the membrane enclosed regions are grouped across sections to extract the three dimensional geometry of neuronal processes. The grouping process takes the geometry of the whole section into account to yield consistent groupings.

4.1 Segmentation

To extract the geometry of neuronal processes from electron microscopy images, the images have to be segmented into regions corresponding to structures of interest. One possibility is to directly classify pixels, e.g. as representing a bouton, dendrite or axon. This multiclass scenario is hard to solve, as the classification of a pixel often cannot be decided by its intensity alone. A bouton for example is in most cases identified by the vesicles inside, which express a prominent texture. But, regions of a bouton with few vesicles are likely confused with the texture of an

axon. To overcome this ambiguity, we propose a different approach, which concentrates on the segmentation of membranes, reducing the segmentation to binary classification. All structures of interest for geometry extraction, are surrounded by membranes. Thus, a complete membrane segmentation results in closed regions corresponding to structures of interest. Functional annotation is then performed based on the characteristics of the whole region instead of the local context around a pixel. Our membrane segmentation approach provides as an additional advantage that image characteristics of membranes are robust with respect to different animal types and staining protocols. Figure 4.1 contains two example images showing the variability of image characteristics for different animal types and staining methods. The left image is from drosophila larva brain and contains thick, but sometimes blurry membranes with weak contrast. The image from cat brain shown on the right side expresses strong contrast and very rich texture. While the texture characteristic changes drastically between the images, the main attributes of membranes, as thin, smooth, and elongated structures remain the same. Thus, the focus on membrane

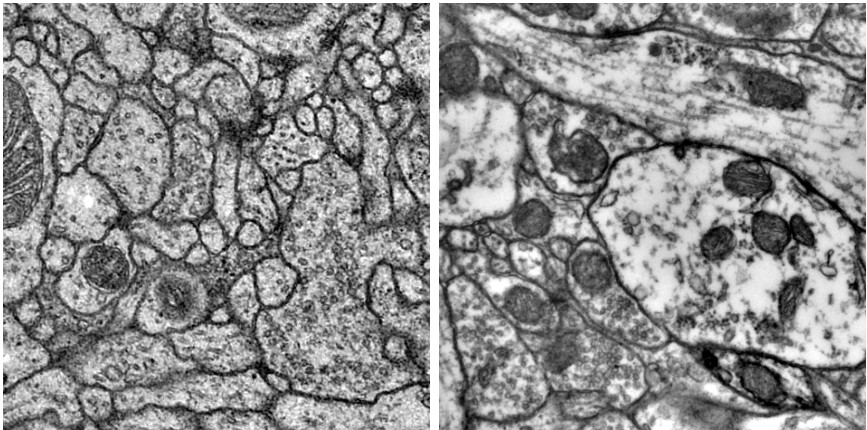


Figure 4.1: Two example images, for different animal types and staining methods. Left: drosophila larva, right: adult cat.

segmentation makes our approach easily adaptable to a wide range of data from different neuroanatomy projects as well as segmentations of other thin elongated structures.

4.1.1 Random Forest Classifier for Membrane Detection

Classification of membranes in electron microscopy images poses a challenging task. The rich and dense structure of brain tissue and other biological samples leads to highly textured images. Objects of interest are typically identified by shape and context rather than by gray value information. As a consequence, state of the art methods train a classifier on manual annotations to capture the image characteristics of membranes. Classifiers employed for this task include neuronal networks (Jurrus et al., 2010; Mishchenko, 2009), boosting (Vitaladevuni and Basri, 2010), convolutional neuronal networks (Jain et al., 2007) and random forest (Andres et al., 2008). For our framework we decided to employ a random forest classifier for several reasons: (i) Random forests are fast in training and prediction, which is ideal for large data sets. (ii) The classifier is robust against overfitting and little manual annotations are necessary to train a random forest for good membrane detection rates. The large variability of image characteristics depending on animal type and sample preparation, requires the retraining of the classifier for data sets from different neuroanatomy projects. With a robust classifier, the manual effort involved for retraining, is kept to a minimum. (iii) The only parameters to tune are the number of trees and the size of the feature subset used to build each decision tree of the forest. The method is not very sensitive to these parameters and default values produce good results for all data sets in our experiments. We employ 500 trees and we set the number of features chose to the square root of the total number of features, which is the default suggested by Breiman (2001). (iv) The vote output of the trees can be interpreted as probability measure for membrane detection. This choice allows us to refine the classification of pixels with a high uncertainty, either by guided user interaction

or with our gap closing framework introduced in the following section. The feature set extracted from the images is designed to capture the characteristics of membranes with little computational cost. Extracted features include the gray value, gradient magnitude, Hessian eigenvalues, and difference of Gaussian for the image smoothed with Gaussian filters of different kernel sizes. In addition we convolve the image with a steerable filter at different orientations. Each filter output serves as a feature, as well as the minimal, maximal and average output of the steerable filter at a pixel position.

Figure 4.2 demonstrates the classifier output with very little training data. The random forest classifier was trained exclusively on the annotated pixels given in the original image shown at the top of the figure. The middle image shows the output of the random forest for the whole image. Pixels with higher membrane probability appear darker in the image. The classification thresholded at a probability of 0.5 is given in the bottom image. Regions with less than 100 pixels are filtered from the output. Even with this small training set and simple postprocessing, the classification captures most of the membranes around neuronal structures in the image. Membranes around mitochondria pose a special problem for classification. On the one hand mitochondria are surrounded by membranes, expressing characteristics identical to other membranes. On the other hand, mitochondria are intracellular structures, and should be distinguished from neuronal processes. In many cases, the distance between the outer cell membrane and the mitochondrion is so small, that the two membranes are merged in the segmentation, rendering the removal of mitochondria membranes from the segmentation output difficult.

4.1.2 Segmentation with Graph Cut

As demonstrated in Figure 4.2, the random forest classifier captures the main image characteristics of membranes with little manual annotations data. But, the segmentation also shows false negative classifications, leading to gaps in the detected membranes. One reason for missing membrane detections are membranes, which run non orthog-

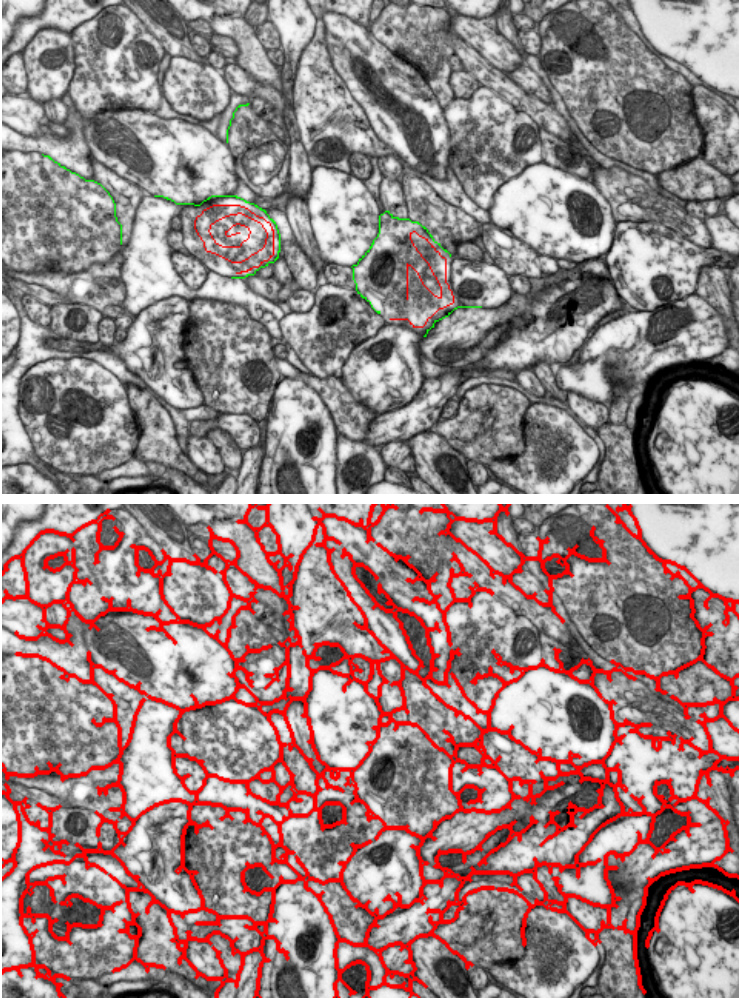


Figure 4.2: The random forest classifier captures the membrane characteristic with very little training data. Top: original image with training annotations (green: membrane, red: background), bottom: overlay with detected membrane, classification is filtered to delete small regions and skeletonized.

onal to the cutting direction and thus get blurred by the projection of the physical sample section to the image. In addition, membranes can be obscured by mitochondria or staining artifacts. Manual interaction is necessary to correct these false classifications. Graph cut provides an excellent framework for semi-automatic segmentation (Rother et al., 2004). Typically, the probabilistic output of a classifier and manual user annotations are combined with a smoothing constraint to reduce false positive detections (Wels et al., 2008).

For thin and elongated structures like membranes, graph cut is well known to have problems with “shrinking bias”. Current state of the art segmentation methods overcome this problem by combining overall smoothness with gradient flux, to enhance the segmentation result (Vasilevskiy and Siddiqi, 2002; Boykov and Funka-Lea, 2006). Vu and Manjunath (2008) employ gradient flux to segment the interior region of a dendrite. But, in images with textured background, like electron microscopy images, gradient flux leads to false positive detections, due to the high gradient in the background.

We improve the segmentation of thin elongated structures by enhancing gap completion. The energy term is submodular and thus can be efficiently globally optimized using max-flow/min-cut computation. The novel energy framework combines a discriminative model for membrane appearance learned by a random forest classifier with perceptual grouping constraints for contour completion in a single energy minimization framework. The gap completion term follows the principle of good continuation, which states that elongated structures, which form a continued visual line should be grouped together. Thus, the proposed energy term focuses on the main characteristics of membranes as thin elongated structures, which are biologically given and therefore not influenced by different sample preparations. We also take information of adjacent sections into account to support the segmentation of membranes which are not prominent in one image, but better detectable in corresponding regions of nearby sections. This corresponds to the principle of non accidental occurrence, which states that elements should be grouped, if their configuration is unlikely to

occur by chance.

The framework is evaluated on two different data sets of conventional ssTEM images from neuroanatomy. The image stacks differ not only in the type of animal brain shown (mammal and insect), but also in the staining protocols used, leading to very different image characteristics. On both data sets, the proposed cost function with perceptual grouping constraints outperforms the state-of-the-art segmentation using gradient flux. These results point out the robustness of the proposed perceptual grouping constraints to different staining protocols and animal types. The high quality of the membrane segmentations supports fully automatic 3D reconstructions of neuronal structures. To demonstrate the wide applicability of the proposed framework we also tested the constrained segmentation method on street detection in satellite imagery and we present convincing results.

4.1.3 Graph Cut with Gradient Flux

In the graph cut framework each pixel p is mapped to corresponding labels $y_p \in \{0, 1\}$ such that the entire labeling y for all pixels minimizes a given energy function $E(y)$. Typically the energy function $E(y)$ consists of a summation over the data term $E_d(y_p)$ and a smoothness term $E_s(y_p, y_q)$ over neighboring pixels:

$$E(y) = \sum_{p \in P} E_d(y_p) + \lambda \sum_{p \in P, q \in N_2(p)} E_s(y_p, y_q), \quad (4.1)$$

where P denotes the set of all pixels and $N_2(p)$ the set of all pixels adjacent to a pixel p in the 2D image plane. As long as E_s is regular, i.e. $E_s(0, 0) + E_s(1, 1) \leq E_s(1, 0) + E_s(0, 1)$, the global minimum of $E(y)$ can be efficiently found by max-flow/min-cut computation (Kolmogorov and Zabini, 2004; Boykov and Kolmogorov, 2004). For this purpose, a graph $G = (\mathcal{V}, \mathcal{E})$ is defined. The set of graph nodes \mathcal{V} consists of all pixels $p \in P$ and two additional terminal nodes s and t which represent foreground and background in the segmentation. As

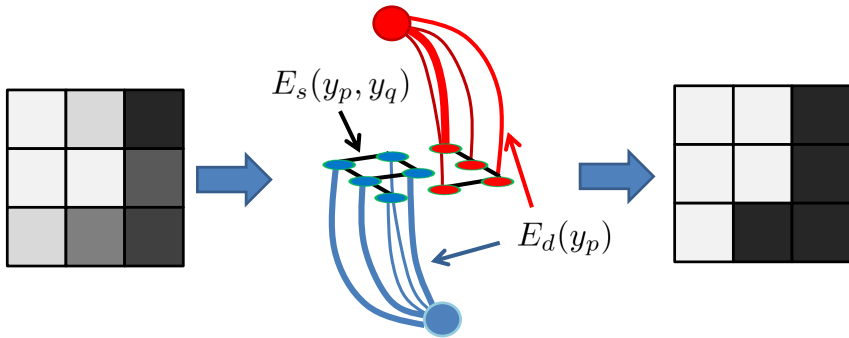


Figure 4.3: Illustration of the standard graph cut framework. The unary potential defines edge weights from all pixel nodes to the source (red) and sink (blue). The binary potential defines edge weights between pixel nodes (black lines). A valid cut has no path between source and sink.

illustrated in Figure 4.3, the set of directed edges \mathcal{E} connects all pixels p to their neighbors $q \in N_2(p)$. Weights are assigned to these edges as specified by the smoothness term $E_s(y_p, y_q)$. In addition the set of edges \mathcal{E} connects each pixel to the two terminal nodes s and t with weights specified by $E_d(y_p)$. Minimizing $E(y)$ corresponds to finding the optimal cut $\mathcal{C} \subset \mathcal{E}$ such that no path exists between the terminal nodes s and t in the graph $G_{cut} = (\mathcal{V}, \mathcal{E} - \mathcal{C})$. The cut is optimal in the sense that the sum of all edge weights of all edges included in the cut is minimal.

