Eye-Trace: Segmentation of Volumetric Microscopy Images with Eyegaze

Thomas Templier, Kenan Bektas, Richard H.R. Hahnloser
Institute of Neuroinformatics
University of Zurich and ETH Zurich
Neuroscience Center Zurich, Zurich, Switzerland
{templier,rich}@ini.ethz.ch

ABSTRACT
We introduce an image annotation approach for the analysis of volumetric electron microscopic imagery of brain tissue. The core task is to identify and link tubular objects (neuronal fibers) in images taken from consecutive ultrathin sections of brain tissue. In our approach an individual 'flies' through the 3D data at a high speed and maintains eye gaze focus on a single neuronal fiber, aided by navigation with a handheld gamepad controller. The continuous foveation on a fiber of interest constitutes an intuitive means to define a trace that is seamlessly recorded with a desktop eyetracker and transformed into precise 3D coordinates of the annotated fiber (skeleton tracing). In a participant experiment we validate the approach by demonstrating a tracing accuracy of about the respective radiuses of the traced fibers with browsing speeds of up to 40 brain sections per second.

Author Keywords
Connectomics; brain mapping; array tomography; neural circuit reconstruction; segmentation; annotation; user interface design; eye tracking.

ACM Classification Keywords
H.5.2. User Interfaces: Input devices and strategies interfaces and presentation.

INTRODUCTION

Connectomics
Cellular connectomics (watch a video introduction from pioneer Jeff Lichtman [29]), a field of neuroscience that aims to decipher the organization of brain neural networks, is facing a data analysis challenge. Dozens of terabytes of high resolution volumetric electron microscopy (EM) images of brain tissue need to be analyzed in order to detect neuronal fibers and their contacts with other fibers across thousands of consecutive images (neuron tracing). Successful tracing of these fibers provides the neuron morphologies and is a prerequisite for revealing the wiring diagram of the neural network (illustrated in a short animation from the online game Eyewire [12]). Such neural circuit reconstruction is an impressively challenging and yet unsolved computer vision problem which has resisted a decade of sustained efforts in developing automated algorithms [17, 23–25, 33, 44]. All recent discoveries in the field [5, 6, 8, 22, 26, 27, 43] have extensively relied on large numbers of trained humans who annotated manually or semi-automatically these large data sets. For instance the study by Takemura et al. [43] necessitated about 15,000 person-hours of manual annotation and proofreading. In this project we explored a multimodal interaction (eyetracker and gamepad controller) approach for cellular connectomics with the aim to increase the tracing throughput of single individuals. We do so by letting users navigate through the data at high speed while continuously recording their decisions at the first easily measurable source: the eye gaze.

CHI'16, May 07-12, 2016, San Jose, CA, USA.
 ISBN 978-1-4503-3362-7/16/05…$15.00.
DOI: http://dx.doi.org/10.1145/2858036.2858578

RELATED WORK

Current Annotation Methods
Neuronal fibers are tubular objects that appear as roundish structures in high-resolution EM images taken from consecutive ultrathin sections (of 5 to 50 nm thickness) cut from a small piece of animal brain tissue (Figure 1).

Typically, skeletonization of neuronal fibers is used in purely manual methods [21, 36] and in semi-automated methods [4]. It consists of identifying fiber cross-sections with mouse clicks near their center of mass in every or in every few consecutive sections (Figure 1b). Each click can be decomposed into a) locating the fiber, b) locating the mouse cursor, c) positioning the cursor and d) clicking on the fiber. According to Fitts´s Law [14], the time required to finish the mouse based annotation operation (steps b and c) is proportional to the logarithm of the ratio of target distance to target size. Put simply, more time is required for small targets which are further away.

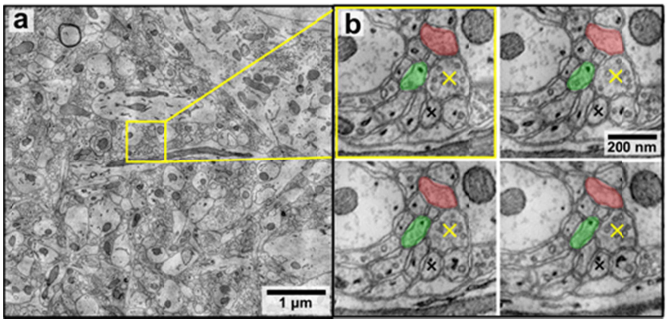

Figure 1. Electron micrographs of brain tissue. a: micrograph of a part of a section cut from a piece of mouse brain tissue. b: zoomed inset across 4 consecutive sections. Two labeled fibers are shown in red and green. The yellow and black crosses represent centers of mass of two separate fibers (skeletonization).

Also, in semi-automated methods aided by machine learning [19, 26, 27, 42], a candidate annotation that fills the inner part of each putative fiber is produced by means of computer vision algorithms (such as the green and red labels in Figure 1b). These annotations are then overlaid on the data and finally are proofread by humans.

To be able to deal with increasingly large volumes, such annotation pipelines need to be scaled up accordingly, which can be achieved by a) a better automated assessment of the locations that require manual intervention to efficiently guide proofreading, b) increasing the tracing throughput of single individuals and c) making the manual tasks easy to learn to minimize overhead from training new annotators. Both proofreading and skeletonization tasks suffer from the lack of intrinsic incentives, implying that human annotators typically must get paid or the task must be gamified [27].

Eyegaze as Input

Eye tracking consists of measuring the geometrical orientation of the eyegaze. When an individual is looking at a screen, the gaze orientation can be converted into the coordinates of the point on screen where the user is looking at.

Recently, various applications using the eye gaze as input have emerged, such as: selection of large menu items in desktop and mobile applications [28], scrolling, typing, emotion monitoring, gaze contingent visualization [3] and immersive interaction in online virtual reality environments and games. Most of these applications require a certain level of concentration known as overt attention, which occurs when an individual has the sensation of mentally focusing on a point and the measured eyegaze orientation points precisely at that same point [18].

Considering the fundamental limitations of using gaze as an input channel, Zhai et al. [45] combined gaze with mouse input by warping the pointer position to the gaze position. Their experimental study demonstrated that this system, taking advantage of Fitt´s law [14], can reduce manual effort. In an experimental study among different age groups, Murata [31] has compared the target pointing performance of participants as they use either mouse or gaze as the input modality. The results showed that with gaze input the pointing times of participants were up to 50% shorter than with mouse only. Especially for older participants with declined motor skills the target pointing with gaze was easier and simpler. Other studies with comparable setups showed that pointing with gaze is as accurate as conventional mouse pointing, and exhibits significantly shorter dwell times [13, 30, 39]. A user-centered design approach and a formative user study [41] integrated gaze as an input modality to a multi-modal environment. The findings indicate that gaze input is an intuitive modality for tasks that are in pointing nature.

