The Connectome
discovering the wiring diagram of the mammalian brain

Launched 2007
Status Ongoing
Participants
Johanna Beyer, VRVis, IIC and Technische Universität Wien
Rusty Blue, Kitware Inc.
Kenneth Blum, Center for Brain Science
Michael Cohen, Microsoft Research
Markus Hadwiger, VRVis
Won-Ki Jeong, IIC
Charles Law, Kitware
Jeff Lichtman, Center for Brain Science
Hanspeter Pfister, IIC
Clay Reid, Center for Brain Science
Will Schroeder, Kitware
Amelio Vázquez-Reina, Tufts University and IIC
Bo Wang, University of Pittsburgh and IIC

An initial rough segmentation distinguishes the axons in a slice of data from electron microscope imaging of thin films of brain tissue. This segmentation, done by an application developed at Microsoft Research, is a starting point for the Active Ribbon method of segmenting and tracking finer features in neural tissue.

Selected accomplishments

- Development of NeuroTrace, a system for scalable electron microscope image segmentation and interactive visualization using graphics processing units (GPUs)
- Development of SSECRETT, the Serial SECtioning and REconstruction Tool (by Kitware Inc.)
- A new algorithm to trace neural pathways, presented at the 2009 Conference on Vision and Pattern Recognition
- Application of HDView (by Microsoft Research) to stitch together and interactively view images containing billions of pixels over the Web
- Substantial progress on an algorithm to automatically detect synapses in electron microscope data
- Ongoing work on novel segmentation, visualization, and analysis algorithms for multi-terabyte neural image data
The Harvard Center for Brain Science (CBS) and the IIC have been working together since 2007, in collaboration with Microsoft Research and Kitware Inc., on the Connectome Project. This ambitious effort aims to apply biology, computer science and software engineering to the grand challenge of determining the detailed neural circuitry of the brain.

The goals of this project have been:

- Close collaboration between scientists at the CBS and computer science researchers at the IIC
- Development of stitching and visualization tools for gigapixel-size images
- Software for segmentation of neural processes in multi-terabyte image stacks
- Tools to visualize and “fly through” images stacks of neural circuitry
- Algorithms to automatically detect neural features such as synapses
- Tools to visualize extracted neural processes and features in two and three dimensions
- Outreach to the scientific neuroscience community, including availability of data and open-source software, and workshops and seminars at Harvard.

A boundary-spanning problem rich in challenge

The nervous system is unique among the organ systems in animals because of the vast number of interconnections between its individual cells (synapses) and the diversity of its cell types (neurons). One cubic millimeter of cerebral cortex contains roughly 50,000 neurons, each of which establishes approximately 6,000 synapses with neighboring cells. These 300 million interconnections are highly specific: Neurons innervate some target cells but avoid others. The complexity is further amplified by the fact that neurons come in many kinds. For example, some neurons make excitatory connections, while others establish inhibitory ones. It is thought that there are well over 100 types of neurons, differing in shape, neurochemistry, and function. In action, each neuron integrates the signals from hundreds or thousands of synaptic signals, and this history determines whether or not it will send an electrical signal to its target cells. A cubic millimeter is but a minuscule part of the full circuitry, which is estimated to contain $60 \times 10^{12}$ synaptic connections.

A cubic millimeter [of cerebral cortex, containing roughly 50,000 neurons,] is but a minuscule part of the full circuitry, which is estimated to contain $60 \times 10^{12}$ synaptic connections.

