

PROJECT DESCRIPTION

A. Introduction

Innovative high-throughput and high-resolution three-dimensional imaging instruments have made the acquisition of unprecedented amounts of sub-cellular level of biological tissue data possible [14, 45, 52]. The *Knife-Edge Scanning Microscope* (KESM, Fig. 1), developed and hosted by the Texas A&M-side PI(Choe)'s research laboratory, is one of the first instruments producing massive three-dimensional data sets [13, 40, 45, 47, 49]. The KESM can section whole small animal organs (e.g., a mouse brain) at sub-micron resolution, generating data at a rate of 180 megabytes/s. There are several mouse brain atlases available to the research community [3, 37, 55, 62], but (1) their resolution perpendicular to the sectioning plane (z -axis) is limited (50 to 60 times less than that of the KESM), thus fine anatomical structures are hard to disambiguate and trace, and (2) the complexity and amount of data made available pose serious challenges to automated geometric reconstruction, morphological analysis, and dissemination. High-resolution collections are available, however the amount of data that can be imaged is very limited (few neurons from small local regions of the brain) [9, 10, 12, 24, 38].

Data like those made available by the KESM are extremely valuable in understanding the computational function of the brain, and together with the manner in which the data are analyzed, can lead to a paradigm shift in computational thinking. The first step in utilizing such data is to obtain an accurate geometric reconstruction of the neuronal and vascular components. However, analyzing the raw data to extract structured geometrical knowledge and connectivity is nontrivial, because automated reconstruction algorithms do not have high enough accuracy and it is hard to validate the results. Most recent automated tracing algorithms can reach nearly 95% accuracy [2, 11, 23, 31, 34, 53]. However, for accurate subsequent functional analysis and computational simulation, (1) higher accuracy, and (2) some form of strong validation process are needed (see 57, 71, 73, on the need and difficulty of validation).

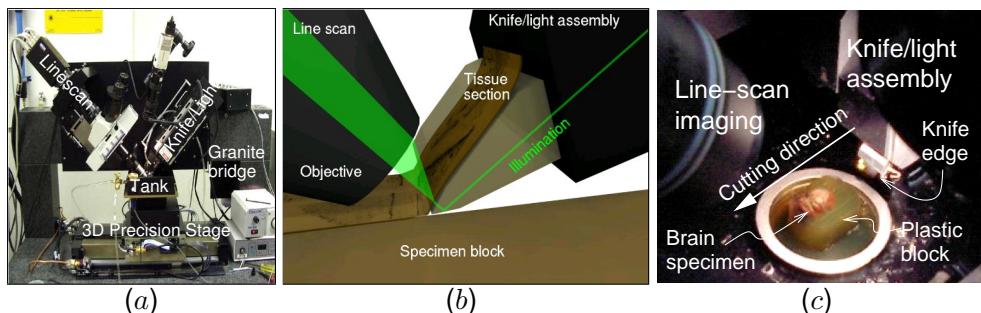


Figure 1: **The Knife-Edge Scanning Microscope.** (a) Photo of the KESM instrument showing line-scan/microscope, knife/light assembly, granite bridge, and 3D precision stage. (b) Specimen undergoing sectioning by knife-edge scanner (thickness of section is not drawn to scale). (c) Close-up photo of the line-scan/microscope assembly and the knife/illumination.

Approach: Recent years have seen an increasing interest in the potential of the human perceptual and cognitive system and utilizing it as part of an integrated computing framework: human computing, or distributed human computing [15, 60, 67, 68]. However, human computing has yet to harvest the rich capability of the human perceptuo-motor system. The human computing framework can be extended to include more dynamic interaction involving the human motor/perceptuo-motor system, not just the perceptual system, e.g., through the use of the rapidly evolving eye

tracking technology and novel analysis techniques associated with the technology. The Clemson-side PI (Duchowski) has extensive experience with this technology [16, 17, 18, 21, 70]. The main idea of this project is to integrate eye tracking and automated tracing algorithms to trace and validate structures embedded in the unique 3D mouse brain data from the KESM while humans are interacting with the data volume in realtime.

Research Goals: The research goal of this project is to develop a seamlessly integrated computational framework of eye tracking and tracing algorithms, for accurate tracing and validation of mouse brain microstructures. Specific objectives are as follows:

1. Acquire vascular and neuronal morphology data from whole mouse brains, using the KESM.
2. Develop vector-based tracing algorithms for rapid and accurate tracing of fiber-like structures.
3. Develop a framework for the use of eye-tracking technology in interactive tracing.
4. Develop a confidence-based rapid validation method using eye-tracking.
5. Conduct large-scale statistical analysis of morphological variability in neuronal and vascular networks in the mouse brain.

Education Goals: The project team will train graduate (through this grant) and undergraduate students (through the REU mechanism) in an interdisciplinary curriculum, and organize annual high school contests, by modifying the research platform into an interactive tracing game.

Innovations in Computational Thinking: The most fundamental computation is the computation by the human brain. This project will take a two-pronged approach for the further understanding of computation in the brain.

1. *From data to knowledge*: Designing human-machine hybrid systems, and solving interface and integration issues that arise will allow us to gain deeper insights into the nature of computation in humans and in machines.
2. *Understanding complexity in natural systems*: Mining the system-level blueprint of the complex neuronal circuitry and vascular networks that supply energy to the network will make possible a quantitative analysis of the architectural underpinning of neural computation, and utilize that in turning data into knowledge.

These two approaches are expected to lead to novel innovations in computational thinking.

Intellectual Merit: The new tracing algorithms, coupled with the use of the human perceptuo-motor system (eye tracking) in a unique human-machine system will help advance our understanding of computations in the brain, and lead to robust engineered systems.

Broader Impact: The neuronal and vascular data and the extracted structural information will serve as an invaluable resource for neuroscience and computer science research (data and tools will be publicly released). Graduate, undergraduate, and high school students will be involved in this unique interdisciplinary project.

B. Background and Prior Work

B.1. Physical Sectioning Microscopy for Volume Imaging

Critical to our proposed work are massive biological data sets sampled at sub-cellular resolution to which our computation will be applied. The technologies for producing such massive data are actively being developed. Such methods can lead to the complete description of the *connectome*, the study of neural connectivity at the whole-brain scale, identified as a *grand challenge* in neuroscience [65]. Examples of such methods include our Knife-Edge Scanning Microscopy (KESM) [40, 45, 46, 48, 49] (see §B.2), All-Optical Histology [66], Array Tomography [52], Serial-Block-Face Scanning Electron Microscopy (SBF-SEM) [14], and Automatic Tape-Collecting Lathe Ultramicro-

tome (ATLUM) [27, 28]. Unlike Array Tomography and SBF-SEM, KESM can survey large volumes of biological tissue (whole small animal organs), and KESM is an order of magnitude faster than All-Optical Histology or ATLUM. SBF-SEM has the advantage of ultra high resolution, and Array Tomography is ideal for repeated imaging of a single specimen with multiple immuno stains.

B.2. Imaging with the Knife-Edge Scanning Microscope

A prototype Knife-Edge Scanning Microscope (KESM, US patent #6,744,572) [45, 46, 49] has been designed at Texas A&M University with support from the National Science Foundation (MRI award #0079874; McCormick, PI), the Texas Higher Education Coordinating Board (ATP award #000512-0146-2001; Keyser, PI), and the National Institute of Neurological Disorders and Stroke (Award #1R01-NS54252; Choe, PI). The instrument, shown in Fig. 1a, is capable of scanning a complete mouse brain ($\sim 310 \text{ mm}^3$) at 300 nm sampling resolution within 100 hours when scanning in full production mode. The instrument comprises four major subsystems: (1) precision positioning stage, (2) microscope/knife assembly, (3) image capture system, and (4) cluster computer. The specimen, a whole mouse brain, is embedded in a plastic block and mounted atop a three-axis precision positioning stage. A custom diamond knife, rigidly mounted to a massive granite bridge overhanging the three-axis stage, cuts consecutive thin sections from the block. Unlike block face scanning, the KESM concurrently cuts and images (under water) the tissue ribbon as it advances over the leading edge of the diamond knife. A white light source illuminates the rear of the diamond knife, and in turn illuminates the brain tissue at its leading edge, with a strip of intense illumination reflected from the beveled knife-edge as illustrated in Fig. 1b. The microscope objective, aligned perpendicular to the top facet of the knife, images the transmitted light. A high-sensitivity line-scan camera repeatedly samples the newly cut section, imaging a stripe 1 to 1.5 mm wide across the tissue ribbon and just beyond the knife-edge, prior to subsequent deformation of the tissue ribbon after imaging. Finally, the digital video signal is passed through image acquisition boards and stored in a dedicated cluster computing system.

B.3. Imaging Results of Neurons and Neurovasculature Using KESM

The prototype KESM has been validated on Golgi-, Nissl-, and India-ink-stained mouse brain specimens, and is currently producing high-quality 3D data. *Nissl staining* dyes the RNA in the cytoplasm of all neurons and the DNA in cell bodies in all cells. However, the dendritic arbors and axons remain unstained. Thus, Nissl staining allows us to reconstruct the distribution of all cell bodies in the mouse brain. Of particular importance is their distribution within the six layers of the cerebral cortex. *Golgi staining*, in contrast, reveals the entire structure of neurons and, as it stains just 1% of the neurons in the tissue, individual neurons can be seen clearly, permitting reconstruction. Finally, *India ink* stains all blood vessels in the tissue sample. Major arteries and veins down to the smallest capillaries are stained, revealing a complex network. Fig. 2 shows volume rendering of the data. The resulting data were presented in [13, 39, 40, 41, 42].