To our knowledge, there is only one report of the use of eyegaze input for image analysis tasks [37]: there, a user interactively segments an object in a static (typically radiologic) image. This segmentation is achieved by feeding the eyegaze inputs to a segmentation algorithm that updates the predicted segmentation of the object of interest.

EYEGAZE FOR NEURON TRACING

In our approach, microscopy images from consecutive sections of brain tissue are displayed on a screen one after the other at a high frame rate (10 to 40 sections per second). As the sections are only a few nanometers thick, flying through them produces a pleasant impression of smooth navigation because the visible cellular structures (membranes, mitochondria, microtubuli, vesicles, myelin) evolve slowly. An observer can then perform a smooth eye pursuit and follow seamlessly these well delineated objects that they keep in their attentional spotlight. As skeletonization of neuronal tubular fibers requires a single localized label every section or every few sections, we realized that an eyegaze readout would be well suited as input to a neuron tracing system. We therefore let humans browse through the consecutive images at a constant speed and record their eyegaze with an eye tracker, infer the location on screen at which the user is foveating, and use that location as the answer to the question “Where is the cross section of the fiber currently located?”. To our knowledge, this is the first attempt of performing frame-based image annotation using eyegaze input.

EXPERIMENT

After extensive tests, a short pilot experiment with 6 participants confirmed our impression that the eyegaze tracing approach was feasible and accurate enough even by novices who are unfamiliar with the data. We therefore refined the design of the pilot experiment and conducted a within subject design experiment with the goal of answering the following main question:

Q1. What is the accuracy of tracing neuronal fibers as a function of browsing speed and fiber size?

Additionally we sought answers to the following annex questions:

q2. Can neuron tracing tasks be performed by novice untrained individuals?

q3. Does the performance improve when a user repetitively traces the same fiber?

q4. Can users maintain high accuracy while tracing long segments of fibers (1200 sections)?

Finally and more generally we also sought to know whether eyegaze based neuron tracing would potentially be accurate and reliable enough to feed data into semi-automated machine vision pipelines.

Setup

The experimental setup is shown in Figure 2. Eyegaze is measured with a desktop eyetracker (Eyetribe, 30Hz sampling rate, 0.5-1° accuracy). The navigation is controlled with a gamepad (Logitech F310) with the following functionalities: Left joystick: panning in the x-y plane; Button 1: start of browsing; Button 2: task termination. A primary display (17 inch LCD, 1280x1024, 75 Hz), presents images to the user while a secondary one, invisible to the participant, is used by the experimenter to monitor the procedure (Figure 2). The head of the user is kept with a chinrest at about 60 cm distance from the display, leading to an accuracy of the eyetracker of 40 screen-pixels. An orchestrator script (Matlab, MIJI) loads tasks consecutively by calling a custom Fiji [38] plugin for display and navigation of the image stacks. Custom scripts (python) monitored the communication with the eye tracker and the gamepad. The browsing speed reported in frames per second (fps) indicates the number of consecutive sections displayed per second.

Neuroanatomical Data

The data was generated from consecutive scanning electron micrographs of a piece of mouse neocortex brain tissue with a voxel size of 3 nm x 3 nm x 29 nm [26] downloaded from the openconnectome.org website [9] (Figure 1). Images were displayed in a Fiji canvas with a fixed magnification factor appropriately set for each task, depending on the initial size of the fiber to be traced. The magnification varied between 200% for large fibers up to 500% for small fibers. At 500% magnification, 8 image pixels correspond to 1 screen pixel. Six image stacks were extracted from the large online dataset: five stacks were of dimensions 2000 x 2000 x 300 voxels (6 µm x 6µm x 8.8 µm) and one stack, containing long fibers, was of dimensions 2000 x 2000 x 1200 voxels (6 µm x 6 µm x 35.3µm).

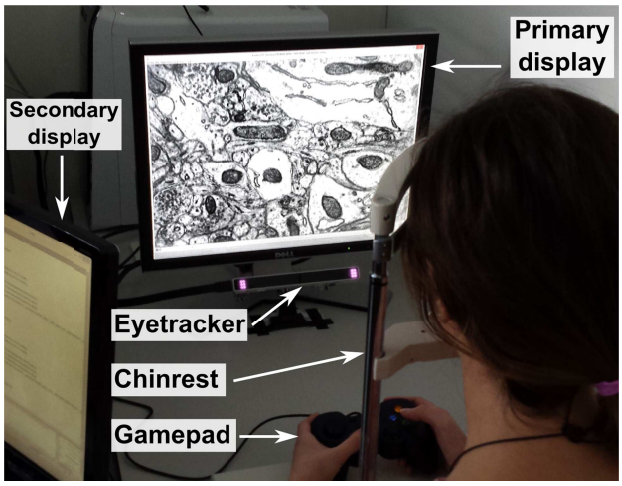

Figure 2. Experimental setup.

From these image stacks we chose 18 unique fibers to be used within the study tasks, described below. The chosen fibers did not terminate before reaching the boundary of the stack volumes. To avoid a possible bias towards more easily traceable fibers that run along the browsing axis (i.e., perpendicular to the tissue sectioning plane), we took care of selecting fibers that exhibited sharp turns.

In order to reveal the relationship between tracing speed and tracing accuracy, we classified fibers based on their diameter size. Small fibers have an average diameter of about 45 pixels (135 nm), medium fibers of 45 to 65 pixels, and large fibers of more than 65 pixels.

Participants

20 students (10 female, MSc. and PhD levels, and ages 23 to 33 years, normal or corrected-to-normal vision) participated in the experiment, provided their written consent and received a small compensation.

Procedure

Guided by the experimenter (2nd author), each participant underwent the following procedure: 1) welcoming and giving consent; 2) initial training (eye tracker calibration and 4 training tasks); 3) experiment: consecutive tracing tasks interleaved with calibrations, breaks and difficulty ratings; 4) system usability scale (SUS) rating; and 5) short debriefing.

Tasks

The experiment was divided into 129 tasks. We analyzed only the first 62 tasks, the remaining ones were designed to test further functionalities that are not reported here (branching points, complex terminations). When a task started, the initial image of a stack of sections was displayed on screen with a red cross placed in the center of the fiber that the participant had to trace. When the participant was ready, she pressed a button on the gamepad triggering the start of the browsing through the stack at a given frame rate that depends on the task. During the browsing, the participant followed the fiber of interest with

her eyes and adjusted the field of view in the x-y plane with the joystick if necessary.

The task could terminate in two ways: 1) the participant followed the fiber until the end of the stack was reached (typically after 300 sections, that is, after 15 seconds if browsing was done at 20 fps); or 2) the participant pressed the termination button before the end of the stack was reached to signify that she was not able to follow the fiber any more for whatever reason.