Often graph cut approaches use a definition of E_s which penalizes for discontinuities in the segmentation for neighboring pixels of similar intensities (Boykov and Funka-Lea, 2006):

$$E_s(y_p, y_q) = \exp\left(-\frac{(x_p - x_q)^2}{2\sigma_s^2}\right) \cdot \frac{\delta(y_p, y_q)}{dist(p, q)}, \quad (4.2)$$

where x_p is the gray value of the image at pixel p and $dist(p, q)$ takes the distance between neighbored pixels into account. The Kronecker

delta function $\delta(y_p, y_q)$ equals 0 if $y_p = y_q$ and 1 otherwise. This ensures that the energy term is regular.

For the segmentation of thin and elongated structures, like blood vessels, it is common to use an additional term $E_{gf}(y_p)$ that incorporates the flux of the gradient vector field into the segmentation. It has been shown that gradient flux can overcome the problem of “smoothing away” thin structures (Vasilevskiy and Siddiqi, 2002). Flux is defined according to

$$F(p) = \sum_{q \in N_2(p)} \langle u_{pq}, v_q \rangle, \quad (4.3)$$

where u_{pq} is a unit vector oriented from pixel p to the neighboring pixel $q \in N_2(p)$ and vector v_p corresponds to the gradient vector at pixel p . This term can be seen as the flow of the gradient vector field through the contour of the segmented region. The corresponding unary potential $E_{gf}(y_p)$ is defined as:

$$E_{gf}(y_p) = \begin{cases} \max(0, F(p)) & \text{for } y_p = 1 \\ -\min(0, F(p)) & \text{for } y_p = 0 \end{cases} \quad (4.4)$$

Kolmogorov and Boykov (2005) provide a detailed description on how to define edge weights for flux in graph cut.

In a simple setting, the term $E_d(y_p)$ of Equation (4.1) can be defined as relying directly on the pixel intensities in the original gray value image. But, structures in electron microscopy images are often only recognizable by their texture in the local context. Therefore, we use the probabilistic output of a random forest classifier trained on annotated data for membrane detection, similar to the approaches in (Wels et al., 2008; Dollar et al., 2006). To account for the random forest classifier, we rename the data term to $E_{rf}(y_p)$ throughout the paper.

Taking the details explained above into account, our implementa-

tion of the state of the art segmentation method looks as follows:

$$E(y) = \sum_{p \in P} E_{rf}(y_p) + \lambda_s \sum_{\substack{p \in P \\ ,q \in N_2(p)}} E_s(y_p, y_q) + \lambda_{gf} \sum_{p \in P} E_{gf}(y_p). \quad (4.5)$$

Using gradient flux to enhance the segmentation of thin objects also has a drawback. In textured images the image gradient is not only very high at the desired segmentation borders, but also at other image regions with high contrast. Therefore the gradient flux can cause a large amount of false positives in the resulting segmentation. In addition we want to use the output of a trained membrane detector as data term for the segmentation. Experiments showed that gradient flux and smoothness alone is not sufficient to compensate for weakly detected membranes, as is illustrated in the following toy data setting. We generate an image, of a perfect membrane represented as straight black line on a white background. A weak classifier response is simulated by fading out a section of the line (Figure 4.4). Although the gradient flux and smoothness terms were calculated on the perfect, non-faded line, they cannot compensate for the weak E_{rf} input. The gradient enhances segmentation of the rim of the lines, but any attempt to make the segmented regions solid by using the smoothness term E_s leads to gaps in the membranes segmented. This problem is more aggravated on real data, since weak classifier responses often occur in the case of membranes which appear fuzzy in the image due to non orthogonal cutting or staining conditions. In these cases the gradient along the membrane is small and thus further limits the use of the gradient vector flux in the segmentation. To overcome this problem we introduce a novel energy term, that focuses on the principle of good continuation to close gaps along membranes. To overcome the shortcomings of gradient flux, we introduce a directional energy term that is based on the perceptual concept of good continuation.

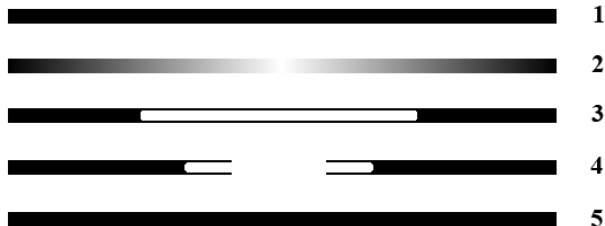


Figure 4.4: Toy example for membrane segmentation. The good continuation energy term is able to produce a solid segmentation where gradient flux fails. From top to bottom: (1) original perfect line, (2) line with a faded out segment as input for the data term E_{rf} , (3) with gradient flux, segmentation of borders is improved, (4) attempt to close segmented structures by additional use of the smoothness term E_s , (5) solid segmentation using only E_{rf} , and the directed term E_{gc} .

4.1.4 Closing Gaps: Submodular Energy Term for Good Continuation

Intuitively, lines as well as membranes are directed structures. By the principle of good continuation well classified parts of directed structures should enforce smoothness in labels along their orientation. This is formulated by $E_{gc}(y_p, y_q)$:

$$E_{gc}(y_p, y_q) = | \langle v_p, u_{pq} \rangle | \cdot \exp \left(- \frac{(x_p - x_m)^2}{2\sigma_{gc}^2} \right) \cdot \frac{\delta_{\rightarrow}(y_p, y_q)}{dist(p, q)}, \quad (4.6)$$

where u_{pq} is a unit vector with the orientation of a straight line between pixels p and q , and v_p is a vector directed along the membrane. The length of v_p reflects the orientedness of the image at p . For this purpose we use a directed filter consisting of a straight line with a thickness comparable to the membrane thickness in the training images. $\langle v_p, u_{pq} \rangle$ is then estimated by the response to this

filter oriented according to u_{pq} . The value of x_m is given as the average gray value of membrane pixels and σ_{gc}^2 can be estimated as the variance of these gray values. Thus, the difference $(x_p - x_m)$ weights the energy term according to the similarity of x_p to the typical gray value of a membrane.

In contrast to Equation 4.2 the factor $\delta_{\rightarrow}(y_p, y_q)$ is not symmetric. Instead $\delta_{\rightarrow}(y_p, y_q) = 1$ for $y_p = 1, y_q = 0$ and $\delta_{\rightarrow}(y_p, y_q) = 0$ for all other cases. This asymmetric definition ensures that E_{gc} only penalizes for cuts that violate the smoothness along the direction of membrane pixels. Although $\delta(y_p, y_q)$ is not symmetric, it is still regular and thus the global optimality of the resulting segmentation is assured (see also (Boykov and Funka-Lea, 2006; Winn and Shotton, 2006)).

In addition we incorporate information from adjacent sections into the segmentation using:

$$E_{na}(y_p, y_q) = m_q \cdot | \langle v_p, v_q \rangle | \cdot \frac{\delta_{\leftarrow}(y_p, y_q)}{\text{dist}(p, q)}, \quad (4.7)$$

where m_q is the probability of pixel q being a membrane and v_p is the large eigenvector of the Hessian at pixel p multiplied by the corresponding eigenvalue. Thus, a high confidence in pixel q being a membrane is propagated to the next section if the corresponding region is similarly oriented. This orientation constraint has the benefit, that it is unlikely for false positive detections to be propagated to the next section, as they will not have a similar oriented correspondence in the other image. $\delta_{\leftarrow}(y_p, y_q)$ again is defined asymmetrically, such that only $E_{na}(0, 1)$ is penalized. In Equation (4.8) the corresponding sum runs over all neighbors $N_3(p)$, which are defined as neighbored pixels in adjacent sections (3 dimensional). To solve the correspondences between images we followed the non-linear warping method described in (Kaynig et al., 2008).

From our experience, the use of gradient flux is likely to lead to false positive membrane segmentations due to texture in the images.

Thus, we decided to omit gradient flux in the final energy term:

$$\begin{aligned}
 E(y) = & \sum_{p \in P} E_{rf}(y_p) + \lambda_{gc} \sum_{\substack{p \in P, \\ q \in N_2(p)}} E_{gc}(y_p, y_q) \\
 & + \lambda_s \sum_{\substack{p \in P, \\ q \in N_2(p)}} E_s(y_p, y_q) + \lambda_{na} \sum_{\substack{p \in P, \\ q \in N_3(p)}} E_{na}(y_p, y_q).
 \end{aligned} \tag{4.8}$$

Although this energy term incorporates information from adjacent sections, the main focus of the segmentation is two dimensional. This is due to the fact that the resolution of TEM images is high (about 5 nm per pixel), but along the vertical direction of the image stack, the resolution is limited by the section thickness of the sample. Even very skilled human operators can at best cut sections of 40 nm thickness. Thus, resolution along the z direction is an order of magnitude lower than the resolution along the x-y plane (see also Figure 4.9). This anisotropic resolution strongly favors a 2D segmentation approach.

4.1.5 Experiments and Results

We evaluate the proposed method on two different neuroanatomical data sets of ssTEM images. Data set 1 shows part of the dorsolateral fasciclin-II tract of the ventral nerve cord of the first instar larva of drosophila, at abdominal segment 5. It consists of 40 images with 512x512 pixels, divided into two sub volumes of 10 and 30 sections. The resolution is 3.7 nm per pixel in the image plane and section thickness is 50 nm. Data set 2 was taken from layer 4 of Area 17 (primary visual cortex) of one adult cat. The data set consists of 9 images with 4312x3018 pixels. Resolution is 1.38 nm per pixel in the image plane and section thickness is 40 nm. Both data sets resemble average image quality from neuroanatomy projects and were fully manually segmented by human experts using *TrakEM2* (Cardona, 2006), a free plugin for *ImageJ* (Abramoff et al., 2004). As can be seen in Figure 4.8, the membranes of data set 1 appear very dark in the images, but also fuzzy in a lot of areas. Data set 2 contains

considerably more texture caused by sub cellular elements like vesicles, microtubules and mitochondria inside the cells. Despite these different challenges, the new approach yields good segmentations on both data sets, demonstrating the great robustness against varying image characteristics.

In addition the proposed framework was applied to satellite images of San Francisco. The extracted features and the classifier employed for the segmentation of streets are the same as for the membrane segmentation, as the focus of the evaluation is on the different graph cut energy terms and not the quality of the classifier.

For the evaluation of the perceptual grouping framework all data sets were split into training and test sets. For the drosophila data set, the small volume was used for training and the large volume for testing. For the cat data set only nine annotated images are available, therefore leave one image out cross validation was used in this case. The random forest classifier ensemble consists of 500 trees. The trees were build with 10 out of 116 features randomly selected for each split. The plots in Figure 4.5 show the precision and recall of the segmentations on all test images. Here, precision can be seen as the probability that a pixel classified as foreground by the automatic segmentation is also marked as foreground in the hand labels given. Recall corresponds to the probability that a foreground pixel is detected. For the membrane segmentation on both data sets the perceptual grouping framework was evaluated with $\lambda_s = 0.6$ and for the evaluation of E_{na} , $\lambda_{gc} = 1.6$. For the state of the art segmentation with gradient flux λ_{gf} was set to 5. For the San Francisco street data set the parameters are $\lambda_s = 0.8$ and $\lambda_{gf} = 10$. The street data set contains no 3d information, therefore E_{na} is not included in the evaluation.

In all three data sets the good continuation energy term E_{gc} leads to a considerable improvement in recall. As can be seen in the example segmentations in Figure 4.8 the loss in precision is mainly caused by thicker membrane segmentations. For the 3d reconstruction of neuronal structures, high recall with closed contours is more desirable

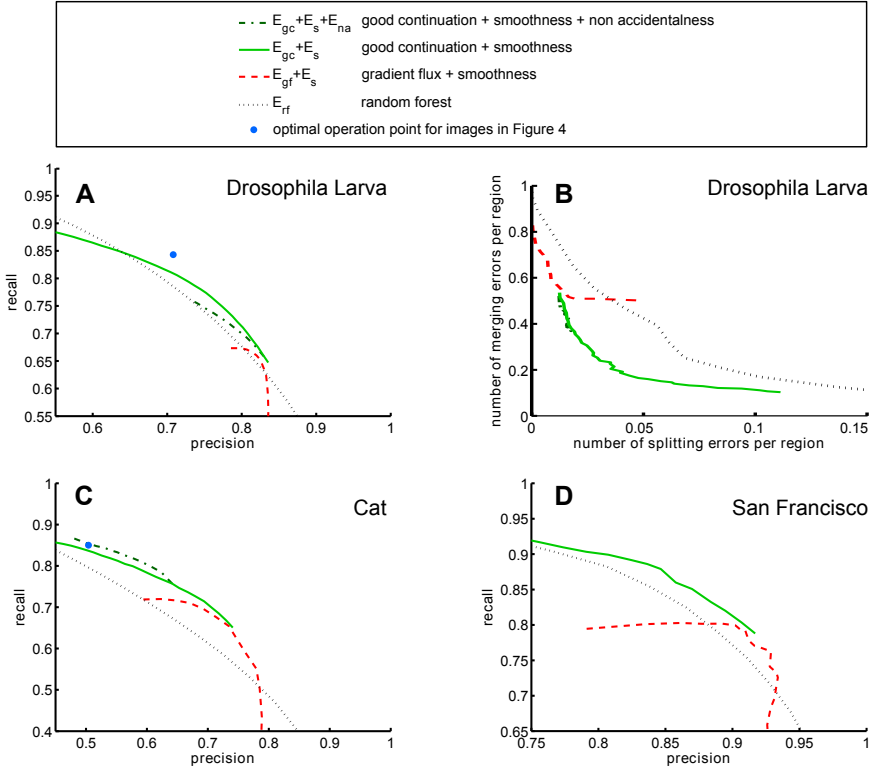


Figure 4.5: Comparison of the proposed framework $E_{rf} + E_{gc} + E_s + E_{na}$ against the state of the art gradient flux energy term $E_{rf} + E_{gf} + E_s$. The plots depict the precision and recall performance per pixel over all test images. The combination of random forests with perceptual grouping constraints yields a considerable improvement in recall. The split and merge error plot (B) demonstrates that the improved recall is caused by gap completion which is highly desirable for 3d reconstructions of neuronal structures.