B.4. Rapid Tracing of Fibrous Matter

In order to turn the raw data into a geometric description of the objects of interest (i.e., reconstruction), we are currently developing rapid tracing algorithms for fibrous matter such as neuronal processes and neurovasculature [13, 39, 42, 44]. Fig. 3 shows an overview of our approach, which is broadly classified as a vector tracing method [2, 8]. The vector tracing method is superior to traditional methods based on segmentation or region growing, since the local coordinate system adjusts to the inherent direction of the traced structures so that many special cases are eliminated. Our method has been used successfully on KESM and on Array Tomography data. See Fig. 3d for preliminary results.

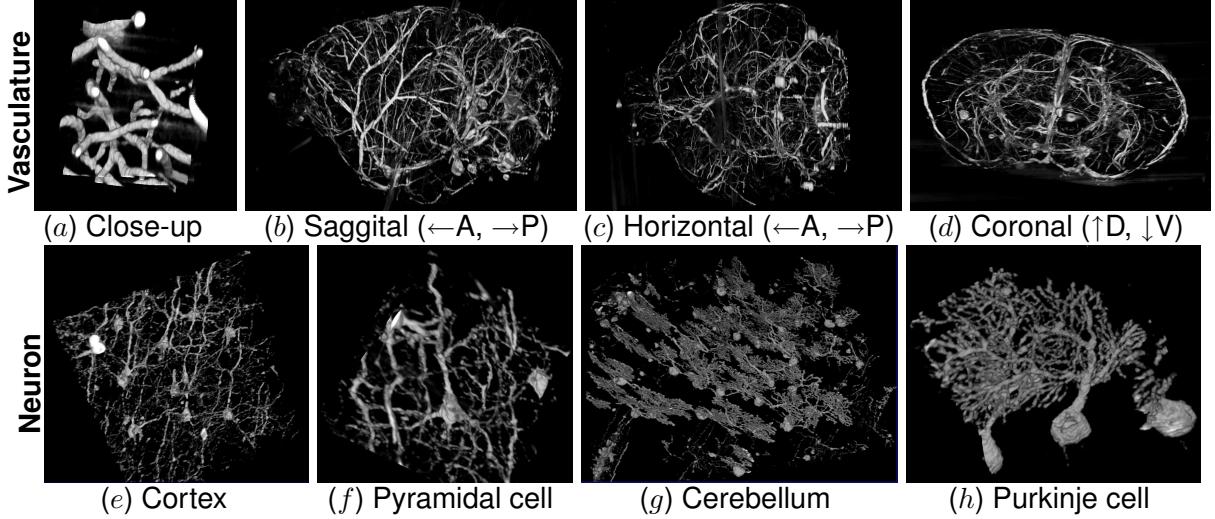


Figure 2: KESM Data. Volume visualizations of KESM data stacks are shown for the vascular data set (top row, India ink stain) and the neuronal data set (bottom row, Golgi stain). (a) Close-up of the vascular data. Width $\sim 100 \mu\text{m}$. (b–d) Three standard views of the whole mouse brain vasculature (subsampled from high-resolution data). Width $\sim 10\text{mm}$. (e–f) Pyramidal cells from the visual cortex. Width $\sim 300 \mu\text{m}$ (in e) and $100 \mu\text{m}$ (in f). (g–h) Purkinje cells from the cerebellum. Voxel size is $0.6 \times 0.7 \times 1.0 \mu\text{m}^3$. Width $\sim 500 \mu\text{m}$ (in g) and $100 \mu\text{m}$ (in h).

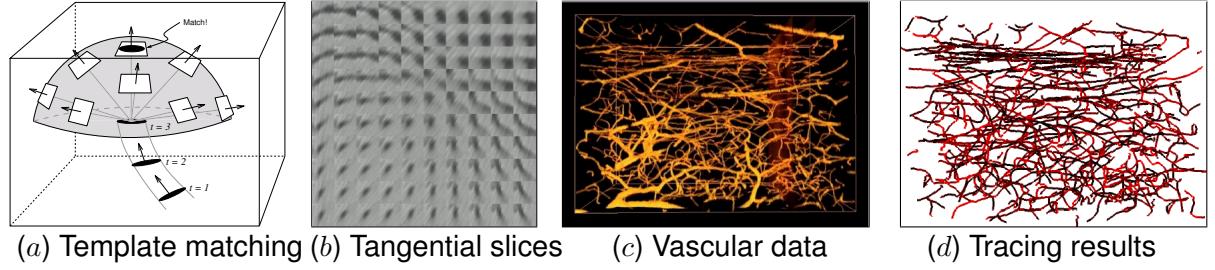


Figure 3: Fast Fiber Tracing. The basic idea behind our fast fiber tracing algorithm is shown. (a) Starting from a seed point ($t = 1$), the direction of the fiber is estimated ($t = 2$ to $t = 3$). The next point in the fiber trajectory is estimated using a template-based approach. Images are formed as interpolated slices (computed via graphics hardware) through the data volume—sampling in several directions around the current point. (Note that the distance between trace points is exaggerated.) (b) A typical array of slices from the tangential planes are shown. Only a small number of these resemble a circular cross-section of a fiber (upper-right corner). The best matching one is selected as the next point in the fiber trajectory. (c) Volume rendering of cortical vasculatures are shown (width = $350 \mu\text{m}$). (d) Tracing results on (c) are shown. (Pilot results presented in [42].)

B.5. Analysis of Eye Tracking Data

Eye tracking has a long history of theoretical and practical uses, such as in visual attention research, and applications to human-computer interaction and collaborative systems [18]. An interesting aspect of eye movement is the *scanpath*, i.e., the path followed in sequence of fixations. Scanpaths have been used for compelling visualizations of captured eye movements as early as the 1970's [56], but its full quantitative potential has yet to be exploited [19]. The Clemson-side PI (Duchowski) is a leading expert in the field of eye tracking, where his latest focus is on scanpath analysis. Coincidentally, the state-of-the-art of eye movement analysis is presently focused on pairwise comparisons of scanpaths [32, 33, 72]. The Clemson team has extended the approach by Privitera and Stark [59] that utilizes string editing, by focusing on the character-labeling por-

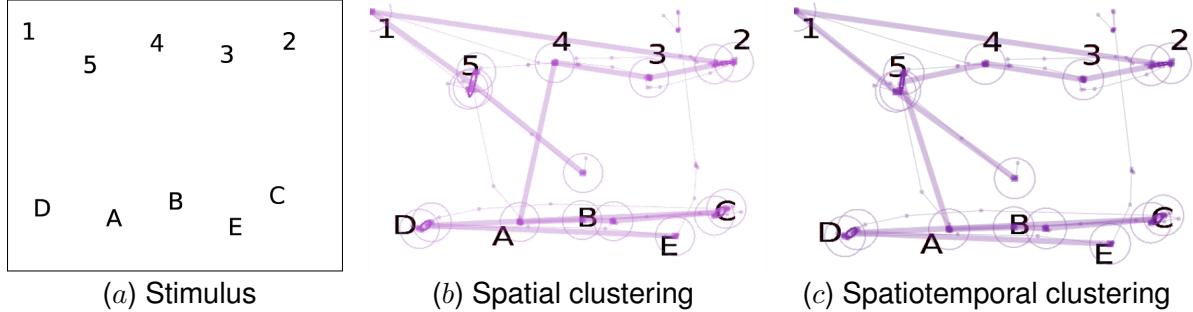


Figure 4: Eye tracking stimulus and scanpath analysis. (a) Eye tracking stimulus (sequence to follow: 12345ABCDE), (b) scanpath analysis using spatial clustering, and (c) scanpath analysis using spatiotemporal clustering are shown. Large circles represent fixation points and the thick edges connecting them show the inferred scanpath. Clustering of fixation points with spatiotemporal information shows more promise (c). For example, in (b), path from 4 to 5 is missing. Adapted from pilot results presented in [19].

tion of the algorithm [19]. Multiple viewers’ scanpaths, each defined as a sequence of clusters (modeled by ellipses), are tested for cluster overlap. Intersecting ellipses are assigned identical character labels. The algorithm’s computational efficiency is drawn from fast proximity queries provided by the *kd*-tree spatial subdivision data structure. The resultant cluster labeling leads to the computation of scanpath similarity (e.g., via string editing comparison). Through this prior work, the importance of both spatial and temporal support during mean-shift fixation clustering was established [19], and the approach successfully validated in the pairwise comparison approach when evaluating still images (Fig. 4). However, to our knowledge, no algorithm currently exists to compare scanpaths captured over dynamic media, such as movies (or, for our purpose, 2D movies from 3D volume data).

B.6. Human Computing

With the dramatic increase in the number of users connecting to the internet and with the rich interaction offered by modern web interfaces, it is natural that ideas emerged to capitalize on the raw human perceptual and cognitive power for solving tough computational problems such as computer vision and natural language processing. Human computing, or distributed human computing, is an emerging field based on this simple yet powerful idea (see [60] for a comprehensive review). For example, perceptual and cognitive skills of internet users have been in image labeling to enhance search [68], digitizing old publications [69], question answering and solving complex tasks (collectively called “crowdsourcing”) [29], and in many other problems. Some uses such as reCAPTCHA has been able to reach 99% accuracy in text transcription, rivaling the level of human experts, based on over 40,000 web sites participating, as of 2008 [69]. However, one other fundamental skill possessed by humans is their motor skill (or perceptuo-motor skill), which has yet to be tapped. Our proposed project will develop a framework for exactly this, by integrating human computing and eye-tracking technology. *Note that due to the involvement of specialized eye-tracking equipment, a large-scale distributed experiment cannot be conducted within the scope of the current project. See near the end of §C.3 for provisions to overcome this concern.*

Results from Prior NSF Support

Texas A&M: For results from the NSF MRI award (#0079874, PI: McCormick), see §B.2–§B.4 and cited references. The NSF-supported MRI project was further recognized by two subsequent

funding from the state and the federal government: the Texas Higher Education Coordinating Board (ATP#000512-0146-2001; Keyser, PI), and the National Institute of Neurological Disorders and Stroke (#1R01-NS54252; Choe, PI). This proposal is distinct from the previous projects, with an added focus on microvasculature (as well as neuronal morphology), and on a neuroinformatics framework focused on tracing and model-based validation. Keyser's NSF supported research (#0220047; Keyser, PI) focused on accurate and robust operations on curved geometry. Efforts from the project resulted in 12 publications (see the CV for details). The PI is currently supported by NSF grant #0905041 (PI: Choe, Co-PIs: Keyser, Abbott), and the Co-PI Keyser by grant #0917286 (PI: Keyser). Both projects have just begun on September 1, 2009, so there is no published result yet (2 papers submitted).