After the termination of a task, the participant had to briefly rate its difficulty on a Likert scale. Thereafter either the next task started five seconds later or a short thirty seconds break was given.

Experimental Design

The participants were exposed to a succession of tasks that were interleaved systematically by breaks, calibrations, or instructions.

Warmup

The first four tasks provided a short initial training and exhibited four different fibers at increasing browsing speeds of 10, 20, 30 and 40 fps. These fibers were not reused in other tasks.

Speed/Size Experiment

Tasks 5-52 constituted the **speed/size experiment** and showed small, medium and large fibers at 10, 20, 30 and 40 fps. Each of the 12 tracing conditions (4 speeds x 3 sizes) was tested 4 times, for a total of 48 tasks.

For these 48 tasks, we used a total of 12 unique fibers: 4 small, 4 medium and 4 large. All 12 fibers were 300 sections long. To prevent a task from displaying the exact same fiber as a previous task, we created four different tasks from each unique fiber by displaying the stack 1) from the first to the last section (forwards); 2) from the last to the first section (backwards); 3) forwards with a 180° rotation around the browsing axis (forward rotated); 4) backwards rotated. The 4 tasks created from a unique fiber were used across the 4 different speeds. For example, fiber number 7 has been traced in the following conditions: (10fps, backwards), (20 fps, forwards rotated), (30 fps, forwards) and (40 fps, backwards rotated).

The 48 tasks were first randomized, then we reordered two tasks to number 5 and 10 for the task repetition experiment described below. This order was the same for all participants.

Repetition Experiments

The two groups of tasks {53,54,55,56} and {58,59,60,61} constituted the **fiber repetition experiment** and displayed 4 times consecutively at 20 fps the same medium and small fiber, respectively.

Task 53 (small fiber, 20 fps) was equivalent to task 5, and task 58 (medium fiber, 20 fps) was equivalent to task 10. The two pairs of identical tasks {5,53} and {10,58} constituted the **task repetition experiment** to assess improvement of the participants between the start and the end of the approximately 1 hour long experiment.

Long Fiber Experiment

Task 57 and 62, that we named **long fiber experiment,** displayed at 20 fps a medium and a small fiber, respectively, that were 1200 sections long.

Calibration

A nine-point eye tracking calibration was performed before the initial training tasks, and before tasks 1, 25 and 49. Calibrations were repeated until an accuracy measure given by the Eyetribe server (v.0.9.49) smaller than 0.4 was achieved. This calibration criterion is slightly more conservative than the "Excellent calibration" criterion implemented in the Eyetribe user interface.

Measures

Completeness

We assessed **completeness** of a tracing task by dividing the number of correctly foveated sections by the total number of sections spanned by the fiber. A section is considered correctly annotated if the participant has not pressed the termination button beforehand. For example if the participant presses the button during section 200 in a task displaying a 300 sections long fiber, then the task completeness is 0.66 (200/300).

Tracing accuracy

Each fiber was skeleton traced by an experienced neuroscientist (1st author) by clicking in each section k $(k = 1, \dots, K)$ on the apparent center f_k of the fiber (in image pixel coordinates). The trace $(f_1, \dots, f_K)$ was defined as the ground truth trace of the fiber. The tracing **accuracy** of a task, expressed in pixels, is defined as the root-mean-square-error (RMSE) $\frac{1}{n}\sqrt{\sum_{k=1}^{n}(f_k - e_k)^2}$ of the gaze positions e_k across all sections, where e_k is the raw unfiltered gaze coordinate fetched from the eye tracker while section k is displayed, and n is the number of validly traced sections (validly traced sections are sections during which the participant was reportedly tracing, that is, he had not pressed the termination button). We computed a **collaborative trace** of a given fiber by averaging gaze coordinates over validly traced portions of fibers across participants. Intuitively, the collaborative trace is the mean trace, which represents the consensus from the crowd. The **collaborative accuracy** is the accuracy of the collaborative trace. We computed a **smoothed bias-corrected trace** by first subtracting from e_k the bias $b = \frac{1}{n}\sum_{k=1}^{n}(e_k - f_k)$ followed by applying two median filters separately on the x and y coordinates, where the respective medians are computed over 10 consecutive sections. The resulting accuracy of the smoothed bias-corrected trace was $\frac{1}{n}\sqrt{\sum_{k=1}^{n}(f_k - M(e_k - b))^2}$, where M represents the two median filters. In our hands, our desktop eye tracker was

sufficiently precise but sometimes introduced some offset. To compensate for such accuracy offsets, we made the bias correction. The median smoothing was computed to improve the precision.

For 40 fps tracing, given that eyegaze sampling is only done at 30 Hz, there was no foveation point measurement for one fourth of the sections. We chose to not interpolate and to discard these sections for RMSE calculation.

Subjective Difficulty Rating

The fiber difficulty was measured with a Single Ease Question (SEQ). After each task, participants were asked to immediately rate the difficulty they encountered in tracing that particular fiber on a Likert scale from 1 (very easy) to 5 (very difficult).

System Usability

The system usability was assessed with a standard system usability scale (SUS) [7] immediately after the end of the experiment.

RESULTS

Speed/Size Experiment

The heat maps in Figures 3 and 4 depict in detail the accuracy and completeness of all participants for the 48 tasks of the speed/size experiment. As can be seen in these heatmaps, the accuracy values roughly oscillate between 20 and 50 pixels and the task completeness is most often 100%. These data illustrate the validation of the eyegaze tracing approach by showing that the error is about the same as the extent of the fibers while the participants rarely lost track during tracing, except for small fibers at speeds 20 fps and higher.

The accuracy tended to decrease with higher flying speeds, as intuitively expected. Somewhat counter-intuitively at low speeds (10 and 20 fps, first two meta-columns on the left, Figure 3) the accuracy decreases with increasing fiber sizes. We attribute this decrease to our definition of accuracy rather than to an inherent difficulty of tracing larger fibers (see discussion).

For a convenient representation, we depict in Figure 5 some accuracy values on representative fibers of our size classes.

Main effects of speed and size

We run a 2-way factorial repeated measures ANOVA (f-rmAN) for 3 size levels and 4 speed levels. The assumption of sphericity was violated (Mauchly´s test) for the main effects of the size, ($\chi2(2) = 6.5$, $p = .04$), for the speed, ($\chi2(5) = 15.1$, $p = .01$), and for the interaction speed/size, ($\chi2(20) = 66.3$, $p = .000$). The Greenhouse-Geisser estimates of corrected degrees of freedom became .76, .66 and .53 respectively.

There was a significant main effect of the size on the accuracy, ($F(1.5, 27.3) = 85.1$, $p = .000$, $\eta2 = .825$). Contrasts with small fibers revealed that the average accuracy for large fibers ($F(1,18) = 112.8$, $r = .37$, $\eta2 = .862$) and medium fibers ($F(1,18) = 21.6$, $r = .67$, $\eta2 = .545$) were larger than that of small fibers.