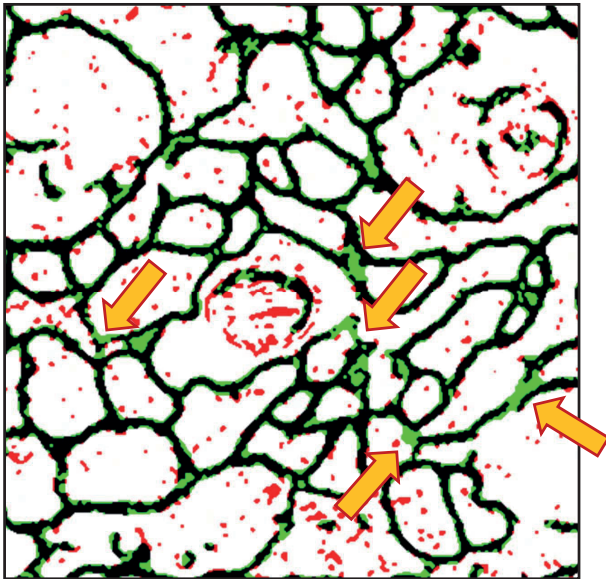


Figure 4.6: Example segmentation at 0.73 precision for an example image of the drosophila larva data set. Green: membrane detection by E_{gc} , red: membrane detection by E_{gf} , black: membrane detection by both methods. Arrows point to gaps which are close by the good continuation term.

than a good precision, as long as no splitting errors are introduced. Therefore, we also evaluate the number of splitting and merging errors per region for the drosophila larva data set (see plot B in Figure 4.5). The plot shows the number of splits and merges per region in the automatically obtained segmentation with respect to the manual ground truth. The splitting error counts the number of times a region from the ground truth segmentation is overlapped by more than one region from the automated segmentation. In order to be significant the split has to be bigger than one percent of the ground truth region. The merging error is the same in reverse. It counts how often a region of the automated segmentation is overlapped by more than one region in the ground truth. Thus, the error is increased if a segmented membrane is not closed and ground truth regions are merged in the automated segmentation. A low splitting and merging error per region preserves the duality between membranes and enclosed regions and thus enables automatic reconstructions of neuronal structures. Plot B in Figure 4.5 clearly demonstrates the substantial improvement in the segmentation by our good continuation term. Figure 4.6 provides a qualitative example for the gap closing performance. The cat brain data set does not contain enough regions to provide meaningful results in terms of splitting and merging errors, due to the large size of the neuronal structures in these images. The term E_{na} that incorporates information from adjacent sections, is very beneficial for the cat data set and leads to an additional increase in recall. For the drosophila data set, the influence of adjacent sections is smaller than for the cat data set because the drosophila images change significantly between sections.

Example segmentations of test images are given in Figure 4.8. Most membranes are correctly segmented and the segmentation is very good with respect to texture caused by vesicles and microtubules. Mitochondria still pose a challenge. They are not only surrounded by a membrane, but also very similar to small dendrites in shape, leading to false positive detections. A possible solution to this problem would be to include extra labels for mitochondria in the training set

and either make the random forest classifier more sensitive to these structures or train a second classifier specifically for mitochondria. This is part of our future research. A segmentation result for the San Francisco data set is given in Figure 4.7. Shown are the segmentation results with 0.85 precision for smoothness combined with good continuation (green) or gradient flux (red). Black pixels were marked as streets by both methods. Although this image is from a completely different domain, the segmentation result shows the same characteristic for both methods as for the electron microscopy images. The good continuation constraint leads to thicker segmentations, but improves the segmentation by gap completion, whereas the gradient flux gives false positive responses at background pixels with high contrast.

The split and merge error of our cost function is low enough, to obtain fully automatically reconstructed dendrites over several sections. An example reconstruction is shown in Figure 4.9. The five dendrites are segmented over 30 sections. Regions were automatically grouped between sections by maximum overlap. This simple tracking method will fail if the structures of interest are not orthogonal to the cutting direction. Also shown in this Figure are cutting planes through the image volume. The very low resolution of the volume in the direction orthogonal to the cutting plane is clearly visible. Because of this difference in resolution we decided to focus our segmentation on the image plane.

4.1.6 Conclusion

The framework introduced in the sections above addresses one of the main bottlenecks for 3D reconstructions in neuroanatomy: the fully automated segmentation of membranes in ssTEM images. The architecture comprises a random forest for classifying single pixels, and novel energy terms for membrane segmentation with graph cut optimization. Large scale quantitative evaluation experiments demonstrated the algorithms performance on cat and drosophila larva brain.

In summary the proposed framework is characterized by the following benefits: (i) *local to global optimization*: a random forest clas-

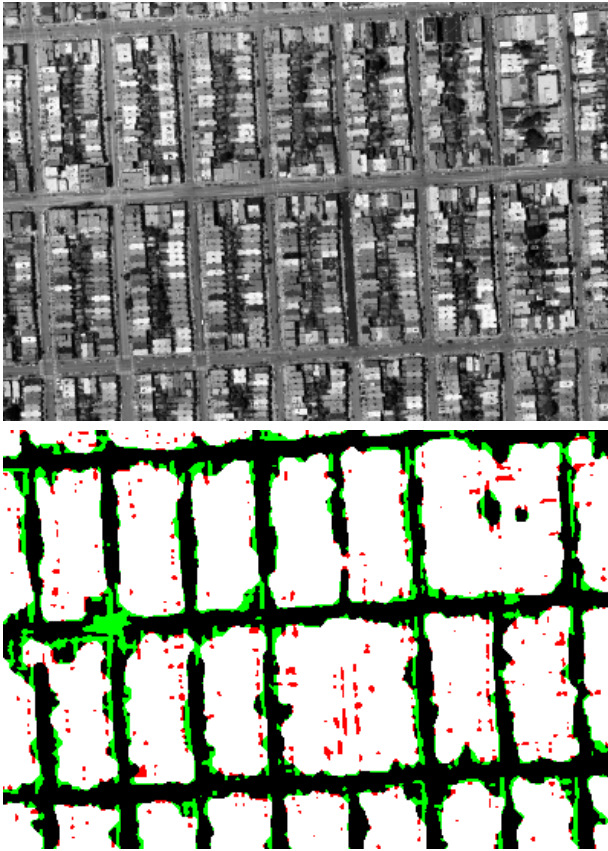


Figure 4.7: Example segmentation at 0.85 precision for the San Francisco street data set. Green pixels are positive detections with the good continuation constraint, red pixels are positive detections by smoothness and gradient flux, black pixels were marked by both methods as streets. Segmentation by good continuation loses precision by thickening the detected streets, but gains additional recall by gap completion. Gradient flux loses precision by false positive detections at high gradient contours.

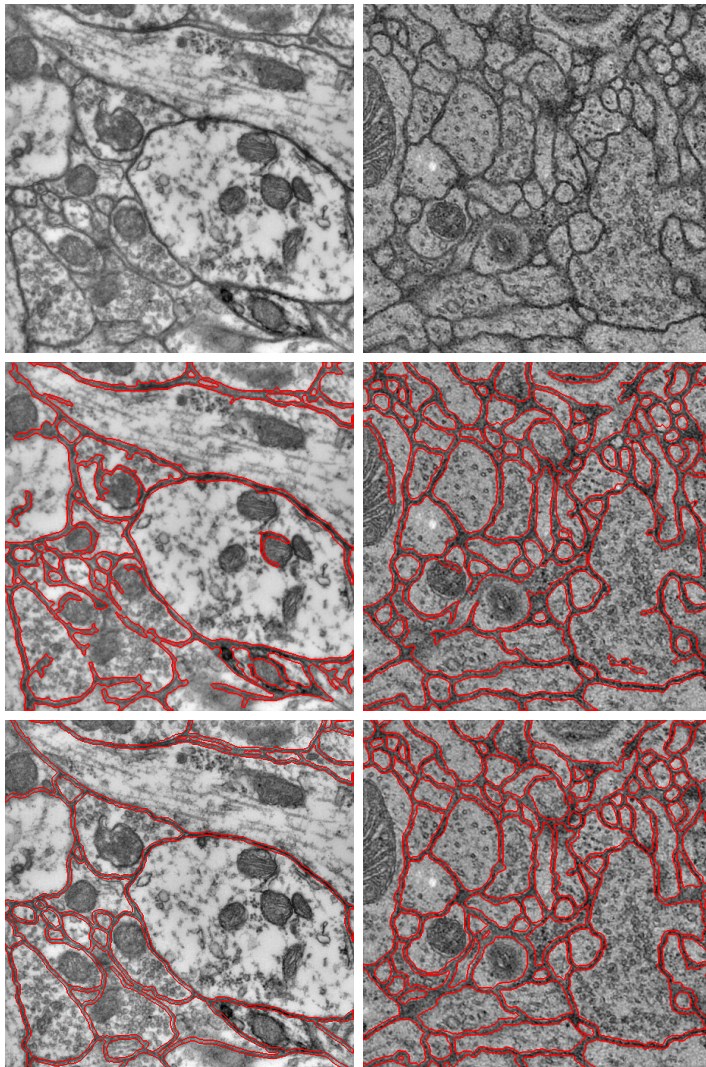


Figure 4.8: Example images and segmentations from two data sets. **Left:** drosophila larva, **right:** cat. From top to bottom: original image, automatic segmentation with perceptual grouping constraints, manual labels.

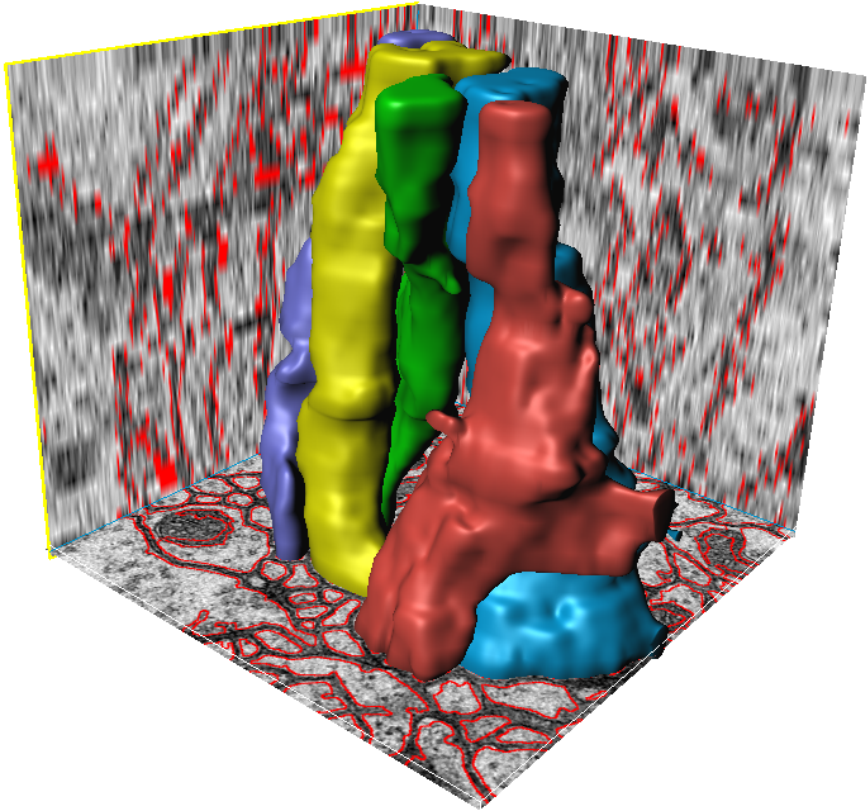


Figure 4.9: Five dendrites fully automatically reconstructed over 30 sections from drosophila larva. The cutting planes surrounding the dendrites show the good resolution of the data volume in xy-direction in contrast to the low resolution in z-direction.

sifier estimates the probability for a membrane locally, while a regular cost function guarantees a global optimum employing graph cuts. (ii) *good continuation*: novel energy terms stabilize contour completion in situations where gradient flux based methods fail. (iii) *robustness*: the algorithms produces proper results even on different animal species. (iv) *consistency*: we have successfully reconstructed a 3D model of dendrites based on the consistent segmentation of an image stack with 30 slices. (v) *excellent performance*: the presented algorithm outperforms the state of the art on all quantitatively evaluated real world scenarios.

4.2 Region Grouping

Segmentation of neuronal structures from ssTEM images is usually performed in two dimensions, using the fine resolution to identify membranes of neuronal processes like dendrites and axons (Kaynig et al., 2010c; Mishchenko, 2009; Reina et al., 2009; Kannan et al., 2009). The regions surrounded by the detected membranes then need to be grouped over consecutive sections to extract the geometry of neuronal processes. Figure 4.10 depicts examples of correspondent regions from adjacent sections. The grouping of segmented regions to 3d objects is challenging, as all structures change in their appearance between the sections. Thin processes are especially difficult, because their flexibility is larger than the variability of large regions. In some cases the correspondent regions do not overlap between the sections. Previous work has addressed the grouping problem by tracking single processes through the image stack (Jurrus et al., 2008, 2009). We extend the previous approaches with respect to three important points: (i) instead of tracking single processes the labeling of the whole data volume is optimized, allowing for neuronal processes to start or end inside the volume, (ii) similarity of regions is learned from annotated data, (iii) geometrical consistency between whole sections is taken into account.