Connectome: Timeline

<table>
<thead>
<tr>
<th>date</th>
<th>milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007–08</td>
<td>Microsoft funds the Connectome collaboration with a $2 million grant to the Center for Brain Science and IIC</td>
</tr>
<tr>
<td></td>
<td>Kitware Inc. starts development on 2D slice viewer (SSECRETT)</td>
</tr>
<tr>
<td></td>
<td>First Connectome @ Harvard meeting with ~25 participants</td>
</tr>
<tr>
<td>2008–09</td>
<td>Won-Ki Jeong and Amelio Vázquez-Reina join Connectome team</td>
</tr>
<tr>
<td></td>
<td>Development of active-contours segmentation method</td>
</tr>
<tr>
<td></td>
<td>Development of interactive 3D visualization and segmentation application using GPUs</td>
</tr>
<tr>
<td></td>
<td>Research on automatic synapse identification method</td>
</tr>
<tr>
<td>2009–10</td>
<td>Research on simultaneous segmentation of multiple neural processes</td>
</tr>
<tr>
<td></td>
<td>Parallelization of 3D visualization and segmentation application on GPU cluster with client-server architecture</td>
</tr>
</tbody>
</table>
Given these large numbers, it is not surprising that the circuits underlying even the simplest of behaviors are not understood. Until recently, attempts to describe fully such circuits were never even seriously entertained, as it was considered too numerically complex. Now, however, new forms of laser-scanning optical microscopy and semi-automated electron microscopy allow high resolution imaging of “connectomes”—that is, the full sets of neural wires that connect neurons and their targets. Determining the detailed connections in brain circuits is a fundamental unsolved problem in neuroscience. Understanding this circuitry will enable brain scientists to confirm or refute existing models, develop new ones, and come closer to an understanding of how the brain works.

**A demand for new technology**

In 2004, Harvard established the Center for Brain Science on its Cambridge campus, an interdisciplinary center aimed at understanding neural circuits—their structure, their function, and how they are changed during development, aging, and disease. Jeff Lichtman and CBS launched the Connectome Project to determine the complete, detailed wiring diagrams of neural circuits. Support to initiate the Connectome Project was generously provided by private gifts to Harvard and CBS.

Initial efforts to generate experimental data sets from simple 3D neural circuits using optical imaging and electron microscopy have been successful. The main bottleneck for the Connectome Project going forward is analyzing the vast streams of image information that we can now acquire. In 2007, led by Clay Reid and Jeff Lichtman, CBS proposed a collaboration with the new IIC to develop the hardware and software necessary to handle and analyze the increasingly large data sets being produced in the Connectome Project. Hanspeter Pfister and the team he assembled at the IIC are addressing the initial hardware needs and exploring software solutions.

**Project ABCs: Acquire, Build, Compute**

CBS efforts are focused on two imaging modalities, one using photons and the other using electrons. With high-resolution optical techniques, investigators study transgenic mice that express a number of different fluorescent proteins in individual neurons. They image the branching patterns and connections of all the axons within a region of the nervous system. They have started with connections between nerve cells and targets in the peripheral nervous system, because these are simpler and more accessible than the synapses in the brain. By now they have reconstructed all of the axons that project to small muscles and are beginning similar efforts in autonomic ganglia, where the axons are making synaptic connections with other neurons rather than muscle fibers. The size of these data sets is in the range of hundreds of gigabytes, and the CBS researchers have found that current tools are unlikely to scale well to the larger data sets that are now being acquired.

At the same time, CBS is developing tools to automate the acquisition of sequential high-resolution images of the central nervous system using electron microscopy. While electron microscopy is well known for its high-resolution images of materials (fractions of a nanometer, nm), the approach they wish to take is a lower-power, large-field image at moderate resolution (5–10 nm). Each thin section of one cubic millimeter of tissue at 5-nanometer resolution would require images with 10 to 20 gigapixels (or 10,000 to 20,000 megapixels). If one were to reconstruct all the synaptic circuitry...
in this one cubic millimeter of brain cortex, one would need a set of serial images spanning a millimeter in depth, each probably no more than 30 nanometers deep. Thus the 1 mm depth would require 33,000 images, each with 10 to 20 gigapixels, or about 1 petabyte (1 million gigabytes) of data. Visualizing, identifying the neuronal structures, and analyzing their connectivity is the largest-scale visual computing problem of which we are aware, approximately five orders of magnitude beyond the current state of the art.

**Visualizing large images: A collaboration with Microsoft Research**

At Microsoft Research, work has proceeded to stitch together and then interactively view images containing billions of pixels. Once these gigapixel-size images are organized into a hierarchical pyramid, the Microsoft HD View application can stream requested imagery over the Web for viewing. This allows exploration of both large-scale and very fine-scale features. The image above shows a walkthrough of the result. Microsoft has also begun preliminary experiments to see if thousands of images from a two-dimensional slice of brain tissue could be assembled using the MSR image stitcher. Once the images are captured and stitched, multiple slices of a sample must be stacked to assemble them into a coherent volume. Perhaps the most difficult task at that point is extracting the individual strands of neurons. This task is known as image segmentation.