There was a significant main effect of the speed on the average accuracy, ($F(1.98, 35.6) = 3.49$, $p = .042$, $\eta2 = .162$). There was a significant interaction effect between the size and speed of tracing, ($F(3.2, 57.2) = 4.98$, $p = .003$, $\eta2 = .217$). The size of the fiber to be traced had therefore a different effect on tracing accuracy depending on the speed.

Interplay of speed and size

In order to characterize the effect of the speed for each different size level we run three separate 1-way repeated measures ANOVA (rmAN) with Bonferroni corrections. Figure 6 shows the accuracy values for the 12 speed/size conditions tested.

The tracing accuracy for **large fibers** (yellow bars, Figure 6) was independent of the speed and its mean was 50 ± 5 pixels ($F(3, 54) = 3358$, $P = 0.25$, $\eta2 = 0.157$), no significant difference between pairs (post hoc comparisons, $p > 0.05$). Because large fibers have a diameter larger than 65 pixels, the gaze input therefore sometimes was located outside of the fiber boundary.

The accuracy for **medium fibers** (red bars, Figure 6) depended on tracing speeds ($F(3, 54) = 53.55$, $p < 0.0005$, $\eta2 = 0.933$) and the average accuracy decreased linearly with speed from 34 ± 8 pixels for 10 fps to 48 ± 7 pixels for 40 fps, as intuitively expected. Post hoc tests (with Bonferroni correction) showed a statistically significant accuracy decrease ($p = .000$) for the speed pairs (10, 30) and (10, 40).

The accuracy for **small fibers** (blue bars, Figure 6) also decreased with increasing speeds ($F(3, 54) = 17.207$, $p < 0.0005$, $\eta2 = 0.489$) from 27 ± 15 pixels for 10 fps to 62 ± 28 pixels for 40 fps. Post hoc tests (with Bonferroni correction) showed a statistically significant accuracy decrease ($p = .000$) between the pairs (10, 30), (10, 40) and (20, 40) fps.

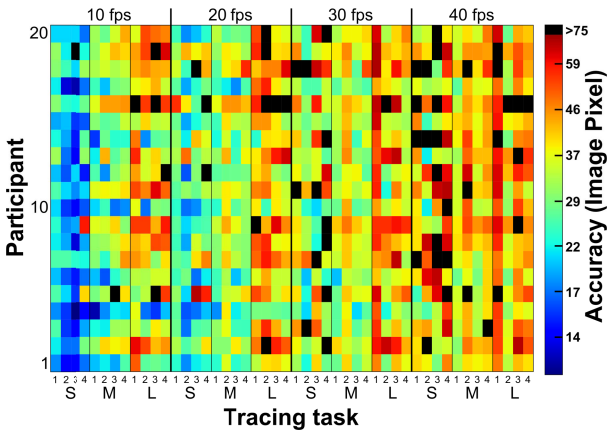

Figure 3. Heat map of the accuracy for each task from the speed/size experiment for all participants. S: small, M: medium, L: large

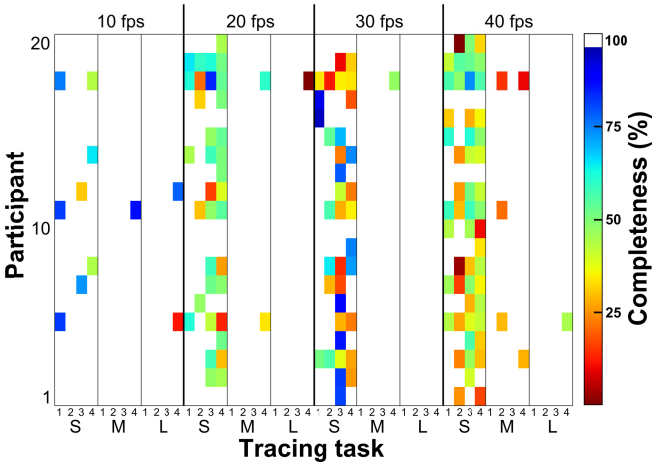

Figure 4. Heat map of the completeness for each task from the speed/size experiment for all participants. S: small, M: medium, L: large

In summary rmAN results are supplementing the general findings we obtained from f-rmAN: rmAN reveals the effect of tracing speed for each fiber size individually and f-rmAN shows that the effect of speed differs depending on size.

Collaborative traces

The accuracies of the collaborative traces for all tasks of the speed/size experiment are shown in Figure 7. In all tasks of the speed/size experiment, the collaborative trace (Figure 7) has higher accuracy than the averaged accuracies of the traces taken individually (Figure 6). This improvement in tracing accuracy is shown in detail in Figure 8 for each of the 48 tasks of the speed/size experiment.

Smoothed bias-corrected traces

The accuracies of the smoothed bias-corrected traces are shown in Figure 9. The bias and smoothing corrections show the greatest effect at low speeds.

Fiber Repetition Experiment

In the repetition experiment, a medium and a small fiber were repetitively traced consecutively 4 times at 20 fps. Some participants noticed the repetition. The average accuracy for the medium fiber was not affected by the repetition (Figure 10 left, red bars, 1-way rmAN, $F(3, 54) = 1.233$, $p = 0.307$). The same was true for the small fiber, (Figure 10 left, blue bars, 1-way rmAN, $F(3, 54) = 1.415$, $p = 0.25$).