Clemson: The Clemson-side PI (Duchowski) has been funded by multiple NSF grants on eye tracking (NSF CAREER, #9984278), visual inspection (NSF ATE, PI: Carey Castle, Greenville Technical College), graphics, and included educational grants such as an REU site (#0850695, PI: Larry Hodges) and curriculum-type grants. Dr. Duchowski's NSF CAREER and ATE grants helped establish the eye tracking laboratory at Clemson, and led to a major book on eye tracking methodology [16]. The ATE grant in particular resulted in an award-nominated ACM SIGCHI publication [63], which showed the efficacy of eye tracking for feedforward training of visual search inspection. Both eye-tracking related efforts have enabled Dr. Duchowski to refine efficient methodological techniques for eye movement data collection and analysis [18]. Teaming with members from Microsoft, he taught a full-day course on these methods at CHI 2008. See the biosketch for the list of publications resulting from prior NSF support.

C. Research Plan

We propose to develop techniques for extracting and validating structural information from the KESM scans of mouse brain neuronal and vascular networks. For the automated extraction of structural information, we will improve upon our vector-tracing algorithm for fast and accurate tracing of the objects of interest. Closely integrated with the automated algorithm, we will develop methods to use eye-tracking to interactively trace and validate tracing results. Combined, we expect to achieve an efficient semi-automated method for tracing and validation. The sections below are organized in the same order as the list of main objectives in the introduction.

C.1. Acquire vascular and neuronal morphology data from whole mouse brains, using the KESM

We propose to collect additional data from the brain of the laboratory mouse (C57BL/6) beyond the pilot data we have collected (one full Golgi-stained brain and one full India-ink stained brain). This task is now a highly routine and fully automated procedure in our lab. Scanning one brain takes less than 2 weeks, so a total of less than 8 weeks (2 mice per stain, for estimating the across-specimen variability and cross-validation) will be needed to complete this task, while other tasks are carried out in parallel.

C.1.1. Stain and prepare specimens

We will examine the brain of the laboratory mouse, as it offers certain advantages for study. The brain is small; its cortex is not folded; tracing experiments and imaging techniques can be performed on live mouse brain; the statistical characteristics of its microscopic features have been quantified carefully and extensively [7]; and the laboratory mouse is the leading mammalian model for genomic modeling and gene modification experiments. C57BL/6J mice have been chosen as representatives of the mammalian brain as they are a very common inbred mouse strain used

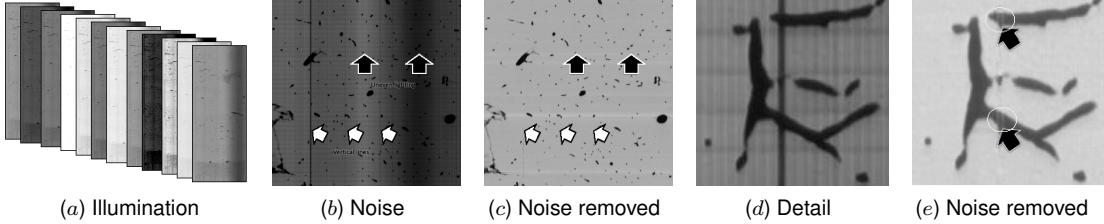


Figure 5: KESM Image Noise and Noise Removal. Various forms of KESM image noise and results of noise removal are shown. (a) Uneven illumination across images due to the cutting speed variation to avoid mechanical chatter. (b) Local illumination bands (black arrows) and knife-defect bands (white arrows). (c) Noise-removed version of b. (d) Close-up of knife-defect band. (e) Noise-removed version of d using selective normalization.

for transgenic studies and aging research. Mice for this study will be obtained from commercial vendors. Brains from two mice will be used for each stain. Note that part of the brain specimen will be used for pilot runs of the KESM, thus those particular data may not be stored for analysis. Data from two mice for each stain will be collected and analyzed for the production runs. We will use two stains, Golgi and India ink. See Table 1 and [1, 30, 61] for the exact protocols. After staining, the whole brains are dehydrated through a series of graded ethyl alcohols (25%–100%) and then infiltrated with araldite and embedded in blocks [1].

Table 1: Specimen Preparation Protocol.

Golgi: The whole brain is saturated in a Golgi-Cox solution [61], and the cells subsequently dyed by immersion in an ammonium hydroxide solution. One advantage of Golgi staining is that it reveals the *entire* neuron structure, but only stains a few neurons (1%) in the tissue. Stained cells are randomly selected, so the process is not reproducible across specimens, and thin axons are not stained well. Thus, we will focus on dendritic morphology and variability.

India ink: The mice are perfused transcardially with saline, then 10% formalin and finally with India ink, which fills the cardiovascular system. The ink stays in the blood vessels during processing of the tissue for embedding. We will perform whole body perfusions with fixatives and India ink such that the entire brain is perfused sufficiently.

C.1.2. Section and image with KESM

The prepared specimens are then scanned by the KESM using a fully automated process. The output of this scanning will be a large volumetric grid of intensity values. Each pass of the knife captures a range of x values (e.g. 2048 or 4096 pixels), across the entire y range (9,000-30,000 pixels); typical scans are $4096 \times 12,000$. Adjacent slices in x are taken in a stair-step manner in order to create a full cross-section of the tissue. Thousands of consecutive slices are also taken in the z direction. The total data produced from a scan of a single mouse brain is currently over 2 TB (using 10X, NA 0.3, Nikon objective).

C.1.3. Imaging artifact removal and image registration

Knife-edge scanning microscopy has its own sources of noise and artifacts. The primary sources of error include: misalignment of the knife edge (causing a uniform gradient to appear in the image), defects in the knife blade (causing “streaks” in the y -axis of the image), oscillations in the light source due to AC power fluctuation (causing a regular pattern of light changes along the y -axis), and knife chatter (causing horizontal lines and loss of most data at those pixels) in the image (see Fig. 5). Image registration issues are minimal due to the way KESM is designed and built.

We have already developed methods for automatically cleaning this data [36, 43] (Fig. 5), retrieving the data-containing region, and aligning the stacks [35]. The full algorithm is as follows:

(1) Normalize pixel intensity $n(x, y)$ by row and then by column: $n(x, y) = \frac{p(x, y)}{\text{median}(\text{row}(y))} \times B$, where $p(x, y)$ is the pixel at (x, y) , $\text{median}(\cdot)$ is the median value, $\text{row}(\cdot)$ is the set of pixels in the row, and B is the desired background intensity. Column normalization is done in a similar manner. Use of B allows background illumination to become the same over all images in the image stack. (2) Apply step 1 using selective normalization, where pixels are restored to their original value when their normalized values occupy the darker-end peak region (i.e., foreground) in the intensity histogram (Fig. 5d&e). (3) Remove local artifacts due to cutting irregularities using adaptive local scaling: $p(x, y) = \frac{p(x, y)}{\text{mean}(L(x, y))}$, where $L(x, y)$ is a local window around (x, y) . (4) Crop non-data portion of the image using a boundary detection template to calculate the cut-off column $x_{\text{cutoff}} = \arg\min_x (\sum_{c=0}^w \sum_{r=0}^h |p(x+c, y+r) - p(x, y)|) + \frac{w}{2} - W$, where $p(x, y)$ is the pixel, (w, h) is the template width and height, and W is the width of the uncropped image. We have found that these algorithms are sufficient in removing most KESM-specific artifacts from the images. Finally, (5) we use stack-by-stack registration using our pattern-matching-based registration algorithm described in [35]. Note that within each stack, images are perfectly aligned due to the way KESM is designed.

C.2. Develop vector-based tracing algorithms for rapid and accurate tracing of fiber-like structures

Our goal is to extract from the KESM volume data set the individual geometric structures of interest, i.e., we want to reconstruct the microstructures based on the raw volume data. We have developed a series of vector-based tracing algorithms [26, 39, 41, 50]. Here, we will discuss [26] in detail, while keeping it open to adopt to this project our other approaches as well.

Our maximum intensity projection(MIP)-based algorithm provides a general framework for 3D tracing, and one unique strength is that it can use any 2D tracing algorithm as a subroutine. Thus, the overall power of our MIP-based algorithm can be arbitrarily increased by the use of more powerful 2D tracing algorithms. Furthermore, our algorithm takes only about 30% of the time to trace a unit block compared to a full 3D algorithm. The algorithm consists of four steps, starting from a seed point: (1) boundary detection and local axis length determination, (2) local volume estimation and local MIP processing, (3) 2D tracing of fiber direction, and (4) 3D fiber direction estimation and adjustment. Fig. 6c shows an overview of our approach, and for details see [26].