**Semi-automatic segmentation with Active Ribbons**

Amelio Vázquez-Reina, a Ph.D. student at Tufts University who is working full-time on the Connectome Project at CBS under the supervision of Hanspeter Pfister, has tackled the problem of segmenting and tracking individual neural connections using geometric deformable models. Geometric deformable models have been successfully applied in the past in medical imaging applications. Their success is mostly attributed to their ability to easily adapt to shapes of unknown topology.

The ability to constrain the geometry of deformable models can be useful when information about the expected shape or positioning of the objects in a scene is known *a priori*. An example of this occurs when segmenting neural cross sections in electron microscopy. Such images often contain multiple nested boundaries separating regions of homogeneous intensities. For these applications, multiphase level sets provide a partitioning framework that allows for the segmentation of multiple deformable objects.

In this project, we show how to define elastic couplings between multiple level set functions to model ribbon-like partitions. We build such couplings using dynamic force fields that can depend on the image content and relative location and shape of the level set functions. To the best of our knowledge, this is the first work that shows a direct way of geometrically constraining multiphase level sets for image segmentation.

We applied our technique to the EM images provided by Jeff Lichtman’s group, which were acquired using the Automatic Tape-Collecting Lathe Ultramicrotome (ATLUM) device built at the lab. In each image section, cells, mitochondria and other intracellular organelles have a membrane of homogeneous intensity and varying thickness. This allows us to make the reasonable assumption that any section can be decomposed into a family of ribbons. Once the deformable model is built, we can generate a pair of force fields that push the ribbon towards the desired features on the images to be segmented. The features of interest in our problem are the cellular boundaries of each neural process.
We accomplish this in our project by estimating the cellular boundaries given the position of the ribbon. Once we have estimated the location of each side of the cellular boundaries, we can generate another force field for our deformable model that will push the active ribbon to these features. The joint action of all the above-mentioned force fields allows our active ribbon model to successfully segment cellular boundaries on the images acquired from the ATLUM device.

**Automatic segmentation with ConnecTrace**

Although active ribbons are a robust method for the segmentation and tracking of individual cells, the deformable ribbon needs to be initialized close to the desired solution to avoid undesired results. For this reason, we have developed the ConnecTrace software, which provides an alternative automatic segmentation of a full data set.

The first step in the segmentation pipeline is denoising. ConnecTrace uses Total Variation Diminishing (TVD) denoising, an edge-preserving denoising method known for its ability to filter out structures with high level-set curvature. This denoising step is implemented in ConnecTrace to prevent clusters of synaptic vesicles from confusing the detection of cellular boundaries. The second step in the processing pipeline is the thresholding of the cellular membranes. Neural membranes can be roughly described as homogeneous elongated dark regions that separate image areas of light intensities. The TVD phase has the secondary effect of smoothing cellular membranes. This smoothing makes it possible to extract most of the cellular boundaries within a section via intensity thresholding. The thresholding step yields a binary mask where the cellular boundaries are separated from the background clutter of the image. Finally, boundary skeletonization provides ConnecTrace with a medial axis representation of the cellular membranes.

ConnecTrace also provides a graphical user interface that allows the user to validate the segmentation results from the above segmentation pipeline. The results are shown to the user in the form of a transparent colorful layer on top of each neural process. ConnecTrace provides several interaction modes that allow the user to fix by drawing or erasing the segmentation errors that resulted from the segmentation pipeline.

The ATLUM device can section and image a block of brain tissue. In such images, the human eye is able to distinguish neurons and small structures such as vesicles.
ConnecTrace has been under test since spring 2009 with as many as five undergraduate students using it for tracing hundreds of neurons in ATLUM data sets. The students have reported that ConnecTrace outperforms other standard software for the tracing of neurons in terms of accuracy and ease of use.

**Automatic identification of synaptic connections**

Another project tackles the problem of automatically identifying synapses (by finding regions of high vesicle densities) in scanning electron microscope (SEM) image data. Synaptic regions have large collections of vesicles, small sacs within cells, which tend to aggregate when a neuron fires a signal. The vesicles store various neurotransmitters that are then released at the synapse. The procedure for distinguishing a synapse typically encompasses the following steps: (1) detection of vesicles (which most commonly appear on SEM images as circles or rings), (2) clustering of vesicles, (3) detection of neuron cell boundaries (which appear as strong edges), and (4) identification of clusters of vesicles juxtaposed with the cell boundaries.