We regard the problem of three dimensional geometry extraction

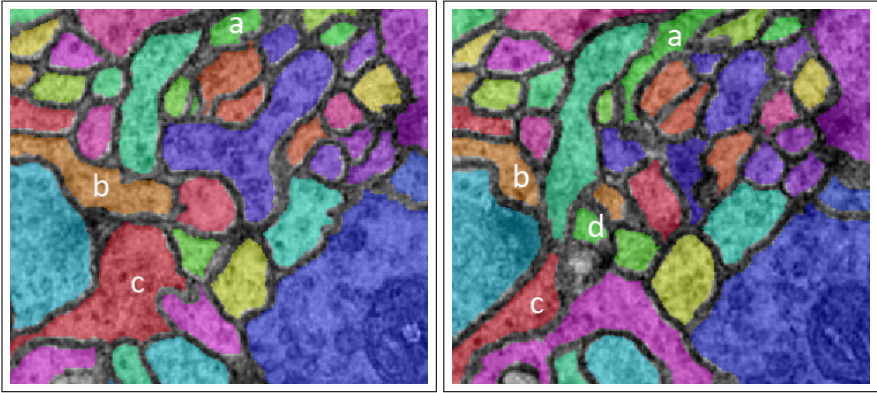


Figure 4.10: Example groupings of regions from two adjacent section images (correspondence is indicated by color). Structures running longitudinal to the cutting plane, express significant changes in appearance between sections (a-c). Example d has no correspondence in the left section.

as partitioning an edge weighted graph into connected components representing an image volume belonging to the same neuronal process. The regions are represented by the vertices V of the graph and the set of edges E connects each region to all regions of the two adjacent sections. Each edge is assigned a weight w_{ij} according to the similarity between regions i and j .

We propose the following processing pipeline to build the edge weight matrix W and to find connected components representing neuronal processes (see Figure 4.11). First, a set of weight matrices based on features like region overlap or similarity of texture is created. A detailed description of the features is given in Section 4.2.1. A random forest classifier is trained on manual annotations to predict the similarity between two regions based on the extracted features. The weight matrix predicted by the random forest classifier only captures the similarity of pairwise regions and does not take the geometry of contiguous processes into account. Therefore, a further step refines

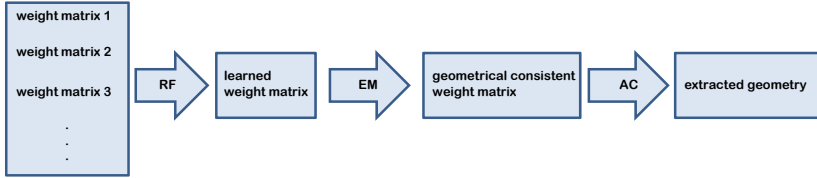


Figure 4.11: Processing pipeline for the extraction of 3d geometry of neuronal processes as proposed in this paper. First, a similarity matrix of pairwise regions is learned by a random forest classifier. The learned weight matrix is then combined with geometrical constraints, taking the geometry of all neuronal processes from the whole section into account. Optimization is performed by expectation maximization. Finally agglomerative clustering is used to extract continuous neuronal processes.

the weight matrix using geometrical consistent constraints that take the geometry of all neuronal processes included in the section into account. Finally, agglomerative clustering is employed to partition the graph into connected components representing neuronal processes. The hierarchical clustering scheme starts from individual objects and then progressively merges the regions which are most similar to each other. This system mirrors the approach of the neuroanatomist, who first establishes correspondences between regions that are easy to detect and then refines the partitioning.

4.2.1 Similarity Measures between Regions

The following paragraphs describe the features that are used to train the random forest classifier from manually annotated data. For each feature we build a weight matrix W , each entry representing the edge weight of the corresponding edge in the graph.

Euclidean distance of region center: Each region i is represented by its center of mass $c_i \in \mathbb{R}^3$. The distance of two regions is then given by the Euclidean distance between the two centers:

$$W_{distance}(i, j) = \sqrt{(c_i - c_j) \cdot (c_i - c_j)^T} \quad (4.9)$$

Overlap of region areas: For each region i , the set P_i contains the position of all pixels belonging to the region ($P_i \in \mathbb{R}^3$). The overlap of two regions is measured by projecting both regions orthogonally to the same plane and building the intersection of both projections:

$$W_{overlap}(i, j) = \#(P_i \cdot A \cap P_j \cdot A), \text{ with } A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (4.10)$$

Difference in region size: Neuronal processes have only smooth variations in diameter. Therefore the size of corresponding regions should be similar to each other as long as the neuronal process is running in a direction orthogonal to the cutting plane.

$$W_{size}(i, j) = \frac{(\#P_i - \#P_j)^2}{\#P_i + \#P_j} \quad (4.11)$$

Here $\#P_i$ describes the size of region i in number of pixels. The size difference between two regions is measured by the squared difference in pixels divided by the total size of both regions. This normalization accounts for the comparability of processes with large or small diameter.

Texture similarity: For the neuroanatomist, texture is an important clue for the extraction of neuronal processes. Intracellular structures like mitochondria, vesicles or microtubules provide information about the type of neuronal process, e.g. bouton or axon, and about the consistent grouping of regions. Following the approach described in (Jurrus et al., 2008), we measure the similarity in texture

by the cross correlation coefficient of two regions

$$W_{xcorr}(i, j) = X_{\max}(r_i, r_j). \quad (4.12)$$

Where r_i represents the gray values of region i and X_{\max} denotes the maximal cross correlation between the two regions. For computational efficiency the cross correlation is computed in the Fourier domain.

Smooth continuation: This feature weights the connection between two regions i and j according to the smoothest continuation to the next sections. The smoothness of a possible continuation is given by the angle θ_{hij} between the three region centers c_h, c_i and c_j (see Figure 4.12).

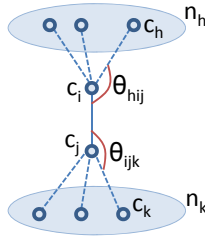


Figure 4.12: Illustration of the smooth continuation feature. The connection between region i and j is weighted according to the smoothest continuation to the next sections. The smoothness of a possible continuation is given by the function $\theta(c_h, c_i, c_j)$ which measures the angle between the three region centers c_h, c_i and c_j .

$$W_{smooth}(i, j) = \frac{1}{2} \cdot (\min_{h \in n_h} \theta(c_h, c_i, c_j) + \min_{k \in n_k} \theta(c_i, c_j, c_k)), \quad (4.13)$$

with $\theta(c_h, c_i, c_j) = \text{abs}(\pi - \angle(c_h, c_i, c_j))$. The set n_h contains all regions from the section above region i and the set n_k contains all

regions from the section below region j . Thus, the weight between two regions is small when there exist smooth continuations at both ends of the connection. Reflection is employed as border treatment to compute the smooth continuation feature for the first and last section of the stack.

4.2.2 Global Geometrical Consistency

So far the single edge weights defined by each feature are focused on the two regions being connected. But, region correspondences should also be assigned in consistency with the overall geometry changes from one section to the next. For example, a clear shift of several adjacent processes from one section to the other cannot be detected by local features and thus would likely lead to false correspondences. We address this problem by establishing geometrical consistency of the correspondences between sections. The approach implements a non-linear but smooth transformation between sections to match correspondent points. Similar to the warping method described in Section 3.6, correspondences are not fixed beforehand, but obtained during the optimization. For the non-linear transformation we use an explicit polynomial kernel expansion to map the points c_i into a higher dimension

$$\phi(c_i) = [1, c_{i1}, c_{i2}, c_{i1}^2, c_{i1}c_{i2}, c_{i2}^2, \dots, c_{i2}^d]^T \quad (4.14)$$

A transformation matrix β is defined to project these points back into the image plane, leading to a non-linear transformation. Correspondences are assigned by a binary matrix M whose entry m_{ij} is one, if point c_i in one section corresponds to point c_j in the adjacent section and zero otherwise.

The energy function to be optimized depends on the similarity of the correspondent regions as classified by the random forest, as well as on the quality of the geometric fit:

$$E(\beta, M) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} -m_{ij} \|\phi(c_i)\beta - c_j\|^2 + m_{ij} \cdot \ln(W(i, j)) \quad (4.15)$$

Here the index i runs over the number of regions n_i from one section and j over the number of regions n_j from the adjacent section. The variable m_{ij} contains the associated value of the assignment matrix M and $W(i, j)$ corresponds to the similarity edge weight given by the random forest classifier. Maximizing this energy function can be interpreted as maximizing the data likelihood $p(C_i, C_j | \beta, M)$ where C_i and C_j are matrices containing all points from two adjacent sections. We use expectation maximization to optimize the joint log-posterior, treating the correspondences as unobservable. The algorithm iterates between estimating the expectation of the latent variables m_{ij} while keeping β fix and maximizing the joint log-posterior while keeping the expectation values of M constant.

E-step: In each iteration the variables m_{ij} are replaced by their conditional expectation given β . The expectation values are calculated using the currently optimized β . Under the condition that M is a valid assignment matrix

($\sum_j^{n_2} m_{ij} = 1$, for all $i = 1, \dots, n_2$), we derive the following result:

$$\gamma_{ij} = \mathbf{E}[m_{ij} | C_i, C_j, \beta] = \frac{p(C_i, C_j | \beta, m_{ij} = 1)}{\sum_{l=1}^{n_2} p(C_i, C_j | \beta, m_{il} = 1)} \quad (4.16)$$

M-step: The expectation of the joint log posterior has the same form as the joint log posterior itself, but with m_{ij} replaced by γ_{ij} . Under the assumption that β is smooth, i.e the components of β are assumed to be normally distributed, maximizing for β yields a weighted ridge regression problem with weights γ_{ij} :

$$\beta \leftarrow (\phi(\widetilde{\mathbf{C}}_i)^T \mathbf{\Gamma} \phi(\widetilde{\mathbf{C}}_i) + 2\lambda \mathbf{I})^{-1} \phi(\widetilde{\mathbf{C}}_i)^T \mathbf{\Gamma} \mathbf{C}_j \quad (4.17)$$

where Γ is a $(n_i \cdot n_j) \times (n_i \cdot n_j)$ -dimensional diagonal matrix of the weights γ_{ij} . The $(n_i \cdot n_j) \times 2$ matrix $\widetilde{\mathbf{C}}_i$ contains n_j copies of each center point c_i from the first section and the $(n_i \cdot n_j) \times 2$ matrix \mathbf{C}_j contains n_j possible correspondence points from the adjacent section for each point c_i . The parameter λ is the regularization parameter defined by the prior distribution $p(\beta)$. In our experiments λ is set to 0.001.

4.2.3 Evaluation

The proposed method is evaluated on neuroanatomical ssTEM images, resembling average image quality from neuroanatomy projects. The data set depicts part of the dorsolateral fasciclin-II tract of the ventral nerve cord of the first instar larva of drosophila, at abdominal segment 5. It consists of 30 images with 512x512 pixels. The resolution is 3.7 nm per pixel in the image plane and section thickness is 50 nm. The whole data set was annotated exclusively by a neuroanatomist, providing the ground truth for the evaluation. The random forest classifier was trained on this data set using ten fold cross validation to obtain the test error. The remaining pipeline is free of tuning parameters and therefore just applied to the test results of the classifier.

As demonstrated by the plots in Figure 4.13, each step of our processing pipeline yields significant improvement for the geometry extraction in terms of split and merge error per object. The split error counts the number of ground truth labels assigned to each cluster by the automatic approach. A perfect solution would assign exactly one label per ground truth cluster. For each additional label a split error is counted. A merge error occurs when two clusters from the ground truth labeling are assigned the same label. If two ground truth clusters are merged more than once, we follow the definition of Turaga et al. (2010) and count this as one error as the same two objects are involved.

The agglomerative clustering is restricted to establish a maximum

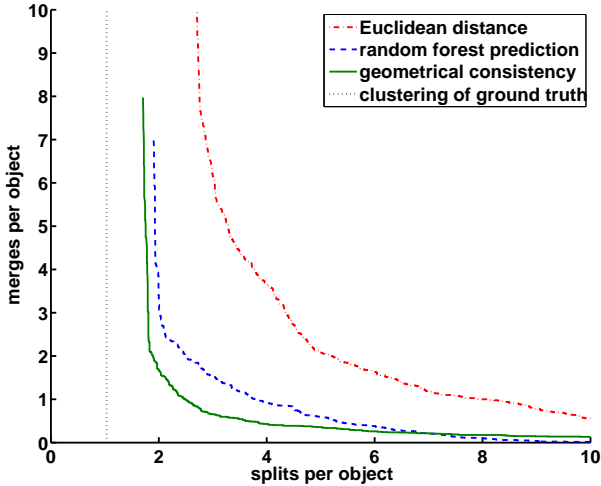


Figure 4.13: Evaluation of clustering results according to split/merge error per neuronal process. Depicted are the results for different weight matrices: (i) Euclidean distance of region centers only (dot dashed line), (ii) weights learned by the random forest classifier (dashed line), and geometrical consistent weights (solid line). The random forest is evaluated using ten-fold cross validation. The plot clearly demonstrates that each step of our processing pipeline yields significant improvement for the geometry extraction. The dotted line corresponds to the best result obtainable without considering branching processes.

of two correspondences for each region, one to the upper and one to the lower section. Thus, our model allows for starting and ending of new neuronal processes inside the volume, but does not account for branching of processes. The dotted line in Figure 4.13 marks the best clustering performance achievable by this model.

Examples of extracted geometries are given in Figure 4.14. The examples demonstrate, that the proposed method is capable of extracting correct geometries also in difficult cases of neuronal processes running longitudinal to the cutting plane and in cases of discontinuities in the geometry due to alignment errors.