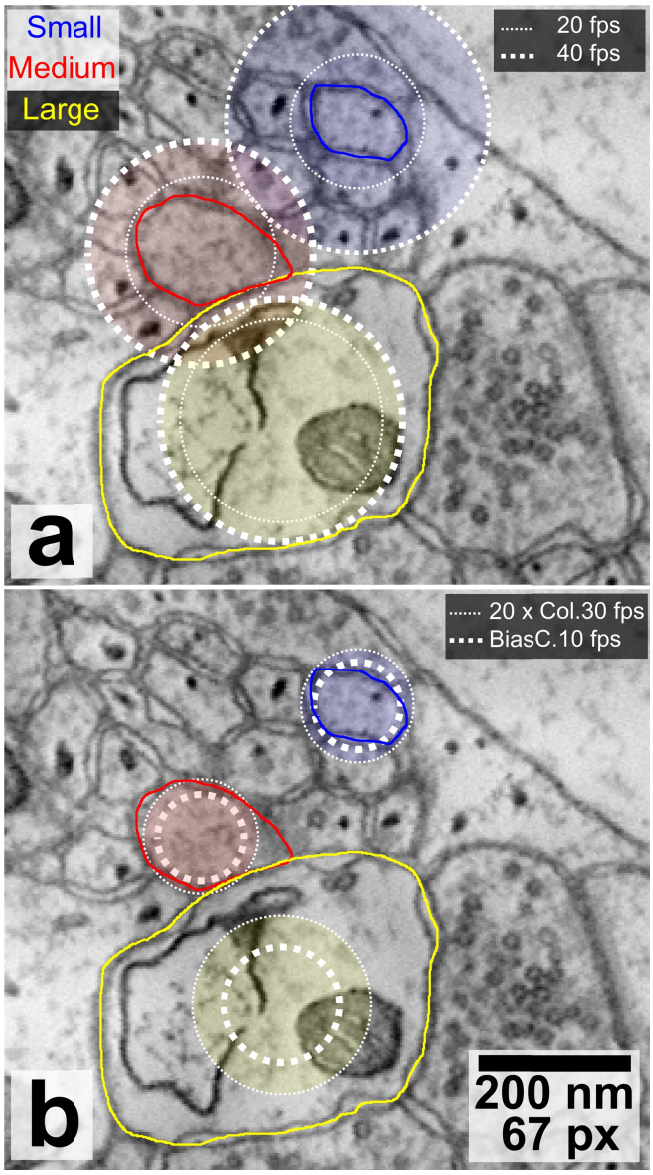

Figure 5. Accuracy of eyegaze tracing as a function of fiber size and browsing speed. Discs with a radius equal to the accuracy are centered on an example small (blue), medium (red), and large fiber (blue). Fiber boundaries are highlighted in corresponding colors. Px:image pixels (3 nm/pixel). a: fine and coarse dotted circles depict the average accuracy of all participants at 20 and 40 fps, respectively; b: fine dotted circles depict the collaborative RMSE at 30 fps. Coarse dotted circles depict the average accuracy of the smoothed bias corrected traces at 10 fps.

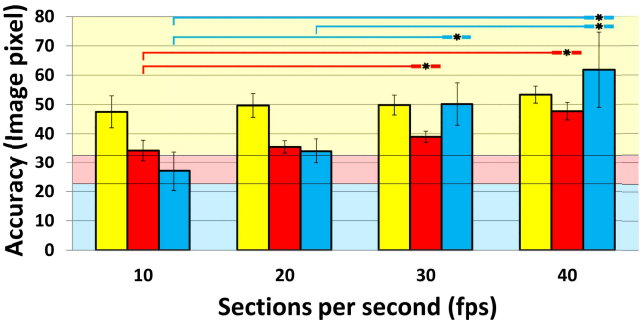

Figure 6. Accuracies for all browsing speeds (x-axis) and fiber sizes (color code as in Figure 3: Yellow: large, red: medium, blue: small.). Error bars show two standard errors of the mean. The colored background indicates the three size classes with the same color code. Stars indicate significance (* = p<0.05, post-hoc Bonferroni).

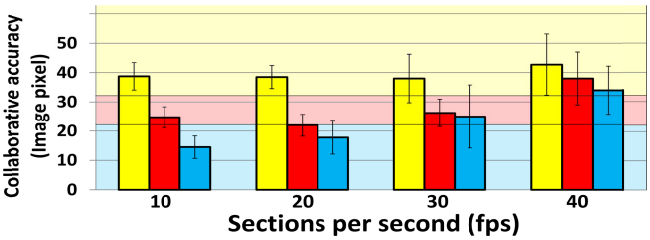

Figure 7. Accuracies of the collaborative trace for all browsing speeds (x-axis) and fiber sizes. Error bars and color codes as in Figure 6.

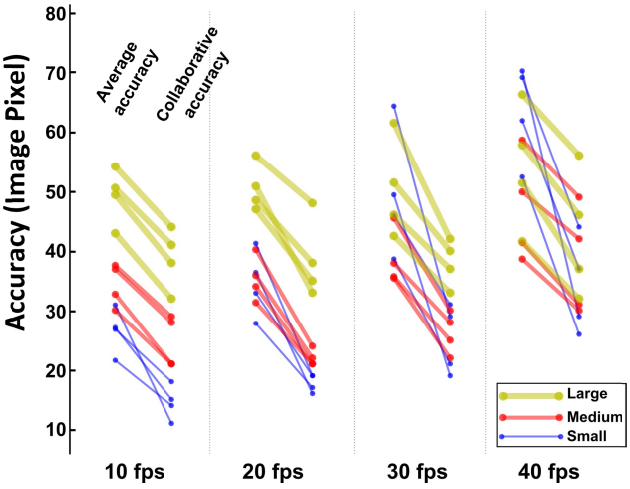

Figure 8. Collaborative tracing. Comparison of the average across all participants of the accuracies of a task with the accuracy of the corresponding collaborative trace.

Task Repetition Experiment

To assess learning of the eyegaze tracing task, participants repeated two tasks (one small, one medium fiber), with the first instance appearing within the first 10 tasks and the second instance appearing more than 40 tasks later. Figure 10 right shows the accuracy for the two repeated pairs of tasks. A pairwise t-test showed a significant increase in accuracy only for the medium fiber ($t(18) = 4.77$, $p < 0.0005$) from 40 ± 6 to 33 ± 6 pixels, and no significant difference for the small fiber ($t(18) = 1.52$, $p = 0.146$). The relatively small differences observed (though statistically significant for the medium fiber) combined with the high accuracies values show that users reach their good performance almost immediately. No participant reportedly noticed that the same fiber was being presented again after 40 tasks.

The task completeness for the repetitively traced medium fiber did not change with repetition and was on average 0.98. (1-way rmAN with Greenhouse-Geisser correction (violated sphericity), $F(1.006, 19.123) = 0.991$, $P = 0.33$). Hence fibers were already completely traced at the first repetition. For the group of small fibers, there was a slight increase in the task completeness after repetition (1st trial: 0.89 ± 0.22 and 4th trial: 0.98 ±0.04) but it was not a significant effect ($F(1.802, 34.244) = 1.264$, $P = 0.293$).

Long Fiber Experiment

The RMSE of the long medium (task 57) and long small (task 62) fibers were 23 ± 3 and 24 ± 3 pixels, respectively (2 traces were excluded as the communication with the eye tracker broke during task 62). These tracing accuracies are comparable with the ones for the short fibers (no tests performed).

Difficulty and Usability Ratings

Figure 11 shows a summary of the difficulty ratings. The subjective difficulty depended on tracing speed (non-parametric Friedman test, $\chi^2(3) = 8.647$, $p = .034$, Figure 11 left) and there was an increase in difficulty as a function of speed, as intuitively expected (though not significant).

The subjective difficulty also significantly depended on fiber size (non-parametric Friedman test , $\chi^2(2) = 24.5$, $p < 0.005$, Figure 11 middle). Small fibers were clearly perceived as the most difficult ones to trace with a statistically significant increase in perceived difficulty between large and small fibers, ($Z = -3.52$, $p < 0.005$) as well as between medium and small fibers, ($Z = -3.52$, $p < 0.005$).