Our software method finds clusters of vesicles by classifying feature vectors extracted from SEM images.
First, it scans localized regions of input images and identifies vesicles via a process known as *histogram of oriented gradients* (HOG) feature extraction. Then a support vector machine (SVM) is used for the classification of the feature vectors. Finally, a low-pass filter and threshold is applied to the output classification images to generate vesicle density maps that indicate regions with high likelihood of containing a synapse.

Once the feature vectors for each individual vesicle have been collected, they are fed into a previously trained SVM classifier that tags each vector with one of two possible classes: vesicle or not-vesicle. Finally, the resulting SVM classification mask, with one value per vector that measures the certainty with which the classification was made, is low-pass filtered and upsampled to match the resolution of the original image.

**Interactive visualization and segmentation of petascale-volume data**

As noted earlier, the ATLUM device at CBS produces petascale high-resolution EM images. Those data sets are crucial for reconstruction of detailed neural connections, but they pose very challenging problems for 3D segmentation and visualization. First, the current practice for segmentation of objects of interest in EM data sets is a mostly manual process that is very labor-intensive and time-consuming. Even though there have been research efforts to develop automated EM segmentation algorithms, they are not robust enough to deal with such common artifacts of real data sets as noise and misalignment. Second, the complex structure of nerve cells makes direct volume rendering of EM data sets very difficult. Transfer functions (mappings of data attributes for visual representation) based solely on image intensity and gradient result in cluttered renderings, degrading the quality of visualization. Finally, it is important that segmentation and visualization algorithms be scalable to cope with ever-increasing data volumes while maintaining interactive performance, so that the user can perform manual modifications at any time if necessary.

Therefore, with our Austrian collaborators, the VRVis visualization group, we have developed NeuroTrace, a hardware-accelerated segmentation and visualization system for neural processes in high-resolution EM data. NeuroTrace employs 2D segmentations and a 3D tracking method with weighted path extrapolation to robustly trace a 3D centerline of a neural pathway along non-axis-aligned slices. NeuroTrace couples this with a novel volume-rendering method with on-demand filtering, which provides better visual cues to find regions of interest in complex EM data sets. These algorithms are efficiently implemented on graphics-processing units (GPUs) for interactive performance. All these steps are combined in an integrated workflow that provides a unified user-interface to easily explore large EM volumes and extract neural processes at interactive rates, a

![Color-mapped results show vesicle detection in sections of EM volumes. Vesicle identifications appear in red and yellow; areas where vesicles are not detected appear in cool blue and green.](image-url)
significant advance over the previous state-of-the-art approach in neurobiology.

**Neural process segmentation with NeuroTrace: The 3D challenge**

It is especially important to be able to track neurons and their connections in a dense 3D volume of tissue image data, a process akin to following a web of overgrown pathways in a dense jungle. In NeuroTrace, we compute 3D segmentations using a combination of 2D neural membrane segmentations and 3D centerline tracking. Because level set segmentation is very sensitive to initialization, we propose a novel active ribbon formulation with an additional constraint based on image correspondence between current and previous 2D slices. The idea is that we add a new force field to the original active ribbon model that maps one image to another using image correspondence. This allows us to robustly initialize the location of the neural membranes on subsequent slices.

Once 2D segmentation is done, we extrapolate the next point along the centerline of the neural process, which may or may not be aligned along an axis. This may seem straightforward, but in fact a challenging problem because the position and the orientation of the next 2D slice should be estimated from the current slice position and the segmentation result.

To tackle this problem, we are exploring a two-step method that consists of estimation and correction steps. In the estimation step, the tangent direction at the last center point is computed using a one-sided finite difference method. We also keep the previous tracking direction. The new tracking direction is then the weighted average between those two vectors where the weight controls the amount of history used to determine the current tracking direction. Once we compute a new tracking direction, a temporary new center position of the next slice can be estimated by extrapolation. The local frame of the previous slice is then projected onto the new center and a new 2D slice is resampled from the volume data to be used for segmentation. Finally, in the correction step, the temporary center point is updated by the correct center of the segmented neural membrane.

For interactive performance, we implemented both a multiphase level set solver and deformable image registration on the GPU. Our GPU level set solver iteratively updates the level set only in active regions using a block-based narrow band until the pre-computed number of iterations is reached or converges to a steady state. In extending the single level set method to multiphase level sets, we need to evaluate the correct distance between two level sets to keep the topology of the active ribbon consistent. To quickly compute the distance fields, we use a GPU-based
parallel Eikonal solver. The deformable image registration method is also implemented on the GPU using fast texture hardware.