(1) First, we define $P_{\text{in}}(\vec{c}, R)$, $P_{\text{out}}(\vec{c}, R)$, and $P_{\text{edge}}(\vec{c}, \tau)$ as the probability of a voxel R -voxel (or τ -voxel) apart from trace center \vec{c} being inside, outside, or on the edge of the fiber, respectively (Gaussian functions are used.) From this, we obtain the edge map $E(\vec{c}, R, \tau) = \frac{1}{3}(P_{\text{in}}(\vec{c}, R - \tau) + P_{\text{out}}(\vec{c}, R + \tau) + P_{\text{edge}}(\vec{c}, \tau))$, and $E(\vec{c}, R) = \max E(\vec{c}, R, \tau)$. The foreground-background boundary along the three axes x , y , and z are found by calculating $\arg\max_{R \in [\vec{c}, \dots, \vec{c} + M\vec{e}]} E(\vec{c}, R)$ where M is the maximum fiber width and \vec{e} is the unit vector along one of the principal axes. (2) From these boundaries, the local MIP volume is created, and MIPs generated on the XY, YZ, and XZ planes. Among these, the projection along the longest fiber direction is ignored since it does not contain fiber direction information. (3) Tracing in 2D is done using a multiscale filter using a Hessian matrix $H_\sigma(x) = \begin{bmatrix} I_{xx}(x) & I_{xy}(x) \\ I_{yx}(x) & I_{yy}(x) \end{bmatrix}$, where $I_{uv}(x) = \sigma^2 \frac{\partial^2 G(x, \sigma)}{\partial u \partial v} I(x)$, where σ is the scale, and $G(x, \sigma)$ is a Gaussian with standard deviation σ , and $I(x)$ is the voxel x in the data volume I . The eigenvector \vec{v}_1 corresponding to the smallest eigenvalue λ_1 indicates the fiber direction. To elaborate, first a fiber-likeness function $V(x, \sigma)$ is defined as 0 when 2nd eigenvalue $\lambda_2 > 0$, and as $\exp\left(-\frac{\lambda_r^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right)$ otherwise, where $\lambda_r = \frac{\lambda_1}{\lambda_2}$, and $S = \sqrt{\sum_{i=1}^2 \lambda_i^2}$, and β and c control the sensitivity of the Hessian filter (0.5 and 0.25, in our experiments). The maximum value along $V(x, \sigma)$ at scale σ gives the eigenvector \vec{v}_1 , which is the fiber direction (see [25] for details). (4) Finally, from the 2D tracing on the two MIP planes, the 3D fiber direction is estimated as follows.

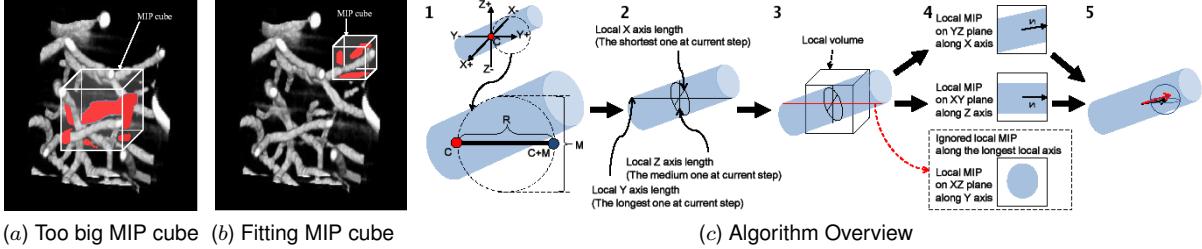


Figure 6: Vector Tracing with Local Maximum-Intensity Projection (MIP). (a-b) Local MIP size has to be small enough for unambiguous, occlusion-free projection (a contains occlusions while b does not). (c) The local MIP-based tracing algorithm is outlined. 1. Starting from seedpoint C, longest-axis boundary C+M is found. 2. Boundaries along the other two axes are found. 3. Local MIP cube size is determined. 4. 2D trace is conducted on two of the three MIPs. 5. Results are combined in 3D.

Assume that XY and YZ were the two chosen planes. Let $\vec{v}_{xy} = (x_1, y_1, 0)$ and $\vec{v}_{yz} = (0, y_2, z_2)$ be the eigenvectors of the two respective MIP planes. The 3D fiber direction is derived as $\vec{v}_{xyz} = (x_1, \frac{y_1+y_2}{2}, z_2)$. From this, the next trace center \vec{c}' can be calculated as $\vec{c}' = \frac{\vec{v}_{xyz}}{\|\vec{v}_{xyz}\|} \rho + \vec{c}$, where ρ is the step size (currently set to 3). The last step corrections for discretization error, where we use a momentum operator to calculate the adjusted center $(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{M_{100}}{M_{000}}, \frac{M_{010}}{M_{000}}, \frac{M_{001}}{M_{000}} \right)$, where $M_{pqr} = \int_z \int_y \int_x x^p y^q z^r I(x, y, z) dx dy dz$ (see [5] for details).

Our algorithm, due to the use of a 2D Hessian matrix, has computational complexity of $O(n^2)$, while full 3D tracing algorithms based on 3D Hessian matrices have $O(n^3)$ complexity, where n is the scale of the Gaussian filter. In practice, our algorithm takes only about 30% of the time of full 3D versions. Initial implementation of this and related approaches has proven successful [26, 39, 41, 50], and we will modify and improve it to trace a wider variety of fibrous data.

C.3. Develop a framework for the use of eye-tracking technology in interactive tracing

Our proposed gaze-directed approach to guiding the tracing of microscopy data for three-dimensional, volumetric representation, as exemplified in Fig. 2, aims to enhance human cognition and generate new knowledge from voluminous digital data. By collecting expert eye movement data for this purpose, we intend to combine or infuse the rendering process with cognitive expertise, which we conjecture will in turn enhance cognition through more meaningful visualization.

To do so we propose to collect and analyze eye movement data over video clips of dynamic media (digital movie clips) recorded during volumetric inspection of microscopic data, e.g., vasculature, as shown in Fig. 2. We intend to record experts' eye movements (e.g., those of neuroscientists) in an effort to select informationally salient topographic regions. In this way we intend to streamline as well as select and highlight informative structures for volumetric reconstruction.

In order to compare expert scanpath with automated tracing results, we need quantitative scanpath comparison metrics. An easy to use computational approach is sought that is analogous to statistical packages that quickly and easily generate tables of means and ANOVA statistics from experimental data. Our most recent efforts revisit the string editing approach, one that computes similarity (or "distance") between pairs of scanpaths along with statistical levels of significance [59]. The resulting metric is similar to Spearman's rank-order coefficient [6] but yields a value $S \in [0, 1]$ instead of $S \in [-1, 1]$ [19]. The metric, while expressive for different combinations of viewer and stimulus, is meant for scanpaths captured over still images. For the work described in this proposal, a new metric designed for video stimulus must be developed. We expect such a metric to have far-reaching application for eye movement analysis over any type of dynamic (video) stimulus, e.g., movies, games, etc., beyond its specific application motivated here.

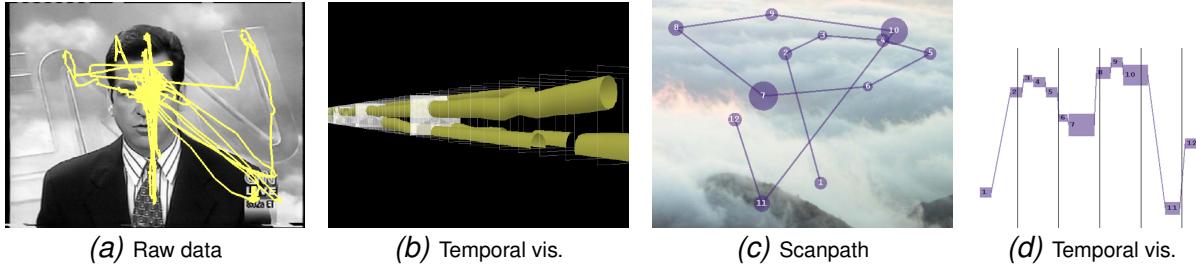


Figure 7: Eye movement and temporal visualization. Raw eye movement data collected over video in (a) with its temporal visualization in (b). Fixation-filtered scanpath over still image visualized in (c) with its temporal projection in (d) where the y -axis denotes vertical gaze position and the x -axis denotes time. Fixation labels are common between (c) and (d); vertical markers denote one-second intervals.

The proposed scanpath similarity metric and subsequent classification algorithm is inspired by an early attempt at spatiotemporal visualization of scanpaths, as shown in Fig. 7. There, we referred to scanpaths as “volumes of interest” [20]. This temporal mapping allows visualization of multiple scanpaths in two and three dimensions. The most relevant aspect of the 3D visualization to the proposed work is the preservation of fixation durations. Conventional views of scanpaths are two-dimensional, projecting the temporal dimension onto the 2D plane. This tends to obscure fixation duration—although the relative durations can be shown by varying the radius of fixations, explicit duration is lacking. This is illustrated in Fig. 7(c)–7(d) where the difference between the standard scanpath representation and a side view of the three-dimensional representation is highlighted. Note that both components of Fig. 7(c)–7(d) agree in their depiction of fixations. Fixations are filtered raw data points, usually by either the well-known velocity-based or the position-variance algorithm, each designed to detect fixations and remove saccades.

From the perspective of a collection of movie frames, each frame can be thought of as a separate stimulus. The scanpath for a single subject, viewing a movie stimulus, can then be broken up into a collection of fixation-frame units. This conceptualization of a scanpath differs from the conventional view, where fixations are projected onto a two-dimensional plane. Our conceptualization avoids this projection entirely. Instead, we produce a three-dimensional scanpath function $f(s, t)$ that returns a fixation (or null for a saccade) for a given scanpath s and frame at time t .

We extend the above definition to the function $f(S, t)$ by changing the single scanpath parameter s to a collection of scanpaths S . This function would then return a collection of fixations for all scanpaths in S at the given timestamp. Then, we may differentiate groups of subjects into their own scanpath sets. For instance, in our experiment, we study the differences between experts and automated tracing results. We may then create an expert scanpath set E and an automated scanpath set N . The functions $f(E, t)$ and $f(N, t)$ would then return collections of fixations at timestamp t for experts and automated traces, respectively. These group-specific collections of fixations for single frames may be clustered by the mean shift approach [64]. The resulting clusters serve as general area of interests (AOIs) for a given frame, describing regions of varying interest for that specific group of individuals. We may then construct a probabilistic model of expected trace for that group. Such a model for a single frame is visualized in Fig. 8(c). The figure also shows the performance of the probabilistic model compared to string-editing-based and random (for baseline) classification methods, for classifying gaze patterns into expert gaze or novice gaze (Fig. 8(d)).