Did the perceived difficulty change when a same tracing task was repeated multiple times? Figure 11, right, shows a slight decrease in difficulty after repetition, however a statistically significant reduction was only found from the first to the fourth presentations for the small fiber ($Z = -3.27$, $p = 0.001$) and no statistically significant changes were found for the repeated medium fiber.

Usability Assessment

The average SUS score of our setup was 68.1, *SD*: 13.2. Based on the literature [2] our score is in the marginal-high part (62.5 to 70 on a scale from 0 to 100) of the *Acceptability Range* and is classified as "good" (between *"ok"* and *"excellent"*) in terms of *Adjective Ratings*.

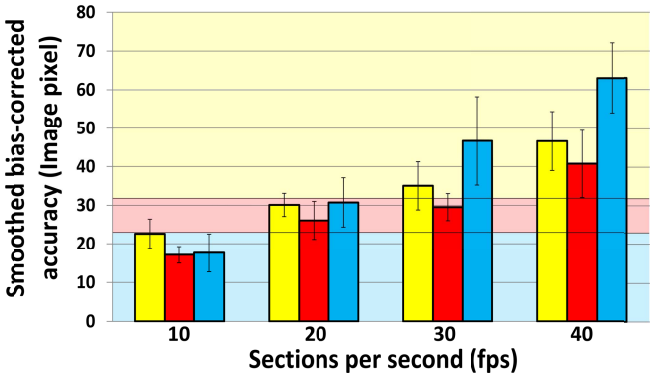

Figure 9. Accuracies of all smoothed bias-corrected traces. Error bars and color codes as in Figure 6.

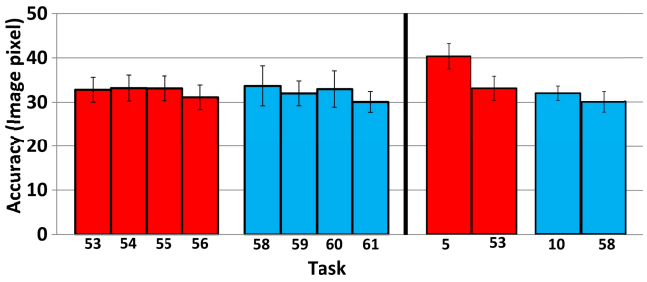

Figure 10. Learning effects with eyegaze tracing. Left: A medium (red) and a small (blue) fibers have been traced four times in a row by each participant; there are no accuracy benefits due to repetition; Right: A same task with a medium (red) and a small (blue) fiber is performed at the beginning of the experiment and more than 40 tasks later.

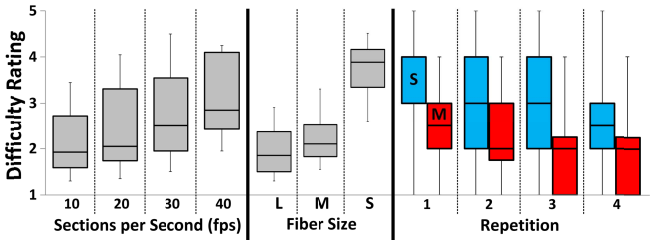

Figure 11. Subjective difficulty ratings (1: very easy, 5: very difficult) as a function of speed, of fiber size, and of repetitions in the fiber repetition experiment.

DISCUSSION

Validity of the Approach

The general impression given by the results of the participant experiment is that our eyegaze tracing approach provides acceptable skeleton traces of neuronal fibers. We found that users can follow fibers (high completeness) at high speed (10 to 40 fps) with reasonable accuracy (accuracy ~ fiber size) and that we see no severe bottleneck in performance that could arise from the technology platform chosen.

Q1. Accuracy

From Figures 3 and 6 we see that for all fiber sizes the gaze remains within the vicinity of the targeted fiber at speeds 10 and 20 fps. The study from Berning et al. [4] reports a state of the art tracing speed of 7.2h/mm, i.e., 1.3 sections/second, using accurate mouse inputs. By comparison, our system is less accurate, but allows for 10 to 40 fps browsing, which is about 15 times faster. To what extent can these numbers be compared? Clearly, the classic mouse-based approach is more accurate because users always click inside the boundary of the fiber of interest. Also, the speed comparison should be interpreted with caution because branching and termination points were not addressed here whereas these are taken into account in the speed reported for the mouse approach. In this sense it might not make sense to compare these tracing speeds.

In our approach we focus on the fact that humans are still unsurpassed in their ability to track intermingled objects and we increase the speed at the expense of the pointer accuracy. A non-automated segmentation system relying exclusively on eye gaze input would probably suffer from the lack of accuracy observed. Therefore, we believe that in future work gaze input can help automated methods to select fiber paths among diverse candidates. For example, if a fiber is erroneously traced as turning to the left by an automated method and that a user is following the fiber to the right, then the rough skeleton of the fiber would be very valuable input to correct the automated segmentation. In this sense, we hope that the achieved accuracy will be acceptable. We therefore do not see this study as a solution ready to be deployed but instead as a proof-of-principle for a scalable system that could eventually be fed into automated machine vision pipelines.

At 10 fps, smoothing and bias-correcting a trace leads to improved tracing accuracy, successfully compensating for jitter and offset produced by the eye tracker. As the ground truth and hence the bias are not known in a real application setting, querying the annotator about the exact location of his eyegaze could be a solution. For example, a salient point such as a bright item could be presented in a way to trigger the foveation of the user for a short moment on that known location, thus inferring the bias. Such a procedure could be seamlessly integrated into the workflow of a game. We note also that the quality of eye trackers is likely to improve quickly in the next few years.

As shown in Figures 3 and 5, at first glance it might be counter-intuitive that the accuracy decreases when the fiber size increases, that is, when the difficulty decreases. Our interpretation is that the gaze actually remains close to the boundaries of the fiber rather than close to the center. Instead of quantifying tracing accuracy as the simple RMSE to the approximate centers of mass, we could have added a penalty cost accounting for when the gaze is located outside of the boundary of the fiber of interest. We believe that this unexpected effect would then have vanished. However we chose to not perform the time-consuming manual contour tracing of all fibers necessary to compute such a refined accuracy.