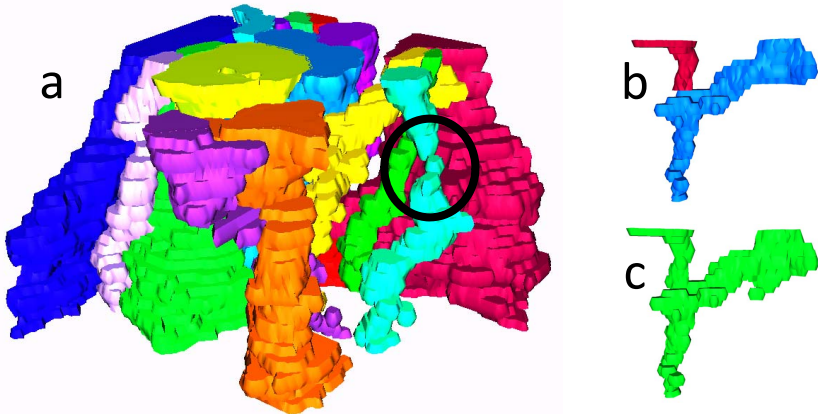


Figure 4.14: The 3d reconstruction shows all parts of neuronal processes that were correctly tracked over all 30 sections from the data set. A black circle marks an example where regions were correctly grouped despite not having any overlap in adjacent sections. The neuronal process shown in detail in Figure b (ground truth) and Figure c (clustering result) shows an example for a split. The large part including regions moving longitudinal to the cutting direction was correctly grouped by the proposed method and the remaining part was also grouped together as one object.

4.2.4 Conclusion

We introduced a novel framework for global tracing of neuronal processes in stacks of serial section transmission electron microscopy images. The setting is formulated as a partitioning problem on edge weighted region-graphs.

The main contributions of this work are threefold: (i) On the modeling side we propose the use of a random forest classifier to learn a predictor for neighborhood relations of regions within the 3d volume. (ii) Predicted region correspondences are refined taking the geometrical consistency of whole sections into account. (iii) The unsupervised clustering approach results in a robust procedure for partitioning the graph.

In depth evaluation of all single steps of the pipeline and cross validation of the similarity classification demonstrate significant improvement in terms of split and merge error per object.

Chapter 5

Functional Structure Annotation

Intracellular structures, like vesicles and mitochondria play an important physiological role in neuroanatomy. Geometry extraction of neurites could be interpreted as extracting the wires from the neuronal network, but only through intracellular structures, the functional parts of the wires can be identified. Vesicles, for example, are a strong indicator that an axon is forming a bouton. Furthermore, the presence of vesicles is a necessary requirement for a synapse. We distinguish two different tasks for structure annotation:

- Annotation of *basic* structure units, for example vesicles, mitochondria, and membranes.
- Classification of *compositional* structures, e.g. synapses or boutons.

The following Chapter describes our framework to solve the tasks above.

5.1 Basic Structure Annotation

There are four basic cellular structures, which are prominent in EM images and are of great importance for neuroanatomy: cell membranes, vesicles, mitochondria, and myelin sheaths. Figure 5.1 depicts examples for these structures.

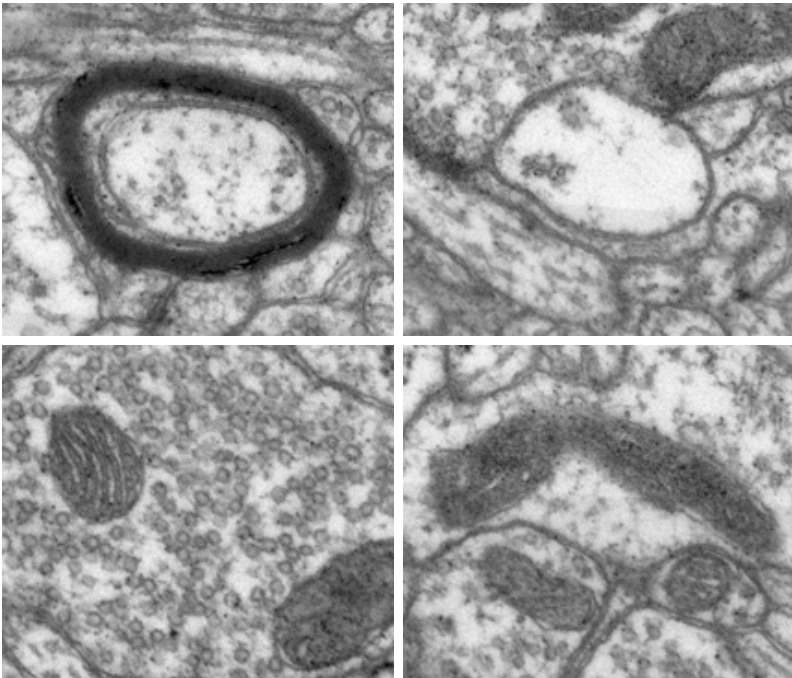


Figure 5.1: Examples for different structures of interest for annotation, upper left: myelinated axon, upper right: dendrite surrounded by normal cell membrane, lower left: patch of vesicles, lower right: mitochondria.

Segmentation of cell membranes, as discussed in detail in Section 4.1, is required for neuron geometry extraction. While mitochondria are on their own objects of interest for neuroanatomy, they do not

directly contribute to functional structure annotation. However, it is beneficial to segment mitochondria in the images to reduce false positive detections of other intracellular structures. Segmentation of mitochondria has been successfully demonstrated using ray and radon like features (Lucchi et al., 2010; Kumar et al., 2010). Myelin sheaths are very pronounced, dark structures and can be filtered with local edge histograms (Jeong et al., 2009). To the best of our knowledge no work has been performed so far and reported in the literature on the detection of single vesicles. The following subsection describes our approach to detect vesicles in TEM images.

5.1.1 Vesicle detection

Vesicles are small spheres filled with neurotransmitters, which can be released at a synapse by the presynaptic neuron. Therefore, vesicle detection is highly significant for the identification of boutons and synapses in EM images. As can be seen in the lower left part of Figure 5.1, vesicles appear as dark circles in the images and often have a bright center. A very simple approach to vesicle detection is to generate a template by averaging different vesicle appearances in the image and then perform template matching based on normalized cross correlation. In our experiments, this approach leads to poor results due to variations in the vesicle shape and false positive responses at membranes and mitochondria. To enhance the detection rate, we employ the following set of features to capture discriminative characteristics of vesicles:

- gray value in the center
- mean gray value of the vesicle area
- difference of center gray value to the vesicle frame value
- distance of center pixel to nearest local gray value minimum along all four vertical axes

- minimal, maximal and average response of the template matching output in the vesicle area
- minimal, maximal and average value of the large Hessian eigenvalue in the vesicle area
- smoothed gradient response at the center pixel
- minimal, maximal and average value of the accumulation area of a Hough transform in the vesicle area

The feature set is fast to compute and focuses on the shape information which discriminates vesicles from other intracellular structures. To train a classifier on this set of features, supervised information is necessary. But, vesicles are very small structures and therefore unguided user annotation is bound to be imprecise. We propose to overcome this problem, by a two stage approach. First, points of interest are identified by a Hough transform looking for circles based on the image gradient. These points of interest are then classified according to the features listed above. The advantage of the interest point detection is that manual labels can be given on the classification of the detected circles. In addition, the estimated radii of the circles can be employed to specify the area of a vesicle and region of interest for features extraction.

Figure 5.2 demonstrates the vesicle detection on a test image. A random forest classifier was trained on 200 annotated examples of circles detected in a training image. The training samples are randomly chosen and typically contain 20-25% positive examples, leading to unbalanced training data. There are different methods to compensate for unbalanced training data when training a random forest classifier. These include reweighting of the samples, oversampling of the small training set or downsampling of the large training set. We follow the approach of Chen et al. (2004) who proposed to combine the down sampling technique with the ensemble idea of the random forest classifier. They down sample the minority class and grow each tree on

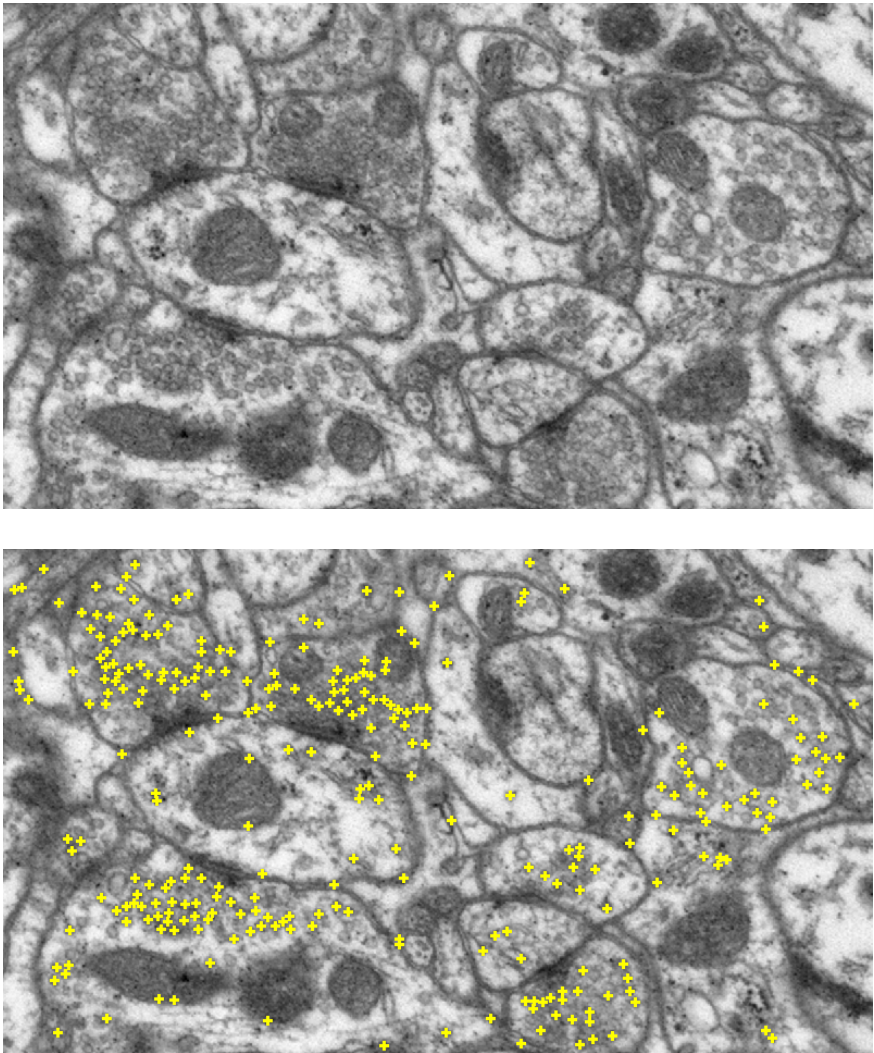


Figure 5.2: Example of the vesicle detection on a test image. Top: original image, bottom: detected vesicles marked with a yellow cross.

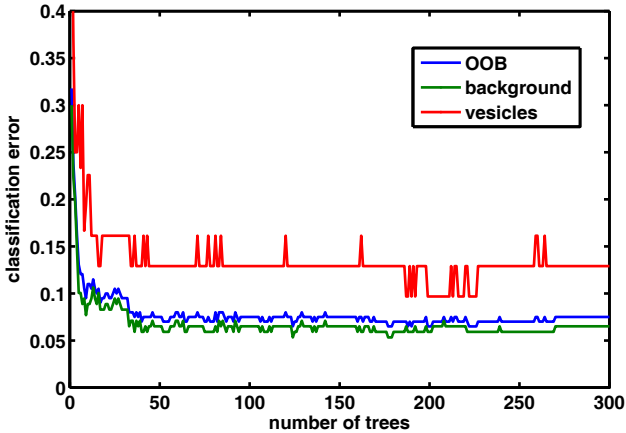


Figure 5.3: The out-of-bag error of the random forest classifier trained for vesicle detection.

a more balanced data set . Figure 5.3 contains the out-of-bag error of the random forest classifier and the class specific error rates for the vesicle and the background class. The detection performance is reasonable, given the limited training data and the image quality. But, the detection rates are not sufficient for automatic counting of vesicles. However, the overall performance is sufficient to provide a valuable feature for the detection of synapses.

5.2 Synapse Detection

At a synapse, the axon terminal of one neuron comes into functional contact with a second neuron. Analyzing these connections is regarded as one of the keys to understanding the functional structure of the brain. Thus, the detection of synapses is crucial for the field of neuroanatomy. A synapse between two neurons typically consists of three parts: the presynaptic side, the synaptic cleft, and the post-synaptic side. The presynaptic part contains the neurotransmitter

enclosed in vesicles. When a nerve impulse reaches the synapse, the neurotransmitter is released into the synaptic cleft, which is a gap between the two neurons. The neurotransmitter then binds to the receptors on the postsynaptic membrane, triggering an action potential on the second neuron.

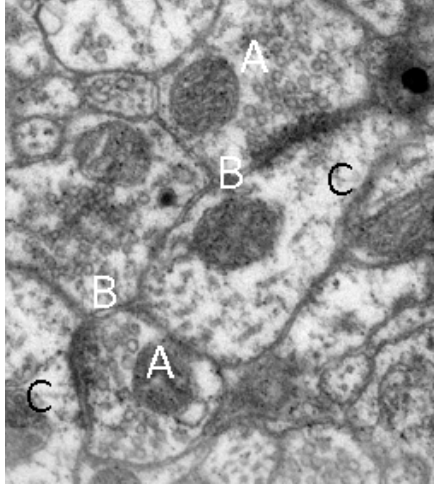


Figure 5.4: Synapses in a TEM image. Annotated are the vesicles (A), the synaptic cleft and postsynaptic density (B), and the receiving part of the second neuron (C)

Figure 5.4 depicts a typical example of a synapse in a TEM image. Only the presynaptic part contains vesicles. The postsynaptic membrane appears thicker and darker in the image than other extracellular membranes. This electron dense region is called the postsynaptic density. Close geometrical configurations of boutons, containing vesicles, and dendrites are possible candidates for contacts between two neurons, but the expression of a postsynaptic density is a necessary requirement for a valid synapse. The synaptic cleft is visible as a thin bright line between the presynaptic and postsynaptic membrane.