The actual application of the above framework to tracing may require domain specific consid-

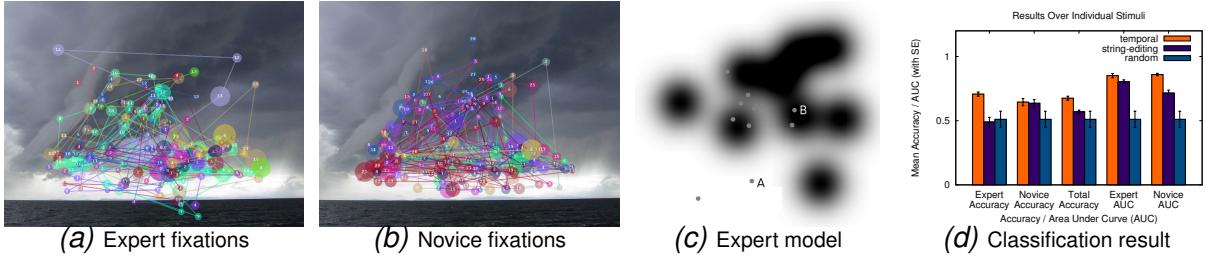


Figure 8: Comparison of scanpath categories using Gaussian mixture model. Fixation patterns of (a) expert and (b) novice are shown. Mixture of Gaussians for expert fixations at a discrete timestamp is shown in (c). Displayed novice fixations were not used in the clustering operation. Note that the fixation labeled ‘A’ is far from the cluster centers, and thus has lower similarity than fixation labeled ‘B’ that is close to a cluster center. Such deviations from the model can be used to distinguish one type of fixation pattern from another (e.g., expert vs. novice). Results of the classification of fixation pattern into expert gaze or novice gaze is shown in (d), compared to string-editing-based classification and random classification ($n = 20$). Note that this kind of comparison can be utilized in validation.

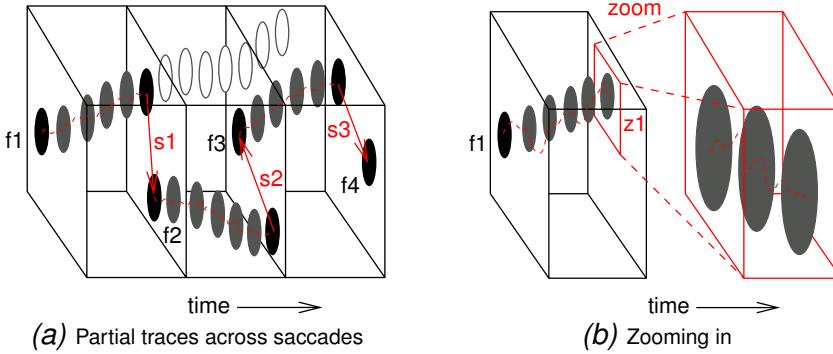


Figure 9: Eye tracking strategy for tracing fiberous matter. Illustration of how eye tracking can be used in tracing 3D objects in an image stack is shown. To avoid clutter, not all objects embedded in the 3D volume are shown. The 3D volume is traversed from left to right, where at each time step, a single 2D image is shown. Dashed meandering curves show the gaze trajectory over time. (a) Humans tend to abruptly change their gaze (saccade), and at first this may seem like a major obstacle in the use of eye tracking for tracing. However, during a tracing event without a saccade, as shown in the traces starting from f_1 until saccade s_1 occurs, partial traces can be done perfectly (the remainder of the object [shown as open ovals] can be traced at a later time). Since saccades (s_1, s_2 , etc.) can be detected easily, the identity of the fibers can be maintained. (b) When the variation in fixation location become too high compared to the diameter of the object being traced, we can zoom into the local region, thus by reducing the variation/diameter ratio. (The gaze trajectories are exaggerated to illustrate the point better.)

erations. First, the 2D slice of the 3D data volume may contain more than one salient objects that grab human attention. As shown in Fig. 9(a), when there are many fibers embedded in the 3D volume, human attention can shift (saccade) almost randomly after following one fiber for a short period of time. However, such saccade events can be easily detected based on the fixation cluster analysis described above, and multiple passes through the 3D volume will produce all fragmented traces that can be stitched together. Second, the variation in fixation position can be greater than the diameter of the traced object. In this case, as shown in Fig. 9(b), the display can be zoomed into the data volume. Other ideas to reduce human workload and exploit automated tracing results is to only show regions with low confidence, measured by automated algorithms (see Fig. 12 and

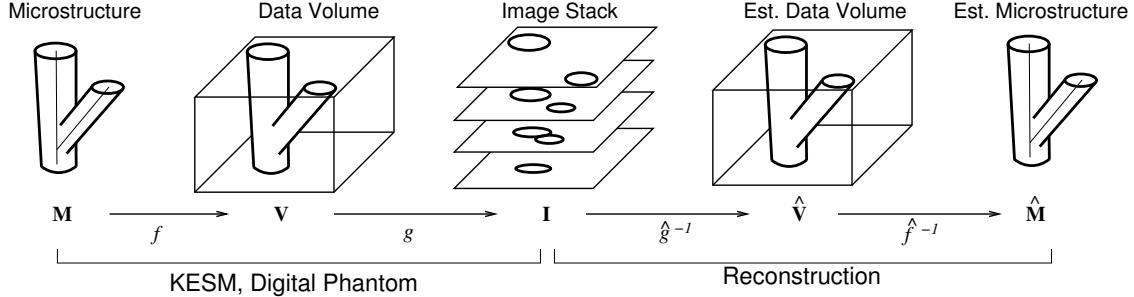


Figure 10: Microstructure-to-image Mapping and Reconstruction. The process by which a microstructure (real or synthetically generated) is turned into a stack of images in KESM and how they are reconstructed is shown. Modeling this process ($g \circ f$, composition of g and f) enables the generation of realistic synthetic image data (digital phantoms). On the other hand, the task of recovering the structural descriptions from the image data is basically the inverse: $\hat{f}^{-1} \circ \hat{g}^{-1}$, a composition of the segmentation (\hat{g}^{-1}) and the 3D reconstruction process (\hat{f}^{-1}). (The “ $\hat{\cdot}$ ” symbol indicates that these functions are estimates.)

§C.4). For the purpose of this and subsequent eye-tracking tasks, we will purchase and install identical eye-tracking systems at Texas A&M and at Clemson. The small-scale test would be to collect eye movements at both locations, aggregate, then compare via the classification scheme given above. Such comparison will also help refine the reconstruction algorithm (by comparing human gaze and algorithmic traces), thus creating a human-computing refinement loop. Finally, to fully harness the potential of the human computing paradigm, we will develop a pilot web-based interface for interactive tracing using the mouse pointing device as a surrogate of the eye-tracking system. We expect our experimentation with the eye-tracking system to help identify key interface characteristics to be included in such a pilot platform (e.g., utilizing free-form mouse movement without any clicking to mimic gaze).

C.4. Develop a confidence-based rapid validation method using eye-tracking

We will develop our validation approach in two stages. In the first, we will use small volumes of data, performing validation over these volumes. Then in the second stage, we will scale up the validation to large-scale volumes.

C.4.1. Small-scale manual and model-based validation

As a first step, we will operate on small volumes of data, for example, a $300\mu\text{m} \times 300\mu\text{m} \times 300\mu\text{m}$ cube of mouse brain. Consider Fig. 10 which illustrates the image acquisition and reconstruction process. Given the data volume (image stack I) obtained from KESM, we will have domain experts (from Co-PI Abbott’s lab) label the data to provide the ground truth \hat{M}_e (note that the domain expert’s morphological labeling is in itself an estimation of the actual ground truth M). We will also use model-based neuron and vascular tree generators to construct realistic digital phantoms for further quantitative validation (see [57] for a review).

Concurrently with manual labeling and digital phantom generation, the data volume I will be put through our segmentation and reconstruction algorithms, producing the estimated microstructure \hat{M} . Validation can now be done by comparing \hat{M}_e (by the domain expert or digital phantom) and \hat{M} (by the reconstruction algorithm). We will use mutual information for the comparison: $I(\hat{M}_e, \hat{M}) = H(\hat{M}_e) + H(\hat{M}) - H(\hat{M}_e, \hat{M})$, where $H(\cdot)$ is the Shannon entropy $-\sum_{p \in P} p \log p$, and $H(\cdot, \cdot)$ the joint entropy. We will also employ measures such as normalized mutual information, entropy correlation coefficient, and gradient [58].

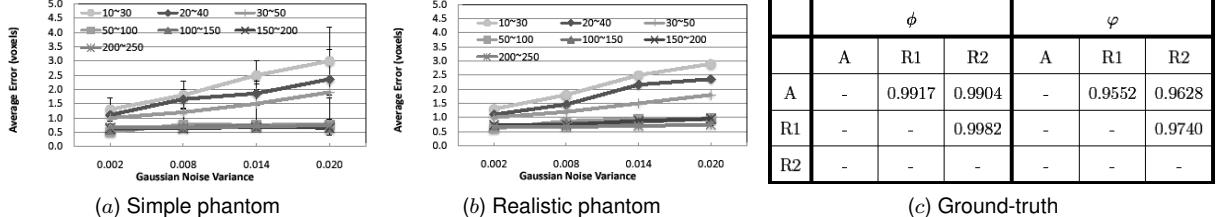


Figure 11: Pilot Validation Results. Systematic validation results with varying image contrast and noise level for digital phantoms made of (a) simple, geometric structures and (b) realistic vascular trees are shown. (c) Validation result A against two manual tracing R1 and R2 on real vascular data are shown (ϕ means length difference, and φ means centerline deviation. Values in the table are the correlation coefficients. Note that the results are very similar between a and b that use different types of phantoms, and tracing is close to perfect as can be seen in c. (Pilot results presented in [26].)