Finally, rather specific to our neuronal reconstruction goal, it was recently demonstrated by Pallotto et al. [32] that a non-conventional brain sample preparation preserving the extracellular space between fibers aided automated machine vision algorithms. Neighboring fibers come indeed into contact much more rarely and are separated by large white gaps. We imagine that this sample preparation could compensate for the relatively lower accuracy of the eyegaze compared to a precise mouse approach.

q2. Crowd Accessibility

Scaling up the annotation task can not only be achieved by increasing the throughput of single individuals, but also by making the task accessible to a large number of individuals. Is our approach accessible to reach a large crowd? As shown in Figure 10, participants reached a high level of accuracy within a few minutes and they did not significantly improve after performing 40 tasks. Also, the heat maps in Figures 3 and 4 show qualitatively that all participants except one performed roughly equally well. From the questionnaires we learned that 14 out of the 20 participants had never manipulated such microscopy data. This therefore lets us encouragingly conclude that our approach is broadly feasible.

q3. Learning a fiber

At 20 fps, there was no significant improvement in the accuracy of participants annotating repetitively four times the same fiber, showing that the accuracy is already close to its maximum at the first attempt. This confirms also that the eyegaze approach relies on a robust mechanism of human vision to identify moving targets. A minor positive effect was observed for the completeness of these repetitively traced fibers showing that users were less likely to lose track of the fiber as they were tracing the fiber again.

q4. Long fiber segments

Our results show that users can continuously maintain attention with an appropriate foveation for at least one minute without problem (1200 sections at 20 fps). The accuracy of the long medium fiber (Task 57) was even slightly higher than the average of the short medium fibers tested in the speed/size experiment (although this long fiber might have inherently been easier to trace compared to the set of medium fibers from the speed/size experiment). We believe that this can be valuable information for the implementation of our approach for large datasets, namely to calibrate the size of the tasks presented to users. Nevertheless, such long fibers are not likely to be presented in a proofreading context as most computer errors are local and therefore users are more likely to be asked to trace short segments for local disambiguation.

Choice of the fibers

Finding a large number of fibers that were contained in a small stack from the first to the last section revealed to be a difficult task, therefore we chose to select a few (12 unique fibers for the size/speed experiment) and to apply the backward and/or the 180° rotation transformations. We were concerned that fibers could be learned, which would have biased our speed/size analysis. However we observed that the fibers displayed at such a high speed in such different settings cannot be learned. The repetition experiments tend to show that there is no learning effect even when the same fiber is repeated 4 times in a row. Also the four tasks based on a same fiber were conducted at four different speeds and therefore helped to reduce bias between different speed conditions that could have arisen from diverse difficulty levels (we assume that a fiber exhibits the same difficulty when displayed in the 4 different settings).

Another limitation maybe arose from the choice of fibers that were present from the first to the last section of the 2000x2000x300 voxels stacks which represent volumes of 6 µm x 6 µm x 9 µm. The fibers might therefore have exhibited a bias towards evolving in the direction perpendicular to the imaging plane, thus appearing more like roundish structures and maybe easing the tracing. Nevertheless, though a subjective precaution, among the candidate fibers we found, we chose on purpose some that looked tortuous to us to compensate for this possible bias. Nonetheless, our results should be considered with some degree of caution, given the limited number of fibers which have been tested.

Navigation

Magnification

In a preliminary pilot experiment, we observed that accuracy tends to increase at high magnification factors. Therefore in a different user interface version we implemented an interactive zoom to let the user adapt the display to the size of the traced fiber. However we thought that this interactive zoom would have made the task too difficult to learn as the participants were present for only about an hour.

Browsing speed

For convenience we decided to simply let the participants browse at a predefined frame rate to focus on the speed and the size of the fibers for the first characterization of our eyegaze approach. In a real setting, these speed parameters could be made easily adjustable, namely accelerating and breaking as in a racing game could also be used to navigate through the data conveniently.

Eyegaze panning

Panning laterally in the x-y plane was performed with a joystick actuated with the left thumb. It could also be done with gaze input only with appropriate treatment of the gaze signal, so that whenever the fiber deviates from the center of the display, the field of view is translated to bring the fiber back to the center of the display. Such an implementation would be desirable in mobile devices to free the hands of the user.

Complex controllers

Improving navigation probably calls for more complex control mechanisms to manipulate the data in a more efficient way, see the virtual reality system from Cali et al. [10] for example. Special handheld controllers, wearable controllers or gesture sensing devices might provide seamless navigation through such complex datasets. However a tradeoff arises between the efficiency and complexity of the human machine interaction and the accessibility of the setup to large crowds.

Collaborative Annotation

Figure 8 convincingly shows a significant increase in accuracy when averaging the trace of a fiber across the 20 participants of the study, even without removing outliers. This improvement means that if a certain level of accuracy is desired, for example as requirement from an automated pipeline, more annotators can be recruited to trace independently the same fibers. As eye tracking is about to reach mass markets, crowdsourcing our proposed eyegaze approach with the aid of online tools appears feasible.

Semi-Automated Eyegaze Tracing

Some semi-automated pipelines require that a proofreader focuses on a given fiber, follows it while the data are browsed for a couple of sections, and finally indicates the new location of the fiber. Using eyegaze information during this procedure might add some noise to the final location of the fiber provided by the annotator, however the complete approximate path would also be recorded, which could also be a valuable input to the algorithms. In a similar fashion, [4] has developed a semi-automated method that currently takes as inputs manually generated skeletons of fibers. The SOPNET framework for neuron circuit reconstruction [15–17] also interestingly makes use of candidate 2-dimensional segmentations of the fibers and could integrate our eyegaze data as input. It will be interesting in the future to assess the minimal accuracy required by these pipelines and whether our eyegaze traced fibers can be fed to them.

From the point of view of the management of the tasks given to annotators, short segments that need to be traced could be distributed among different users or players, as currently implemented in the Eyewire game and other applications [4, 20, 27]. A possible very large scale implementation of that approach could be a login procedure for new generation wearable glasses [35] during which the user has to foveate on a fiber across a few sections while his gaze is recorded. Such login procedure could constitute a new generation of recaptchas [1] with a currently unsolved complex computer vision problem that would immediately receive a great attention from the machine vision community, maybe leading to a resolution faster than without this publicity.

General Applications

The eyegaze tracing approach might be suitable for any visual data in which spatial or temporal trajectories have to be drawn: tracing of tubular objects (neurons, vasculature) in various microscopy modalities, tracking of objects over time (dividing cells, growing neurons, interacting animals, evolving particles in a gas, vehicles and humans in surveillance imagery). Automated tracking or guidance of fast moving or occluded military targets such as soldiers, vehicles or missiles could also be enhanced with gaze input.

We think that our application in itself could also be seen as a novel contribution to the field of human machine interaction as to our knowledge, it uniquely (apart from games) challenges individuals with complex tasks at an unprecedentedly high temporal rate, in some way similar to a trained pianist hitting more than 10 keys per second. Complex high throughput human computer interaction systems may not have appeared yet because the applications were missing.

Future Work

In other tasks of the experiment not analyzed in this article, participants were instructed to press a button whenever the currently traced fiber was splitting into two fibers or was merging with another one. In future work the collaborative detection of branches or objects of interest such as synaptic contacts between fibers might be achieved to create wiring diagrams of the neuroanatomical networks.