To capture the main components of a synapse, we designed a

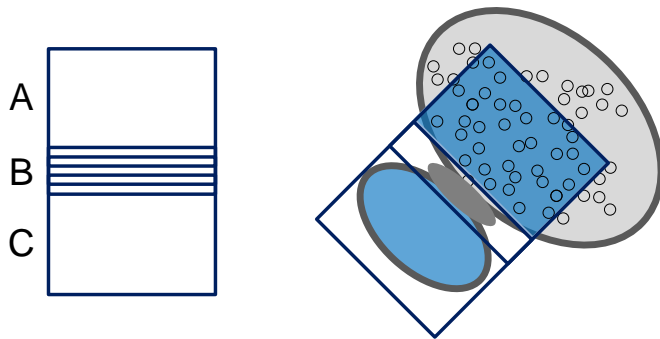


Figure 5.5: Feature extraction template for synapse detection.

feature extraction template, as depicted in Figure 5.5. We extract features to identify the presynaptic and post synaptic neuron parts for regions A and C of the template. Region B covers the area of the postsynaptic density and the synaptic cleft. As demonstrated in the right part of Figure 5.5, the template is aligned along to the orientation of the intracellular membrane. If region A or C are larger than the neuronal region adjacent to the membrane, features are only extracted from the small area of this neuronal region. If the neuronal region is larger than area A or C of the template, features are extracted from the area covered by the template. This way, vesicles are only considered for a synapse if they are in close proximity to the membrane. The following list contains the features extracted from region A and C:

- size of the feature extraction area in pixels
- mean, variance, minimum, maximum, and histogram bins of gray values
- weighted histogram of gradient orientations (HOG)

- mean, variance, minimum, and maximum of the structure tensor at different scales
- mean, variance, minimum, and maximum of the gradient magnitude
- mean, variance, minimum, maximum and sum of the vesicle detector output and the output thresholded at 0.5
- squared difference between the features listed above for region A and C of the feature extraction template

In addition to the output of the vesicle detection, the features for template regions A and B focus on the gradient directions. For a patch of vesicles the orientation of the gradient vector field is uniformly spread over the whole orientation range. Microtubuli instead often give rise to elongated textures, leading to dominant directions in the gradient orientation.

To capture the characteristics of the postsynaptic density and the synaptic cleft, we employ the following feature set for template region C:

- mean gray value of the whole region
- median gray value along the single stripes parallel to the membrane (see Figure 5.5)
- position of first local gray value minimum on the first half of the stripes
- position of first local gray value minimum on the second half of the stripes
- distance between the two local gray value minima
- maximal gray value in between the two local gray value minima

The postsynaptic density gives rise to a wide dark region, which is captured by the mean gray value of region C. The rest of the features set is designed to capture the orthogonal profile of the synaptic cleft. The presynaptic and postsynaptic membranes give rise to local minima in the gray value profile, whereas the cleft itself appears as a bright stripe between the two membranes.

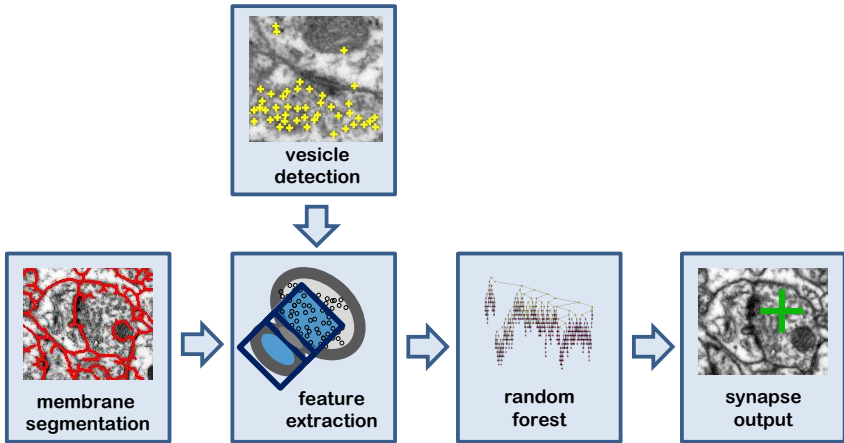


Figure 5.6: Workflow for our synapse detection framework.

The diagram in Figure 5.6 contains the whole workflow for the synapse detection. First, intracellular membranes are segmented according to our method discussed in Section 4.1. Then, the features listed above are extracted and a random forest classifier is trained on manual synapse annotations. As described already for the vesicle detection in Section 5.1.1, the training data is biased, as only about 20% of all membrane pixels belong to synapses. Again we employ sub sampling of the majority class independently for each tree of the random forest to compensate for this imbalance.

To evaluate our synapse detection framework we employ a dataset

of 40 consecutive cat brain TEM images of 1654×1051 pixels. The data set contains a minimum of 10 and a maximum of 20 synapses per image, with a median of 15. Manual annotation was independently performed by two neuroanatomists. We take one set of labels as gold standard and take the second set of labels as comparison for human performance on the data set. The images were corrected against lens distortions, stitched and aligned. On five images from the stack we made annotations to provide training labels for membrane and vesicle classification. The segmentation of membranes was enhanced by enforcing gap completion (see Section 4.1.2). As the output of the classifier gives a synapse probability for each membrane pixel, positive detections have to be grouped together to count for a synapse. Thus, we introduce a postprocessing step, in which detected synapse membrane pixels are grouped together and single detections are deleted by morphological operations. The values for the range of the grouping and the small region filtering are estimated using cross validation. From our 40 image data set, nine randomly chosen images are taken as a validation set for the postprocessing parameters and one image as test data. The remaining 30 images form the training data for the classifier. Figure 5.7 demonstrates the median precision and recall performance taken over all images.

We compare our feature set to state of the art synapse detection methods, which solely concentrate on the postsynaptic density as the identifying feature Mishchenko et al. (2010); Kreshuk et al. (2011). While these features are very fast to compute, our experiments demonstrate that taking pre- and postsynaptic regions into account leads to a significant gain in the detection. At the precision rate achieved by an independent human expert (green dot), our method outperforms the state of the art approach by a factor of two. Figure 5.8 depicts the detection performance on an example test image from the data set. There are two false positive detection examples in this image. The red dot in the lower right part of the image marks a false positive detection where part of a mitochondrion in a bouton was confused with a postsynaptic density. In the upper right corner

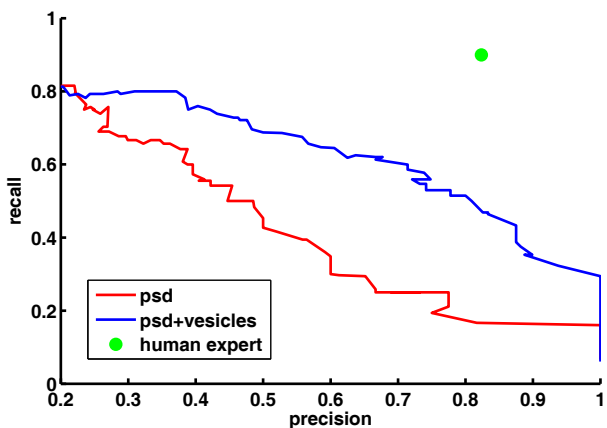


Figure 5.7: Precision recall curves for synapse detection. Red curve: performance of state-of-the-art method, concentrating on postsynaptic density and the synaptic cleft only (psd), blue curve: our complete feature set taking characteristics of pre- and postsynaptic regions into account, green dot: human expert performance

the false positive detection is caused by a fuzzy membrane between a bouton and a dendritic process. Figure 5.9 depicts some additional

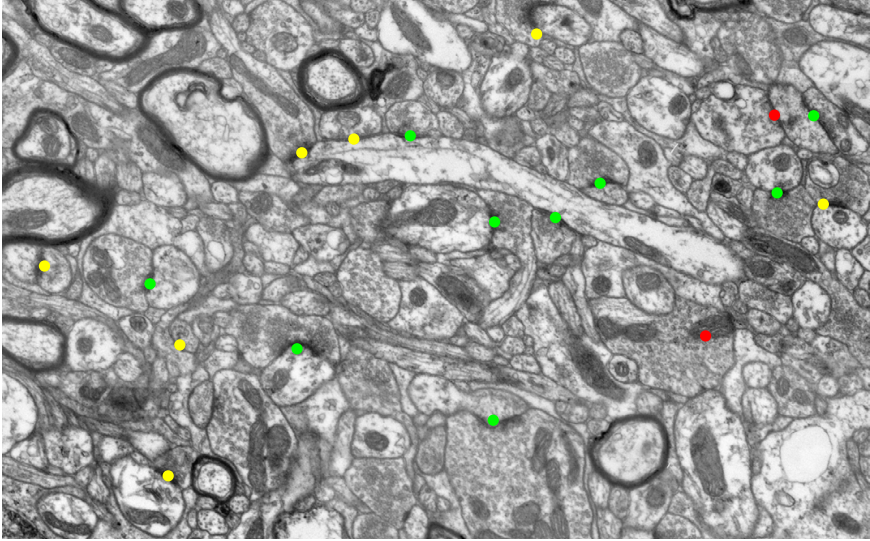


Figure 5.8: Example image with detected synapses (precision: 0.7, recall: 0.43) . Green: true positive detections, yellow: false negative detection, red: false positive detections.

examples of false positive detections caused by fuzzy membranes. The examples in the top row contain all a bouton and dendritic part of a synapse, but do not express a postsynaptic density. For comparison the examples in the bottom row show true positive detections with postsynaptic densities.

While the performance of the synapse detection is very promising, the results are not good enough for a fully automatic synapse detection. We are planning to employ the synapse detection framework so far as an interest point detector in a semi automatic scenario to speed up the labeling of large data sets.

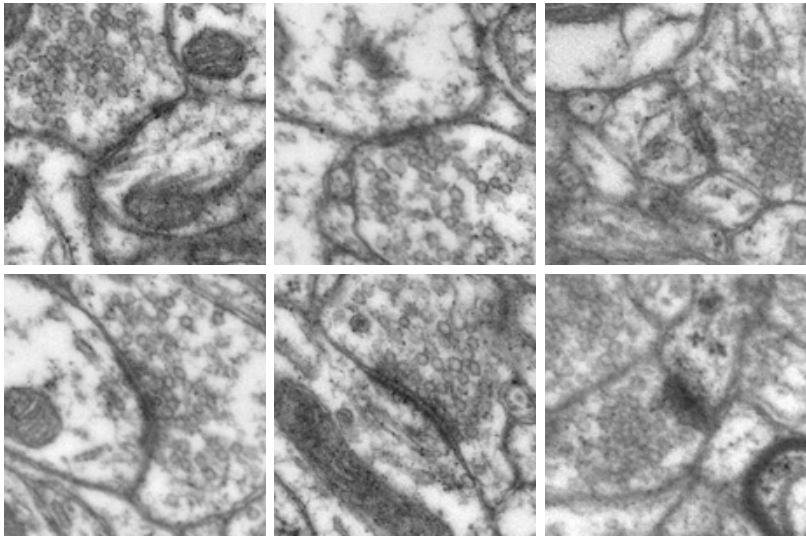


Figure 5.9: Examples of false positive (top row) and true positive (bottom row) detections thresholded at 0.5. For the examples shown, both neuroanatomists independently provided the same label.

Chapter 6

Conclusion

In this thesis we address the whole image processing pipeline for computational neuroanatomy and connectomics. For image stitching and alignment we propose unsupervised approaches to identify artifact signals like lens distortions or staining blurs in the images. For membrane segmentation and geometry extraction we propose feature sets, which can be employed to train a random forest classifier with very little user annotations, rendering our framework adaptable to different data sets with minimal effort. The output of the classifier is combined with different smoothness constraints, which are biologically inspired and therefore independent of staining protocols or animal types. The developed approaches are applicable to a wide range of data sets and have been demonstrated to outperform state of the art methods for all steps of the image processing workflow. The proposed pipeline yields fully automatic reconstructions of neuronal processes over 30 sections. However, dense reconstruction of a volume needs additional proof reading steps for the segmentation of membranes and the extraction of the geometry from the segmentations. In a software implementation the costs of the geometrical consistent region assignment matrix can be employed to detect inconsistencies in the segmentation in order to guide the user interaction. Since manual annotation of electron microscopy images is very time-consuming and

identification of synapses needs to be done by trained experts, the annotated data sets currently available for neuron reconstruction and synapse detection are very limited. With the methods proposed in this thesis, manual user interaction is greatly reduced, enabling annotation of large data sets with little effort compared to manual annotations. For geometry reconstruction, the segmentation of membranes is the part of the processing pipeline with the greatest potential to improve the overall performance for dense reconstruction. As adjacent sections contain valuable information, future work could compose the non-linear registration and membrane segmentation into a unified framework. In addition the grouping of corresponding regions across sections needs to be enhanced to account for merging and splitting of neuronal processes. With the recent developments in the field of electron microscopy the section thickness is greatly reduced, leading to an increase in the similarity of adjacent sections. As a consequence, the image alignment becomes easier than for the 50 nm z-resolution used in our experiments. With increased z-resolution the synapse detection can be extended to a full 3d approach.

Appendix A

Derivation of Closed Form Solutions

A.1 Distortion Correction Transform

This supplement material describes the closed form optimization of the non-linear transformation matrix α in the distortion correction and stitching formula as described in Section 3.3.