The key research issue in this task is how to map corresponding parts in the geometric microstructure model of the ground truth \hat{M}_e and that of the estimate \hat{M} in order to estimate the joint probability. We will employ two approaches, here.

First, we will use an approach for **1. Morphology comparison**. For this, we will consider the geometric microstructure models as graphs of vertices and edges, and employ approximate graph matching (often called “attributed graph matching”) algorithms (e.g., [22, 51]) to align the models being compared. Then, we will treat the morphometry (diameter, in-degree and out-degree, volume, length, etc.) associated with the pairs of vertices and edges as joint events, from which we can calculate the probability distributions. **2. Imaging comparison.** Given reconstruction \hat{M} from KESM images I , calculate $g \circ f(\hat{M})$ to regenerate an estimate of the image stack. This reconstructed volume will be directly compared with the digital phantom of manually labeled ground truth ($= g \circ f(\hat{M}_e)$). In addition, we can compare our reconstruction directly with the original KESM stack of images. For these comparisons, we intend to use one or more of the numerous existing image comparison algorithms: select a good one, rather than develop yet another one on our own.

This second approach has three advantages. First, by employing more standard image-based comparisons, the definition and computation of metrics for comparison is substantially easier. Second, by comparing reconstructions based on stacks, we eliminate the effects of structural variation that occurs at either too high of a frequency or too fine of a precision to actually be contained in the image scan. That is, the method is robust to variations between \hat{M} and \hat{M}_e that cannot be determined from the scanned data. Third, the image-based comparison allows for comparison of the reconstructed data directly to the raw image scans, allowing the iterative improvement process to progress farther than might be possible with synthetic data, and removing the limitation that all data be hand-labeled for verification.

On the other hand, this approach has disadvantages in that finer structures that are known to exist but cannot be known from the raw data will be ignored, even if expert knowledge determines they should be there. That is, there may be some artificial limitation in the amount of detail an expert may provide to the process. Also, there is a heavy reliance on the accuracy of $g(\cdot)$, which means that differences could be due to problems other than in \hat{M} . Fig. 11 shows pilot validation results of our MIP-based vector tracer against digital phantoms and manual ground truth.

C.4.2. Large-scale, confidence-based, selective validation using eye tracking

Manual validation, combined with validation based on digital phantom can be the best approach to validation, but in practice it is impossible to conduct such validation at a large scale. To overcome this limitation, we propose a confidence-based, selective validation that can be done rapidly using eye tracking. This way, regions in the data volume that are most suspected to contain error can be

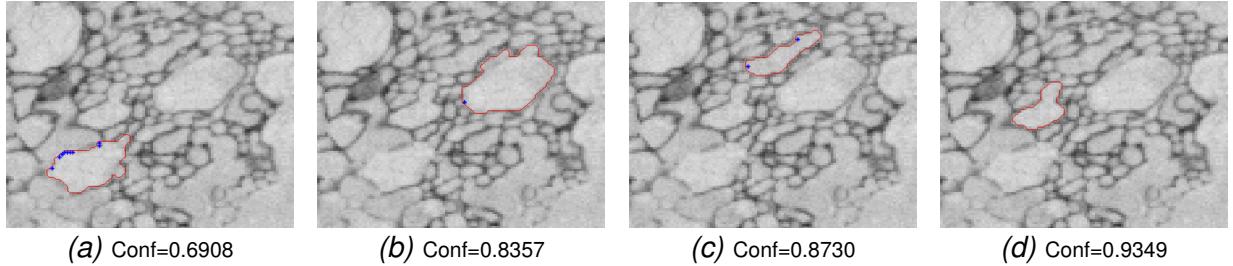


Figure 12: **Segmentation confidence measure.** Local contrast information can be used to give a confidence value for automated segmentation results. Scanning electron microscopy data are shown with the segmentation results (red enclosed contours), with the parts with low confidence marked (blue asterisks), and the overall confidence value shown below (min 0.0, max 1.0).

quickly inspected, and thus large volumes can be covered in a short period of time.

Our confidence measure will be based on local contrast. To measure the intensity differences between foreground and background, χ^2 distance metric is applied, which is given by: $\chi^2(x, y) = \frac{1}{2} \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$, where x_i is the value of the normalized histogram of x at bin i . The segmentation (or object) boundary points are those with great intensity differences between two regions. The contrasts between two regions along the boundary points are computed. Let F be the set of boundary points and R be the foreground, the intensity contrast C of a boundary point s is given by: $C = \max \{c(s, t), t \in N_s, t \notin R\}$, where $c(s, t)$ computes the intensity difference between the two points s and t , and N_s is the neighborhood of s . All computed C values along the boundary are normalized to fit [0...1]. A point is considered to have good segmentation if its C value is greater than a predefined threshold value. These threshold values can be empirically determined by data we obtain from the exhaustive small-scale validation in §C.4.1. Fig. 12 shows pilot results of this approach.

Based on the confidence measure, small volumes containing low-confidence regions will be identified and displayed on the screen. The raw data volume will be shown as a 2D movie, as usual. Repeated eye-tracking traces on the volume are expected to correct potential errors in the automated traces. We expect the amount of data to be inspected in such a way to be less than 5% of the total volume. In parallel to this, we will conduct similar validation runs on randomly sampled high-confidence regions, to ascertain the correctness of the automated tracing algorithm.

C.5. Conduct large-scale statistical analysis of morphological variability in neuronal and vascular networks in the mouse brain

Based on the tracing results on neuronal and vascular data, we will conduct a large-scale statistical analysis of morphological variability. For both neurons and vasculature, morphology contributes greatly to function. Understanding the variability both (1) within one specimen, across different regions in the brain; and (2) across specimen, can lead to insights into the functional modularization of the brain. For example, our initial analysis of a small sample from the rat brain has shown significant differences in vascular morphology across the cortex, cerebellum, and the spinal cord [42]. Such quantitative data can help us understand the metabolic demand of these different regions, and help correlate the anatomy to functional data (such as fMRI). The scope of this project will limit us to focus on the following. (1) Neurons: we will focus on the variability in dendritic morphology in cortical pyramidal cells, since Golgi does not reliably stain axons. We will collect morphological statistics such as segment length and diameter, branch order, number of bifurcations, soma diameter, and other parameters to uniquely specify any given neuron [4]. (2) Vasculature: we will continue our work reported in [42] to collect statistics similar to the ones listed above for neu-

ronal morphology. One distinct feature of vascular networks is that they are basically graphs, not trees. Thus the existence and prevalence of cycles in the vascular network will be quantified and analyzed. Our statistical analysis framework will be made general enough to fully utilize the 3D geometric reconstruction data. This will allow us to provide a user interface to quickly enter custom queries for an open-ended statistical exploration of the data by us and by third party users.

D. Broader Impact Activities

Student training and under-represented groups: This project will support the training of three Ph.D. students (two directly supported, and one other supported by other funds). Our team's previous work in this area also has a strong history of involving Master's-level students (not directly funded; seven M.S. theses have come from the group) and undergraduate students (Pis Keyser and Choe have supervised thirteen undergraduates in the past 7 years) under Research Experience for Undergraduates (REU) program (REU site, #0353957, #0649233, PI: V. E. Taylor), and the Computing Research Association(CRA)'s Distributed Research Experiences for Undergraduates (DREU) that targets under-represented groups in computer science. The Clemson team also has significant experience with undergraduate research and mentoring, through their own REU site (#0850695, PI: Larry Hodges). We will continue our effort in this direction.

The resulting database and geometric reconstructions, and the eye-tracking framework also give us an opportunity to reach out to K-12 students. Pis Keyser and Choe have had significant experience in interactions with high-school level groups, with Keyser having given multiple presentations to such groups, and Keyser and Choe organizing a computer science contest/outreach day (hosted by the department) for high school students each of the last three years. The topic of last year's contest was "Computational Thinking". In relation to the current project, we will organize a interactive tracing contest based on the eye-tracking framework, in year 2 and year 3. We expect 10 to 15 teams from regional high schools to participate, with up to 5 students in each team.

Our past efforts have demonstrated that we are able to incorporate under-represented groups in the project, such as women (currently two of our graduate students are female, and two women graduated recently) and minority students (primarily at the undergraduate level). We will continue to involve and reach out to these groups in our proposed project.

Dissemination: A separate NSF-funded project (#0905041: *CRCNS data sharing: Whole Mouse Brain Neuronal Morphology and Neurovasculature Browser*) is providing a method for distribution, browsing, and basic visualization of whole data sets. In this proposed project, we will add on to that system the geometric reconstructions and validation data (in years 2-3) we develop here. Also, we will release the eye-tracking data, which is expected to revolutionize the way eye tracking is used in scientific data mining. As the reconstruction code becomes mature, we will release it in an open-source environment, such as <http://sourceforge.net>. Software tools will be continually maintained through an active project on SourceForge, which provides support for development activity such as bug tracking and user forums. Existing code (for noise removal) from our group has already been released on SourceForge. We will actively advertise this new resource through various channels: mailing lists, news briefs in scientific publications (such as Science, Nature, and also popular science magazines), personal contact, and demos and exhibitions at scientific meetings. We will also organize short courses and workshops to expand and support the user community. For the above, co-PI Choe will depend on his extensive experience in most of the above activities (Topographica software project <http://topographica.org>, book-related web page and promotion <http://computationalmaps.org> [54], exhibitions as part of the NIH/NIMH Human Brain Project, a minisymposium at the Society for Neuroscience meetings [we organized one in 2008], and a workshop at the Computational Neuroscience meeting in 2010).

COORDINATION PLAN

1. Specific roles of senior personnel

The five major sections in the research plan (§C) are listed below, with PI/Co-PI responsibilities indicated (the abbreviation TAMU will be used to represent Texas A&M University). The approximate timeline of each task is also listed at the end of each task.