**Visualization of EM volume and segmented neural processes with NeuroTrace**

Electron-microscope data is dense and heavily textured, and exhibits a complex structure of interconnected nerve cells, and has a low signal-to-noise ratio. Therefore, standard volume rendering results in cluttered images that make it hard to identify regions of interest (ROIs) or to observe an ongoing segmentation. In addition, naive volume rendering cannot visualize extremely large EM volumes at the rates necessary for interactive exploration.

We have chosen to explore on-demand local filtering with adaptive data structure for fast volume rendering in NeuroTrace. The main motivations are the flexibility offered by being able to change filters and filter parameters on the fly, and avoiding additional disk storage and bandwidth bottlenecks for terabyte-sized volumes of data. Therefore, we perform filtering only on parts (blocks) of the volume that are visible from the current viewpoint, and store the computed data directly on the GPU for later reuse. We have implemented dynamic caching for these pre-computed blocks on the GPU to avoid costly transfers to and from GPU memory while at the same time avoiding repetitive recalculation of filtered blocks. During visualization either the original volume, the noise-reduced data, the computed edge values, or a combination of the above can be displayed.

Our visualization approach supports the inspection of data prior to segmentation, for identifying ROIs, as well as the visualization of the ongoing and final. To improve the pre-segmentation visualization of the raw data we have implemented on-the-fly nonlinear noise removal and edge enhancement to support the user in finding and selecting ROIs. We have implemented 2D and 3D Gaussian and mean filters, as well as a nonlinear median filter with user-adjustable neighborhood sizes. Nonlinear filters have shown especially good noise removal properties without degrading edges in the EM data. Using a local histogram-based edge metric, which is only calculated on demand for currently visible parts of the volume and cached for later reuse, we can enhance important structures (e.g., myelinated axons) while fading out less important regions. In our local histogram-based edge detection approach, we take a block neighborhood around each voxel to calculate the brightness gradient for different directions. We separate the voxels’ neighborhood along the given direction into two halves and calculate the histogram in each half-space. Finally, the histogram difference is calculated using the $\chi^2$ distance metric. A high difference between histograms indicates an abrupt change in brightness in the volume, i.e., an edge. The maximum difference value over all directions is saved as the edge value in the cache block. During ray-casting we use the computed edge values to modulate the current sample’s opacity with different user-selectable opacity weighting modes. To speed up the local filtering, we implemented NVIDIA’s CUDA.
programming environment. In order to visualize and inspect the segmented neural processes in 3D, we depict the original volume data together with semi-transparent isosurfaces that delineate structures such as axons or dendrites. The output image is generated in a single ray-casting pass for both the isosurfaces (surfaces of constant density) and the part of the volume that is shown using direct volume rendering.

The neural process segmentation described above outputs a set of implicit surfaces for each 2D slice. However, in order to make the system scalable for large EM data, we do not store these 2D distance fields. Instead, we convert the segmentation to a compact format by fitting an ellipse to each active ribbon. An entire structure such as an axon is then represented as a simple list of elliptical cross-sections, which reduces the memory footprint significantly. In order to visualize smooth shaded surfaces, we evaluate the distance from a set of ellipses on the fly by interpolating the closest adjacent ellipse pair for any given 3D point.

**Future work**

To improve the current segmentation and visualization system, we would like to implement a greater variety of filters and other edge-detection approaches. Also we would like to automatically adjust our pre-defined filter settings and opacity windowing function depending on the resolution of the input data. The biggest challenge in the future is the extremely large z-slice distances (thicknesses) in EM data sets. The integration of shape-based interpolation or directional coherence methods into the volume rendering might be a promising approach to this problem. We also would like to extend the current segmentation and tracking method to handle merging and branching of neural processes.

The next step of our research and development would be to extend the current segmentation and visualization system for distributed computing systems, such as GPU clusters, in order to handle extremely large EM data sets. The system will be based on a client-server architecture where the server processes large data sets in parallel for interactive performance. Developing parallel applications, such as simultaneous tracking of multiple neural processes or multiple image tiles stitching and registration on a GPU cluster system, would be another interesting future direction.

**Publications and posters**