Definitions:

- $x_n^{(i,j)}$ is an 1×3 vector containing the coordinates of the n th correspondence point in image i to image j in homogeneous coordinates.
- $\tilde{x}_n^{(i,j)} = \phi_d(x_n^{(i,j)})$ is the point $x_n^{(i,j)}$ expanded with a polynomial kernel.
 - d is the degree of the polynomial kernel expansion.
 - $\phi_d(u, v, 1) = (1, u, v, u^2, uv, v^2, \dots, v^d)$.
- α is a $\frac{(d+1)(d+2)}{2} \times 3$ transformation matrix that projects the kernel expanded points back to the 2D image plane in homogeneous coordinates.

APPENDIX A. DERIVATION OF CLOSED FORM SOLUTIONS

- $A^{(i)}$ is a 3×3 affine transformation matrix for image i .
- B is the number of images.
- $N(i, j)$ is the number of correspondence points between image i and image j .

Assumptions:

- For all images i , $A^{(i)}$ is given (estimated by weighted least squares using Huber loss, as described in Section 3.2.1).
- λ is fix (in all our experiments set to 0.01).

Optimization: Correspondence points are automatically detected using SIFT features. For each correspondence point the non-linear transformation is applied to its local image coordinates in image i and image j . After correcting for distortions, the correspondences are mapped by an affine transformation to a common coordinate system corresponding to the mosaic image. For the set of all images B , the squared Euclidean distance between correspondence points should be minimal in the coordinate system of the mosaic image. This goal induces the following optimization problem:

$$\min_{\alpha, A} \sum_{i, j=1, \substack{B \\ j \neq i}}^B \sum_{n=1}^{N(i, j)} \left(\left\| \left(\tilde{x}_n^{(i, j)} \alpha A^{(i)} \right) - \left(\tilde{x}_n^{(j, i)} \alpha A^{(j)} \right) \right\|^2 + \lambda \left\| \tilde{x}_n^{(i, j)} \alpha - x_n^{(i, j)} \right\|^2 \right). \quad (\text{A.1})$$

The quadratic norm can be written as matrix multiplication

$$\begin{aligned}
\min_{\alpha, A} \sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} & \left[\left(\tilde{x}_n^{(i,j)} \alpha A^{(i)} - \tilde{x}_n^{(j,i)} \alpha A^{(j)} \right) \right. \\
& \cdot \left(A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} - A^{(j)T} \alpha^T \tilde{x}_n^{(j,i)T} \right) \\
& \left. + \lambda \left(\tilde{x}_n^{(i,j)} \alpha - x_n^{(i,j)} \right) \left(\alpha^T \tilde{x}_n^{(i,j)T} - x_n^{(i,j)T} \right) \right]
\end{aligned} \tag{A.2}$$

and then solved according to distributive law

$$\begin{aligned}
\min_{\alpha, A} \sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} & \left[\tilde{x}_n^{(i,j)} \alpha A^{(i)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} - \tilde{x}_n^{(i,j)} \alpha A^{(i)} A^{(j)T} \alpha^T \tilde{x}_n^{(j,i)T} \right. \\
& - \tilde{x}_n^{(j,i)} \alpha A^{(j)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} + \tilde{x}_n^{(j,i)} \alpha A^{(j)} A^{(j)T} \alpha^T \tilde{x}_n^{(j,i)T} \\
& + \lambda \left(\tilde{x}_n^{(i,j)} \alpha \alpha^T \tilde{x}_n^{(i,j)T} - \tilde{x}_n^{(i,j)} \alpha x_n^{(i,j)T} - x_n^{(i,j)} \alpha^T \tilde{x}_n^{(i,j)T} \right. \\
& \left. \left. + x_n^{(i,j)} x_n^{(i,j)T} \right) \right].
\end{aligned} \tag{A.3}$$

The set of correspondence points is symmetric, i.e. $N(i, j) = N(j, i)$. Thus, the formula above simplifies to:

$$\begin{aligned}
\min_{\alpha, A} \sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} & \left[2 \cdot \tilde{x}_n^{(i,j)} \alpha A^{(i)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} - 2 \cdot \tilde{x}_n^{(i,j)} \alpha A^{(i)} A^{(j)T} \alpha^T \tilde{x}_n^{(j,i)T} \right. \\
& \left. + \lambda \left(\tilde{x}_n^{(i,j)} \alpha \alpha^T \tilde{x}_n^{(i,j)T} - 2 \cdot \tilde{x}_n^{(i,j)} \alpha x_n^{(i,j)T} + x_n^{(i,j)} x_n^{(i,j)T} \right) \right].
\end{aligned} \tag{A.4}$$

To obtain the closed form solution for the optimal α we now differentiate the formula above with respect to α^T and set the result to zero.

$$\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left[4 \cdot A^{(i)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} - 4 \cdot A^{(j)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(j,i)} \right. \\
 \left. + \lambda (2 \cdot \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} - 2 \cdot \tilde{x}_n^{(i,j)T} x_n^{(i,j)}) \right] = 0 \tag{A.5}$$

In the next step, the formula is divided by two and all α terms are sorted to the left side of the equation.

$$\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left[2 \cdot A^{(i)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} - 2 \cdot A^{(j)} A^{(i)T} \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(j,i)} \right. \\
 \left. + \lambda \alpha^T \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right] = \sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left[\lambda \tilde{x}_n^{(i,j)T} x_n^{(i,j)} \right] \tag{A.6}$$

This equation now has the structure of an encapsulating sum Petersen and Pedersen (2008):

$$\sum_n Y_n X Z_n = C \Rightarrow \text{vec}(X) = \left(\sum_n Z_n^T \otimes Y_n \right)^{-1} \text{vec}(C) \tag{A.7}$$

where $Y \otimes Z$ denotes the Kronecker product of an $m \times n$ matrix Y and an $r \times q$ matrix Z . The result is an $mr \times nq$ matrix, defined as:

$$Y \otimes Z = \begin{pmatrix} Y_{11}Z & Y_{12}Z & \dots & Y_{1n}Z \\ Y_{21}Z & Y_{22}Z & \dots & Y_{2n}Z \\ \vdots & & & \vdots \\ Y_{m1}Z & Y_{m2}Z & \dots & Y_{mn}Z \end{pmatrix} \tag{A.8}$$

After transposing both sides of the equation, the closed form to estimate α is:

$$\begin{aligned} \text{vec}(\alpha) = & \left[\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(2 \cdot A^{(i)} A^{(i)T} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right. \right. \\ & \left. \left. - 2 \cdot A^{(j)} A^{(j)T} \otimes \tilde{x}_n^{(j,i)T} \tilde{x}_n^{(i,j)} + \lambda (I_{3 \times 3} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)}) \right) \right]^{-1} \cdot \\ & \text{vec} \left(\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(+\lambda \cdot x_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right) \right) \end{aligned} \quad (\text{A.9})$$

For the implementation of the distortion correction, it is possible to avoid the homogeneous coordinate system and separate the affine transformation matrix A into a 2×2 matrix R and a 1×2 matrix T :

$$A = \begin{pmatrix} R_{11} & R_{12} & 0 \\ R_{21} & R_{22} & 0 \\ T_{11} & T_{12} & 1 \end{pmatrix} \quad (\text{A.10})$$

Following this notation, the minimization problem is:

$$\begin{aligned} \min_{\alpha, R, T} \sum_{\substack{i,j=1, \\ j \neq i}}^B \left(\left\| \left(\phi_d(X^{(i,j)}) \alpha R^{(i)} + T^{(i)} \right) - \right. \right. \\ \left. \left. \left(\phi_d(X^{(j,i)}) \alpha R^{(j)} + T^{(j)} \right) \right\|^2 + \lambda \left\| \phi_d(X^{(i,j)}) \alpha - X^{(i,j)} \right\|^2 \right). \end{aligned} \quad (\text{A.11})$$

Following the same conversion scheme as described above, the closed form solution for α is:

$$\begin{aligned}
 \text{vec}(\alpha) = & \left[\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(2 \cdot R^{(i)} R^{(i)T} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right. \right. \\
 & \left. \left. - 2 \cdot R^{(j)} R^{(i)T} \otimes \tilde{x}_n^{(j,i)T} \tilde{x}_n^{(i,j)} + \lambda (I_{2 \times 2} \otimes \tilde{x}_n^{(i,j)T} \tilde{x}_n^{(i,j)}) \right) \right]^{-1} \cdot \\
 & \text{vec} \left(\sum_{\substack{i,j=1, \\ j \neq i}}^B \sum_{n=1}^{N(i,j)} \left(- 2 \cdot R^{(i)} T^{(i)T} \tilde{x}_n^{(i,j)} + 2 \cdot R^{(i)} T^{(j)T} \tilde{x}_n^{(i,j)} \right. \right. \\
 & \left. \left. + \lambda x_n^{(i,j)T} \tilde{x}_n^{(i,j)} \right) \right)
 \end{aligned} \tag{A.12}$$

A.2 Non-linear Warping

This supplement material describes the closed form optimization of the non-linear transformation matrix α for the warping of adjacent serial section images, as described in Section 3.6.

Definitions:

- n_1 is the number of reference points in image X
- n_2 is the number of reference points in image Y
- $\varphi_{\mu,\sigma}$ describes a normal distribution with mean μ and standard deviation σ
- $v(x_i)$ gives the gray value of the correspondent image at point x_i
- x_i is a 1×2 vector containing the coordinates of the i th correspondence point in the reference image X

- y_j is a 1×2 vector containing the coordinates of the j th correspondence point in the warping image Y
- $\phi_d(x_i)$ is the polynomial kernel expansion of point x_i
 - d is the degree of the polynomial kernel expansion
 - $\phi_d(u, v) = (1, u, v, u^2, uv, v^2, \dots, v^d)$
- β is a $\frac{(d+1)(d+2)}{2} \times 2$ transformation matrix that projects the kernel expanded points back to the 2D image plane
- n_β is the number of entries for $\beta \Rightarrow n_\beta = (d+1)(d+2)$
- M is a $n_1 \times (n_2 + 1)$ matrix and its entries M_{ij} specify whether x_i is correspondent to y_j
- c is a constant and defines an outlier penalty
- I is a $\frac{(d+1)(d+2)}{2} \times \frac{(d+1)(d+2)}{2}$ identity matrix

Optimization: The goal is to find the optimal transformation matrix β which maximizes the posterior probability

$$p(\beta|X, Y) = \sum_{M \in \mathcal{M}} \frac{p(X, Y|\beta, M) \cdot p(\beta) \cdot p(M)}{p(X, Y)} \quad (\text{A.13})$$

The correspondence matrices M are uniformly distributed and thus, can be omitted under the assumption that M is a valid assignment matrix. In addition, the denominator $p(X, Y)$ is constant for any two given images and thus does not influence the optimal parameter setting. Therefore, the optimization problem simplifies to:

$$\prod_{i=1}^{n_1} \prod_{j=1}^{n_2} (\varphi_{0, \sigma_1}(v(x_i) - v(y_j)) \cdot \varphi_{0, \sigma_2}(\phi(x_i)\beta - y_j))^{M_{ij}} \cdot \prod_{i=1}^{n_1} (\varphi_{0, \sigma_1}(c\sigma_1))^{M_{i0}} \cdot \prod_{i=1}^{n_\beta} \varphi_{0, \frac{1}{\sqrt{\lambda}}}(\beta_i) \quad (\text{A.14})$$

APPENDIX A. DERIVATION OF CLOSED FORM SOLUTIONS

Maximizing the term above corresponds to minimizing the negative logarithm:

$$\begin{aligned}
 & - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} [\ln \varphi_{0,\sigma_1}(v(x_i) - v(y_j)) + \ln \varphi_{0,\sigma_2}(\phi(x_i)\beta - y_j)] - \\
 & \sum_{i=1}^{n_1} M_{i0} [\ln \varphi_{0,\sigma_1}(c\sigma_1)] - \sum_{i=1}^{n_\beta} \ln \varphi_{0,\frac{1}{\sqrt{\lambda}}}(\beta_i)
 \end{aligned} \tag{A.15}$$

For the optimal solution for β only the prior and the geometric fit term are important. The normalization constants can be omitted as they do not depend on β .

$$\min_{\beta} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \cdot \frac{(\phi(x_i)\beta - y_j)(\phi(x_i)\beta - y_j)^T}{2\sigma_2^2} + \frac{\lambda}{2} \text{Tr}(\beta\beta^T) \tag{A.16}$$

The formula above is now differentiated with respect to β^T and the result set to zero:

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \frac{2\beta^T \phi(x_i)^T \phi(x_i) + 2y_j^T \phi(x_i)}{2\sigma_2^2} + \lambda\beta^T = 0. \tag{A.17}$$

We sort all terms containing β to the left side and take the transpose:

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \frac{\phi(x_i)^T \phi(x_i)\beta}{\sigma_2^2} + \lambda\beta = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \frac{\phi(x_i)^T y_j}{\sigma_2^2}. \tag{A.18}$$

So the closed form solution for the optimal choice of β is:

$$\beta = \left[\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \phi(x_i)^T \phi(x_i) + \lambda\sigma_2^2 I \right]^{-1} \phi(x_i)^T y_j. \tag{A.19}$$

Bibliography

- Abramoff, M., Magelhaes, P., and Ram, S. (2004). Image processing with imagej. *Biophotonics International*, 11(7):36–42.
- Andres, B., Köthe, U., Helmstaedter, M., Denk, W., and Hamprecht, F. A. (2008). Segmentation of sbfsem volume data of neural tissue by hierarchical classification. *Pattern Recognition*, 5096:142–152.
- Arganda-Carreras, I., Sorzano, C. O. S., Marabini, R., Carazo, J. M., de Solorzano, C. O., and Kybic, J. (2006). Consistent and elastic registration of histological sections using vector-spline regularization. *CVAMIA*, 4241:85–95.
- Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Surf: Speeded up robust features. *Comput. Vis. Image Underst.*, 110(3):346–359.
- Black, M. J. and Anandan, P. (1996). The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.*, 63(1):75–104.
- Bock, D., Lee, W., Kerlin, A., Andermann, M., Soucy, E., Yurgenson, S., Wetzell, A., Hood, G., and Reid, R. (2010). Local anatomical connectivity of a cluster of physiologically characterized cells in mouse visual cortex. In *Society for Neuroscience*.
- Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient n-d image segmentation. *Int J Comput Vision*, 70(2):109–131.