§C.1 Acquire vascular and neuronal morphology data from whole mouse brains, using the KESM:

- Abbott (TAMU) will be responsible for specimen preparation.
- Choe (TAMU) will be responsible for KESM data acquisition.
- Keyser (TAMU) will be responsible for image processing and noise removal, with the assistance of the TAMU graduate student.
- Year 1

§C.2 Develop vector-based tracing algorithms for rapid and accurate tracing of fiber-like structures:

- Keyser (TAMU) will design and develop the algorithms, with the assistance of the TAMU graduate student.
- Years 1 and 2

§C.3 Develop a framework for the use of eye-tracking technology in interactive tracing:

- Duchowski (Clemson) and Keyser (TAMU) will coordinate this effort, designing and developing the framework. The TAMU and Clemson graduate students will be involved.
- Abbott (TAMU) will serve as a domain expert in the eye-tracking experiments.
- Years 1 and 2

§C.4 Develop a confidence-based rapid validation method using eye-tracking:

- Duchowski (Clemson) and Choe (TAMU) will design and implement the framework, with the assistance of the graduate students.
- Abbott (TAMU) will serve as a domain expert in the eye-tracking experiments, and provide ground truth data on selective samples.
- Years 2 and 3

§C.5 Conduct large-scale statistical analysis of morphological variability in neuronal and vascular networks in the mouse brain:

- Choe (TAMU) will take the lead on this task, with Keyser (TAMU) assisting, in designing and developing the statistical analysis framework.
- Abbott (TAMU) will provide scientific expertise in evaluating the statistical data.
- Year 3

2. Management across institutions and disciplines

Choe (TAMU) will be the main coordinator for the entire project.

Cross-institutional management: Choe (TAMU) and Duchowski (Clemson) will be the main contact across TAMU and Clemson. At TAMU, Choe, Abbott, and Keyser are members of the Brain Networks Laboratory (director: Choe), where a weekly meeting is held on campus to get updates on the progress and plan for the week ahead. The meeting is held in a room equipped

with a networked PC, an LCD projector, and a telephone, which provides sufficient environment for teleconference with the Clemson team. Specific recurring activities will be as follows.

- The TAMU and the Clemson team will hold a bi-weekly teleconference using this facility. All senior personnel and the two graduate students, and other participants (such as REU students) will be present. Each participant will provide a brief report of the past weeks' activity, and present plans for the upcoming week. The meetings are expected to last one hour.
- At the beginning of each semester (Spring, Summer, and Fall), a planning meeting will take place to carve out detailed activities for each semester.
- At the conclusion of each semester (Spring, Summer, and Fall), a three-hour long evaluation meeting will be scheduled, with research presentations from both TAMU and Clemson.

Cross-disciplinary management: The computer science PI/Co-PIs Choe, Keyser, and Duchowski will work closely with the neuroscience Co-PI Abbott. For Choe, Abbott, and Keyser, interaction is already on-going, with at least one meeting per week and more often as needs arise. Abbott and Duchowski will mainly coordinate their efforts through the bi-weekly teleconference.

3. Coordination mechanisms

Annual workshops: We will organize an annual workshop, for year 2 and year 3, on the subject of eye-tracking applied to neuroinformatics. We are already approved by conference organizers to organize a workshop at the Computational Neuroscience meeting in San Antonio (August, 2010). Candidate host conferences include: Computational Neuroscience meeting, Neuroinformatics meeting, and Society for Neuroscience meeting. Choe and Abbott have experience in organizing a minisymposium at the Society for Neuroscience meeting, in 2008, and they are co-organizers for this year's workshop at the Computational Neuroscience meeting.

Site visits: At least once a year, the TAMU and/or the Clemson team will visit the counterpart, to coordinate the project. Specific activities during the site visits will be:

- Discussion on scientific issues.
- Integrate modules developed separately at each side.
- Solve any technical issues (e.g., with the eye tracking instrument).
- Tutorial on software and database usage.
- Research presentation.
- Planning.

Site visits will typically occur during the summer, when the schedule is relatively less restricted than during regular semesters.

Software development and sustainability: In order to ensure continued support of the software platform developed through this project, we will design and implement a protocol for software development, documentation, and education. All members of the development team (both TAMU and Clemson) will be trained to abide by this protocol. Involvement of external developers as well as department-funded graduate students will ensure continued support beyond the funding period. We expect maintenance to require less effort once the initial platform is implemented during the project period, and we believe we can build on the work of our recently funded CRCNS data sharing project to make dissemination easier. We will also seek continued funding from follow-up grants and from collaborations.

Data exchange: Both the TAMU team and the Clemson team will operate Subversion (SVN) servers to exchange code and data. For large volumes of data, we will use the budget requested for storage to purchase hard drives (2TB drives cost about \$200) and directly ship them via courier service.

Target audience: Data exchange will go beyond the immediate project teams at TAMU and Clemson. The target audience of our project is diverse. Stephen J. Smith (Stanford, Array Tomography), George Karniadakis (Brown University, vascular transport modeling), and Graziano Fiorito (Stazione Zoologica, Italy, octopus neuronal morphology) will be the immediate users of our framework. Other groups who have shown interest in our work include Peter Hunter (CellML, Univ. of Auckland, New Zealand), Ching-Long Lin (lung research, Univ. of Iowa), Gro Amdam (honey bee brain research, Arizona State Univ.), and researchers locally at Texas A&M such as Thierry Lints (songbird brain research).

4. Budget line items

Domestic travel budget includes funding to support cross-site visits, at least once a year, and support for conference trips where the TAMU and Clemson personnel can meet.

Telecommunication lines will be established through free internet services such as Skype, thus no budget item is listed for this.

REFERENCES CITED

- [1] Abbott, L. C., and Sotelo, C. (2000). Ultrastructural analysis of catecholaminergic innervation in weaver and normal mouse cerebellar cortices. *Journal of Comparative Neurology*, 426:316–329.
- [2] Al-Kofahi, K. A., Lasek, S., Szarowski, D. H., Pace, C. J., Nagy, G., Turner, J. N., and Roysam, B. (2002). Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Transactions on Information Technology in Biomedicine*, 6:171–187.
- [3] Allen Institute for Brain Science, Allen brain atlas. <http://www.brain-map.org/>.
- [4] Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). NeuroMorpho.Org: A central resource for neuronal morphologies. *Journal of Neuroscience*, 27:9247–9251.
- [5] Boldak, C., Rolland, Y., and Toumoulin, C. (2003). An improved model-based vessel tracking algorithm with application to computed tomography angiography. *Journal of Biocybernetics and Biomedical Engineering*, 3:41–64.
- [6] Boslaugh, S., and Watters, P. A. (2008). *Statistics in a Nutshell*. Sebastopol, CA: O'Reilly Media, Inc.
- [7] Braitenberg, V., and Schuz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer. Second edition.
- [8] Can, A., Shen, H., Turner, J. N., Tanenbaum, H. L., and Roysam, B. (1999). Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *IEEE Transactions on Information Technology in Biomedicine*, 3:125–138.
- [9] Cannon, R. C., Turner, D. A., Pyapali, G. K., and Wheal, H. V., Duke-Southampton archive of neuronal morphology. <http://neuron.duke.edu/cells/index/topindex.html>.
- [10] Cannon, R. C., Turner, D. A., Pyapali, G. K., and Wheal, H. V. (1998). An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods*, 84:49–54.
- [11] Capowski, J. J., editor (1989). *Computer Techniques in Neuroanatomy*. Plenum.
- [12] Cell Centered Database. <http://ncmir.ucsd.edu/CCDB/>.
- [13] Choe, Y., Abbott, L. C., Han, D., Huang, P.-S., Keyser, J., Kwon, J., Mayerich, D., Melek, Z., and McCormick, B. H. (2008). Knife-edge scanning microscopy: High-throughput imaging and analysis of massive volumes of biological microstructures. In Rao, A. R., and Cecchi, G., editors, *High-Throughput Image Reconstruction and Analysis: Intelligent Microscopy Applications*. Boston, MA: Artech House. In press.
- [14] Denk, W., and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biology*, 19:e329.
- [15] DeOrio, A., and Bertacco, V. (2009). Human computing for EDA. In *Proceedings of the 46th Annual Design Automation Conference*, 621–622.
- [16] Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory & Practice*. London, UK: Springer-Verlag, Inc.

- [17] Duchowski, A. T. (2004). Eye tracking. In Bainbridge, W. S., editor, *Encyclopedia of Human-Computer Interaction*, 251–254. Great Barrington, MA: Berkshire Publishing Group.
- [18] Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory & Practice*. London, UK: Springer-Verlag, Inc. Second edition.
- [19] Duchowski, A. T., Driver, J., Robbins, A., Ramey, B. N., Tan, W., and Jolaoso, S. (2010). Scanpath Comparison Revisited. In *Eye Tracking Research & Applications (ETRA)*, (To appear). Austin, TX: ACM.
- [20] Duchowski, A. T., and McCormick, B. H. (1998). Gaze-Contingent Video Resolution Degradation. In *Human Vision and Electronic Imaging III*. Bellingham, WA: SPIE.
- [21] Duchowski, A. T., Medlin, E., Cournia, N., Murphy, H., Gramopadhye, A., Nair, S., Vora, J., and Melloy, B. (2002). 3D eye movement analysis. *Behavior Research Methods, Instruments, and Computers (BRMIC)*, 34:573–591.
- [22] Eshera, M. A., and Fu, K. S. (1986). An image understanding system using attributed symbolic representation and inexact graph-matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:604–618.
- [23] Fiala, J. C. (2005). Reconstruct: A free editor for serial section microscopy. *Jounral of Microscopy*, 218:52–61.
- [24] Fiala, J. C., and Harris, K. M. (2001). Extending unbiased stereology of brain ultrastructure to three-dimensional volumes. *J. Am. Med. Inform. Assoc.*, 8:1–16.
- [25] Frangi, A. F., Niessen, W. J., Hoogeveen, R. M., Walsum, T. V., and Viergever, M. A. (1999). Model-based quantification of 3-D magnetic resonance angiographic images. *IEEE Transactions on Medical Imaging*, 18:946–956.
- [26] Han, D., Keyser, J., and Choe, Y. (2009). A local maximum intensity projection tracing of vasculature in Knife-Edge Scanning Microscope volume data. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 1259–1262.
- [27] Hayworth, K. (2008). Automated creation and SEM imaging of Ultrathin Section Libraries: Tools for large volume neural circuit reconstruction. In *Society for Neuroscience Abstracts*. Washington, DC: Society for Neuroscience. Program No. 504.4.
- [28] Hayworth, K., and Lichtman, J. W. Automatic Tape-Collecting Lathe Ultramicrotome (ATLUM), http://www.mcb.harvard.edu/lichtman/ATLUM/ATLUM_web.htm.
- [29] Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Crown Business.
- [30] Jacobowitz, D. M., and Abbott, L. C. (1997). *Chemoarchitectonic Atlas of the Developing Mouse Brain*. Boca Raton, FL: CRC Press.
- [31] Jain, V., Murray, J. F., Roth, F., Turaga, S., Zhigulin, V., Briggman, K. L., Helmstaedter, M. N., Denk, W., and Seung, H. S. (2007). Supervised learning of image restoration with convolutional networks. In *IEEE 11th International Conference on Computer Vision (ICCV 2007)*, 1–8.

- [32] Josephson, S., and Holmes, M. E. (2002). Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. In *Eye Tracking Research & Applications (ETRA) Symposium*, 43–49. New Orleans, LA: ACM.
- [33] Josephson, S., and Holmes, M. E. (2006). Clutter or Content? How On-Screen Enhancements Affect How TV Viewers Scan and What They Learn. In *Eye Tracking Research & Applications (ETRA) Symposium*, 155–162. San Diego, CA: ACM.
- [34] Jurrus, E., Tasdizen, T., Koshevoy, P., Fletcher, P. T., Hardy, M., Chen, C.-B., Denk, W., and Whitaker, R. (2006). Axon tracking in serial block-face scanning electron microscopy. In Metaxas, D., Whitaker, R., Rittscher, J., and Sebastian, T., editors, *Proceedings of the 1st Workshop on Microscopic Image Analysis with Applications in Biology*, 114–119.
- [35] Kwon, J., Mayerich, D., Choe, Y., and McCormick, B. H. (2008). Lateral sectioning for knife-edge scanning microscopy. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 1371–1374.
- [36] Kwon, J.-R. (2009). *Acquisition and Mining of the Whole Mouse Brain Microstructure*. PhD thesis, Department of Computer Science, Texas A&M University.
- [37] MacKenzie-Graham, A., Jones, E. S., Shattuck, D. W., Dinov, I. D., Bota, M., and Toga, A. W. (2003). The informatics of a C57BL/6J mouse brain atlas. *Neuroinformatics*, 1:397–410.
- [38] Martone, M. E., Gupta, A., Wong, M., Qian, X., Sosinsky, G., Ludscher, B., and Ellisman, M. H. (2002). A cell-centered database for electron tomographic data. *J. Struct. Biol.*, 138:145–155.
- [39] Mayerich, D., Abbott, L. C., and Keyser, J. (2008). Visualization of cellular and microvessel relationship. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization)*, 14:1611–1618.
- [40] Mayerich, D., Abbott, L. C., and McCormick, B. H. (2008). Knife-edge scanning microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain. *Journal of Microscopy*, 231:134–143.
- [41] Mayerich, D., and Keyser, J. (2008). Filament tracking and encoding for complex biological networks. In *Proceedings of ACM Symposium on Solid and Physical Modeling*, 353–358.
- [42] Mayerich, D., Kwon, J., Choe, Y., Abbott, L., and Keyser, J. (2008). Constructing high-resolution microvascular models. In *Proceedings of the 3rd International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB 2008)*. Online.
- [43] Mayerich, D., McCormick, B. H., and Keyser, J. (2007). Noise and artifact removal in knife-edge scanning microscopy. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 556–559.
- [44] Mayerich, D. M., and Keyser, J. (2008). Filament tracking and encoding for complex biological networks. In *Proceedings of Solid Modeling*. To appear.
- [45] McCormick, B. H. (2003). The knife-edge scanning microscope. Technical report, Department of Computer Science, Texas A&M University. <http://research.cs.tamu.edu/bnl/>.

- [46] McCormick, B. H., System and method for imaging an object. USPTO patent #US 6,744,572 (for Knife-Edge Scanning; 13 claims).
- [47] McCormick, B. H., Abbott, L. C., Mayerich, D. M., , Keyser, J., Kwon, J., Melek, Z., and Choe, Y. (2006). Full-scale submicron neuroanatomy of the mouse brain. In *Society for Neuroscience Abstracts*. Washington, DC: Society for Neuroscience. Program No. 694.5. Online.
- [48] McCormick, B. H., Abbott, L. C., Mayerich, D. M., , Keyser, J., Kwon, J., Melek, Z., and Choe, Y. (2006). Full-scale submicron neuroanatomy of the mouse brain. In *Society for Neuroscience Abstracts*. Washington, DC: Society for Neuroscience. Program No. 694.5. Online.
- [49] McCormick, B. H., and Mayerich, D. M. (2004). Three-dimensional imaging using Knife-Edge Scanning Microscope. *Microscopy and Microanalysis*, 10 (Suppl. 2):1466–1467.
- [50] Melek, Z., Mayerich, D., Yuksel, C., and Keyser, J. (2006). Visualization of fibrous and thread-like data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1165–1172.
- [51] Melnik, S., Garcia-Molina, H., and Rahm, E. (2002). Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the 18th International Conference on Data Engineering*, 117–128.
- [52] Micheva, K., and Smith, S. J. (2007). Array tomography: A new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron*, 55:25–36.
- [53] MicroBrightField, Inc., Autoneuron. <http://www.mbfbioscience.com/autoneuron>.
- [54] Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Berlin: Springer. URL: <http://www.computationalmaps.org>.
- [55] Mikula, S., Trotts, I., Stone, J. M., and Jones, E. G. (2007). Internet-enabled high-resolution brain mapping and virtual microscopy. *Neuroimage*, 35:9–15.
- [56] Noton, D., and Stark, L. (1971). Eye Movements and Visual Perception. *Scientific American*, 224:34–43.
- [57] Pham, D. L., Xu, C., and Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337.
- [58] Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2000). Image registration by maximization of combined mutual information and gradient information. In *Lecture Notes In Computer Science, Vol. 1935; Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, 452–461. London: Springer.
- [59] Privitera, C. M., and Stark, L. W. (2000). Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(9):970–982.
- [60] Quinn, A., and Benderson, B. (2009). A taxonomy of distributed human computation. Technical Report HCIL-2009-23, Human-Computer Interaction Lab, University of Maryland, College Park.

- [61] Rhyu, I. J., adn D. B. Walker, L. C. A., and Sotelo, C. (1999). An ultrastructural study of granule cell/Purkinje cell synapses in tottering (tg/tg), leaner (tg^{la}/tg^{la}) and compound heterozygous tottering/leaner (tg/tg^{la}) mice. *Neuroscience*, 90:717–728.
- [62] Rosen, G. D., Williams, A. G., Capra, J. A., Connolly, M. T., Cruz, B., Lu, L., Airey, D. C., Kulakarni, A., and Williams, R. W. (2000). The mouse brain library@www.mbl.org. In *Proceedings of the 14th International Mouse Genome Meeting, Narita, Japan*, C6 [online]. http://www.imgs.org/abstracts/2000abstracts/C6_p039E_RobWilliams.shtml, <http://mbl.org/>.
- [63] Sadashivan, S., Greenstein, J. S., Gramopadhye, A. K., and Duchowski, A. T. (2005). Use of Eye Movements as Feedforward Training for a Synthetic Aircraft Inspection Task. In *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems*, 141–149. Portland, OR: ACM Press.
- [64] Santella, A., and DeCarlo, D. (2004). Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest. In *Eye Tracking Research & Applications (ETRA) Symposium*, 27–34. San Antonio, TX: ACM.
- [65] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1:e42.
- [66] Tsai, P. S., Friedman, B., Ifarraguerri, A. I., Thompson, B. D., Lev-Ram, V., Schaffer, C. B., Xiong, Q., Tsien, R. Y., Squier, J. A., and Kleinfeld, D. (2003). All-optical histology using ultrashort laser pulses. *Neuron*, 39:27–41.
- [67] von Ahn, L. (2006). Games with a purpose. *IEEE Computer*, 39:96–96.
- [68] von Ahn, L., Kedia, M., and Blum, M. (2006). Peekaboom: A game for locating objects in images. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2006)*, 55–64.
- [69] von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321:1465–1468.
- [70] Vora, J., Nair, S., Gramopadhye, A., Duchowski, A. T., Melloy, B., and Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: Presence and comparison studies. *Applied Ergonomic*, 33:559–570.
- [71] Warfield, S. J., Zou, K. H., and Wells, W. M. (2002). Validation of image segmentation and expert quality with expectation-minimization algorithm. In *Lecture Notes In Computer Science, Vol. 2488; Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, 298–306.
- [72] West, J. M., Haake, A. R., Rozanski, E. P., and Karn, K. S. (2006). eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences. In *Eye Tracking Research & Applications (ETRA) Symposium*, 149–154. San Diego, CA: ACM.
- [73] Yoo, T., Ackerman, N. J., and Vannier, M. (2000). Toward a common validation methodology for segmentation and registration algorithms. In *Lecture Notes In Computer Science, Vol. 1935; Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, 422–431. London: Springer.