BIBLIOGRAPHY

- Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE T Pattern Anal*, 26(9):1124–1137.
- Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32.
- Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In Pajdla, T. and Matas, J., editors, *ECCV*, volume 3024 of *LNCS*, pages 25–36, Prague, Czech Republic. Springer.
- Cardona, A. (2006). Trakem2: an imagej-based program for morphological data mining and 3d modeling. In *Proc 1st ImageJ User and Developer conference, Luxembourg*.
- Chen, C., Liaw, A., and Breiman, L. (2004). Using random forest to learn imbalanced data. technical report, Department of Statistics, University of California, Berkeley.
- Chui, H. and Rangarajan, A. (2000). A new algorithm for non-rigid point matching. In *CVPR*, volume 2, pages 44–51.
- Claus, D. and Fitzgibbon, A. W. (2005). A rational function lens distortion model for general cameras. In *CVPR*, volume 1, pages 213–219, Washington, DC, USA. IEEE Computer Society.
- Costa, N. M. M. A. D. and Martin, K. (2010). How thalamus connects to spiny stellate cells in the cat’s visual cortex. *in print J. Neurosci*.
- Denk, W. and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol*, 2(11):e329.
- Devernay, F. and Faugeras, O. (2001). Straight lines have to be straight. *Mach Vision Appl*, 13(1):14–24.
- Dollar, P., Tu, Z., and Belongie, S. (2006). Supervised learning of edges and object boundaries. In *CVPR*.

- Eccles, D. (2009). Diagram outlining the internal components of a basic tem system. http://en.wikipedia.org/wiki/File:Scheme_TEM_en.svg.
- Fiji (2010). an image processing package based on imagej (fiji is just imagej - batteries included).
- Flegler, S. L., Heckman, J. W., and Klomparens, K. L. (1995). *Scanning and Transmission Electron Microscopy: An Introduction*. Oxford University Press, USA.
- Gay-Bellile, V., Perriollat, M., Bartoli, A., and Sayd, P. (2006). Image registration by combining thin-plate splines with a 3D morphable model. In *ICIP*, pages 1069–1072.
- Geyer, C. and Daniilidis, K. (2001). Structure and motion from uncalibrated catadioptric views. In *CVPR*, volume 1, page 279, Los Alamitos, CA, USA. IEEE Computer Society.
- Glocker, B., Komodakis, N., Paragios, N., Tziritas, G., and Navab, N. (2007). Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging*, volume 4584 of *Lecture Notes in Computer Science*, pages 408–420. Springer.
- Gremban, K., Thorpe, C., and Kanade, T. (1988). Geometric camera calibration using systems of linear equations. In *ICRA*, volume 1, pages 562 – 567.
- Hartley, R. and Kang, S. B. (2007). Parameter-free radial distortion correction with center of distortion estimation. *IEEE T Pattern Anal*, 29:1309 – 1321.
- Hastie, T. and Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer.
- Hayworth, K., Kasthuri, N., Schalek, R., and Lichtman, J. (2006). Automating the collection of ultrathin serial sections for large volume tem reconstructions. *Microsc Microanal*, 12(S02):86–87.

BIBLIOGRAPHY

- Hill, D. L. G. and Batchelor, P. G. (2001). Medical image registration. *Physics in Medicine and Biology*, 46(3).
- Jain, V., Murray, J., Roth, F., Turaga, S., Zhigulin, V., Briggman, K., Helmstaedter, M., Denk, W., and Seung, H. (2007). Supervised learning of image restoration with convolutional networks. In *ICCV*, pages 1–8.
- Jeong, W.-K., Beyer, J., Hadwiger, M., Vazquez, A., Pfister, H., and Whitaker, R. T. (2009). Scalable and interactive segmentation and visualization of neural processes in em datasets. *IEEE T Vis Comput Gr*, 15:1505–1514.
- Johnson, H. J. and Christensen, G. E. (2002). Consistent landmark and intensity based image registration. *IEEE Trans Med Imaging*, 21(5):450–461.
- Jurrus, E., Paiva, A. R., Watanabe, S., Anderson, J. R., Jones, B. W., Whitaker, R. T., Jorgensen, E. M., Marc, R. E., and Tasdizen, T. (2010). Detection of neuron membranes in electron microscopy images using a serial neural network architecture. *Med. Image Anal.*, 14(6):770–783.
- Jurrus, E., Tasdizen, T., Koshevoy, P., Fletcher, P. T., Hardy, M., Chien, C., Denk, W., and Whitaker, R. (2009). Axon tracking in serial block-face scanning electron microscopy. *Med. Image Anal.*, 13(1):180–188.
- Jurrus, E., Whitaker, R., Jones, B. W., Marc, R., and Tasdizen, T. (2008). An optimal-path approach for neural circuit reconstruction. In *ISBI*, pages 1609–1612.
- Kannan, U., Paiva, A., Jurrus, E., and Tasdizen, T. (2009). Automatic markup of neural cell membranes using boosted decision stumps. In *ISBI*.

- Kaynig, V., Fischer, B., and Buhmann, J. M. (2008). Probabilistic image registration and anomaly detection by nonlinear warping. In *CVPR*, pages 1–8.
- Kaynig, V., Fischer, B., Müller, E., and Buhmann, J. M. (2010a). Fully automatic stitching and distortion correction of transmission electron microscope images. *J Struct Biol*, 171(2):163–173.
- Kaynig, V., Fischer, B., Wepf, R., and Buhmann, J. M. (2007). Fully automatic registration of electron microscopy images with high and low resolution. In *Microsc Microanal*.
- Kaynig, V., Fuchs, T., and Buhmann, J. M. (2010b). Geometrical consistent 3d tracing of neuronal processes in sstem data. In *MIC-CAI*.
- Kaynig, V., Fuchs, T., and Buhmann, J. M. (2010c). Neuron geometry extraction by perceptual grouping in sstem images. In *CVPR*.
- Knott, G., Marchman, H., Wall, D., and Lich, B. (2008). Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J. Neurosci*.
- Kolmogorov, V. and Boykov, Y. (2005). What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *ICCV*, volume 1, pages 564–571.
- Kolmogorov, V. and Zabini, R. (2004). What energy functions can be minimized via graph cuts? *IEEE T Pattern Anal*, 26(2):147–159.
- König, P., Kayser, C., Bonin, V., and Würtz, R. P. (2001). Efficient evaluation of serial sections by iterative Gabor matching. *J. Neurosci. Methods*.
- Koshevoy, P., Tasdizen, T., Whitaker, R., Jones, B., and Marc, R. (2006). Assembly of large three-dimensional volumes from serial-section transmission electron microscopy. In Metaxas, D.,

BIBLIOGRAPHY

- Whitaker, R., Rittscher, J., and Sebastian, T., editors, *MIAAB, Copenhagen*), pages 10–17.
- Kreshuk, A., Straehle, C., Sommer, C., Koethe, U., Knott, G., and Hamprecht, F. (2011). Automated segmentation of synapses in 3d em data. In *ISBI, in press*, pages 1–4.
- Kumar, R., Reina, A. V., and Pfister, H. (2010). Radon-like features and their application to connectomics. In *MMBIA*.
- Lawrence, A., Bouwer, J. C., Perkins, G., and Ellisman, M. H. (2006). Transform-based backprojection for volume reconstruction of large format electron microscope tilt series. *J Struct Biol*, 154(2):144 – 167.
- Li, W. and Leung, H. (2004). A maximum likelihood approach for image registration using control point and intensity. *IEEE T Pattern Anal*, 13(8):1115–1127.
- Lingua, A., Marenchino, D., and Nex, F. (2009). Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors*, 9(5):3745–3766.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int J Comput Vision*, 60(2):91–110.
- Lucchi, A., K.Smith, Achanta, R., Lepetit, V., and Fua, P. (2010). A fully automated approach to segmentation of irregularly shaped cellular structures in em images. In *MICCAI*, Beijing, China.
- Luther, P. K., Lawrence, M. C., and Crowther, R. A. (1988). A method for monitoring the collapse of plastic sections as a function of electron dose. *Ultramicroscopy*, 24(1):7–18.
- Marsh, B. J. (2007). Reconstructing mammalian membrane architecture by large area cellular tomography. In McIntosh, J. R., editor, *Cellular Electron Microscopy*, volume 79 of *Methods Cell Biol*, pages 193 – 220. Academic Press.

- Mishchenko, Y. (2009). Automation of 3d reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. *J. Neurosci. Methods*, 176:276–289.
- Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K. M., and Chklovskii, D. B. (2010). Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron*, 67(6):1009–1020.
- Ourselin, S., Roche, A., Subsol, G., Pennec, X., and Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image Vision Comput*, 19(1-2):25–31.
- Perperidis, D., Mohiaddin, R. H., and Rueckert, D. (2005). Spatio-temporal free-form registration of cardiac MR image sequences. *Med. Image Anal.*, 9(5):441–456.
- Petersen, K. and Pedersen, M. (2008). The matrix cookbook. Technical University of Denmark.
- Preparation, S. (2010). Thin sectioning for tem. [http://219.221.200.61/ywwy/zbsw\(E\)/edetail2.htm](http://219.221.200.61/ywwy/zbsw(E)/edetail2.htm). last accessed in December 2010.
- Rasband, W. (1997-2010). Imagej. <http://rsb.info.nih.gov/ij/>.
- Reina, A. V., Miller, E., and Pfister, H. (2009). Multiphase geometric couplings for the segmentation of neural processes. In *CVPR*.
- Rother, C., Kolmogorov, V., and Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23:309–314.
- Saalfeld, S., Cardona, A., Hartenstein, V., and Tomancák, P. (2010). As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets. *Bioinformatics*, 26(12):i57–i63.

BIBLIOGRAPHY

- Saalfeld, S. and Tomančák, P. (2008). Automatic landmark correspondence detection for ImageJ. In *Proc of the 2nd ImageJ User and Developer Conference*, Luxembourg.
- Sawhney, H. and Kumar, R. (1999). True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE T Pattern Anal*, 21(3):235–243.
- Shih-Schon, L. and Bajcsy, R. (2001). True single view point cone mirror omni-directional catadioptric system. In *ICCV*.
- Stein, G. P. (1997). Lens distortion calibration using point correspondences. In *CVPR*, pages 602–608.
- Stewart, C. V., Tsai, C., and Roysam, B. (2003). The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans Med Imaging*, 22:1379–1394.
- Strecha, C., Fransens, R., and Gool, L. V. (2004). Wide-baseline stereo from multiple views: a probabilistic account. In *CVPR*, volume 2, pages 552–559.
- Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C. S., and Carragher, B. (2005). Automated molecular microscopy: The new leginon system. *J Struct Biol*, 151(1):41–60.
- Szeliski, R. (2004). Image alignment and stitching: A tutorial. Technical report, Microsoft Research.
- Tola, E., Lepetit, V., and Fua, P. (2010). Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE T Pattern Anal*, 32:815–830.
- Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., and Seung, H. S. (2010). Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.*, 22(2):511–538.

- Vasilevskiy, A. and Siddiqi, K. (2002). Flux maximizing geometric flows. *IEEE T Pattern Anal*, 24(12):1565–1578.
- Veeraraghavan, A., Genkin, A. V., Vitaladevuni, S., Scheffer, L., Xu, S., Hess, H., Fetter, R., Cantoni, M., Knott, G., and Chklovskii, D. (2010). Increasing depth resolution of electron microscopy of neural circuits using sparse tomographic reconstruction. In *CVPR*, pages 1767–1774.
- Vitaladevuni, S. N. and Basri, R. (2010). Co-clustering of image segments using convex optimization applied to em neuronal reconstruction. In *CVPR*, pages 2203–2210.
- Vu, N. and Manjunath, B. (2008). Graph cut segmentation of neuronal structures from transmission electron micrographs. In *ICIP*, pages 725–728.
- Wang, A., Qiu, T., and Shao, L. (2009). A simple method of radial distortion correction with centre of distortion estimation. *J Math Imaging Vis*, 35:165–172.
- Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., and Comaniciu, D. (2008). A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-d mri. In *MICCAI*.
- White, J., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *c. elegans*. *Phil trans R Soc London*, 314:1–340.
- Winn, J. and Shotton, J. (2006). The layout consistent random field for recognizing and segmenting partially occluded objects. In *CVPR*.
- Yang, H. and Choe, Y. (2009). Cell tracking and segmentation in electron microscopy images using graph cuts. In *ISBI*.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE T Pattern Anal*, 22(11):1330–1334.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.

Curriculum Vitae

Name Verena Sabine Kaynig-Fittkau
Date of birth September 29, 1979, in Lüneburg, Germany

08/1992 - 07/1998 High school: Gymnasium Scharnebeck

07/1998 **Abitur** at Gymnasium Scharnebeck

10/1998 - 07/2000 Basic studies (\sim B.Sc.) in sign language interpreting
University of Hamburg

10/2000 - 09/2002 Basic studies (\sim B.Sc.) in computer science
University of Hamburg

10/2002 - 07/2006 Advanced studies in computer science
University of Hamburg

07/2006 **Diplom (\sim M.Sc.) in Computer Science**
University of Hamburg

08/2006 - 03/2011 **Doctoral studies at ETH Zürich**