The current most promising way to scale up this approach to the analysis of large volumes of electron microscopy imagery of brain tissue would be to feed semi-automated pipelines [4, 17] with eyegaze-traced skeletons.

CONCLUSION

The analysis of increasingly large and complex data requires not only the improvement of automated methods such as machine learning but also new ways for individuals to visualize and manipulate them, making the link between the data and the automated algorithms.

In this paper we designed and implemented a novel image annotation method for the analysis of large volumetric imagery of brain tissue. By displaying dynamically the data, we enabled the readout of the decisions of the individuals directly at the earliest possible sense: the eyegaze. We performed a participant experiment to validate the approach and showed acceptable accuracy during fast visualization of the data. We also showed that our approach is scalable to be used by crowds of novice users, making our approach a good candidate for analyzing increasingly large datasets.

ACKNOWLEDGMENTS

We thank the openconnectome project, the Lichtman laboratory for making the data available and the participants. We thank Andrew Duchowski and Bart Knijnenburg for comments.

REFERENCES

1. Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. 2008. reCAPTCHA: human-based character recognition via Web security measures. *Science (New York, N.Y.)* 321: 1465–1468. http://doi.org/10.1126/science.1160379

2. Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3: 114–123.
3. Kenan Bektas, Arzu Cöltekin, Jens Krüger, and Andrew T Duchowski. 2015. A Testbed Combining Visual Perception Models for Geographic Gaze Contingent Displays. *https://diglib.eg.org/handle/10.2312/eurovisshort.20151127.067-071*.
4. Manuel Berning, Kevin M. Boergens, and Moritz Helmstaedter. 2015. SegEM: Efficient Image Analysis for High-Resolution Connectomics. *Neuron* 87, 6: 1193–1206. http://doi.org/10.1016/j.neuron.2015.09.003
5. Davi D Bock, Wei-Chung Allen Lee, Aaron M Kerlin, et al. 2011. Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471, 7337: 177–182. http://doi.org/10.1038/nature09802
6. Kevin L Briggman, Moritz Helmstaedter, and Winfried Denk. 2011. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* 471, 7337: 183–188. Retrieved from http://dx.doi.org/10.1038/nature09818
7. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189: 194.
8. Daniel J Bumbarger, Metta Riebesell, Christian Rödelsperger, and Ralf J Sommer. 2013. System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell* 152, 1-2: 109–19. http://doi.org/10.1016/j.cell.2012.12.013
9. Randal Burns, William Gray Roncal, Dean Kleissas, et al. 2013. The Open Connectome Project Data Cluster: Scalable Analysis and Vision for High-Throughput Neuroscience. *International Conference on Scientific and Statistical Database Management*: 1–11. http://doi.org/10.1145/2484838.2484870
10. Corrado Calì, Jumana Baghabra, Daniya J Boges, et al. 2015. Three-dimensional immersive virtual reality for studying cellular compartments in 3D models from EM preparations of neural tissues. *Journal of Comparative Neurology*: n/a–n/a. http://doi.org/10.1002/cne.23852
11. K R Cave and N P Bichot. 1999. Visuospatial attention: beyond a spotlight model. *Psychonomic bulletin & review* 6, 2: 204–223.
12. Eyewire. 2014. *Reconstructing a Neuron in 3D*. Retrieved from https://www.youtube.com/watch?v=noDx7TmMr8Q
13. Ribel Fares, Shaomin Fang, and Oleg Komogortsev. 2013. Can we beat the mouse with MAGIC? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*: 1387. http://doi.org/10.1145/2470654.2466183
14. P M Fitts. 1992. The information capacity of the human motor system in controlling the amplitude of movement. 1954. *Journal of experimental psychology. General* 121, 3: 262–269.
15. Jan Funke, Bjoern Andres, Fred a. Hamprecht, Albert Cardona, and Matthew Cook. 2012. Efficient automatic 3D-reconstruction of branching neurons from em data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 1004–1011. http://doi.org/10.1109/CVPR.2012.6247777
16. Jan Funke, Björn Andres, Fred Hamprecht, Albert Cardona, and Matthew Cook. 2011. Multi-Hypothesis CRF-Segmentation of Neural Tissue in Anisotropic EM Volumes. i: 1–8. Retrieved from http://arxiv.org/abs/1109.2449
17. Jan Funke, Julien N P Martel, Stephan Gerhard, and Bjoern Andres. 2014. Candidate Sampling for Neuron Reconstruction from Anisotropic Electron Microscopy Volumes. *MICCAI Supplement*, 1: 0–5. http://doi.org/10.1007/978-3-319-10404-1_3
18. Wilson S Geisler and Lawrence K Cormack. 2011. Models of overt attention. In *The Oxford handbook of eye movements*. 439–454. http://doi.org/10.1093/oxfordhb/9780199539789.013.0024
19. Daniel Haehn, Seymour Knowles-barley, Mike Roberts, et al. 2014. Design and Evaluation of Interactive Proofreading Tools for Connectomics. http://doi.org/10.1109/TVCG.2014.2346371
20. Moritz Helmstaedter, Kevin L Briggman, and Winfried Denk. 2011. High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nature neuroscience* 14, 8: 1081–1088. http://doi.org/10.1038/nn.2868
21. Moritz Helmstaedter, Kevin L Briggman, and Winfried Denk. 2011. High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nat Neurosci* 14, 8: 1081–1088. http://doi.org/10.1038/nn.2868
22. Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. 2013. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* 500, 7461: 168–74. http://doi.org/10.1038/nature12346
23. Moritz Helmstaedter and Partha P Mitra. 2012. Computational methods and challenges for large-scale circuit mapping. *Curr Opin Neurobiol* 22, 1: 162–169. http://doi.org/10.1016/j.conb.2011.11.010

24. V Jain, J F Murray, F Roth, et al. 2007. Supervised Learning of Image Restoration with Convolutional Networks. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 1–8. http://doi.org/10.1109/ICCV.2007.4408909

25. Viren Jain, H Sebastian Seung, and Srinivas C Turaga. 2010. Machines that learn to segment images: a crucial technology for connectomics. *Curr Opin Neurobiol* 20, 5: 653–666. http://doi.org/10.1016/j.conb.2010.07.004

26. Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, et al. 2015. Saturated Reconstruction of a Volume of Neocortex. *Cell* 162: 648–661. http://doi.org/10.1016/j.cell.2015.06.054

27. Jinseop S. Kim, Matthew J. Greene, Aleksandar Zlateski, et al. 2014. Space–time wiring specificity supports direction selectivity in the retina. *Nature*, 4. http://doi.org/10.1038/nature13240

28. Manu Kumar, Andreas Paepcke, and Terry Winograd. 2007. EyePoint: practical pointing and selection using gaze and keyboard. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 430. http://doi.org/10.1145/1240624.1240692

29. Jeff William Lichtman. 2014. *Brain Connectomics?* Retrieved from https://www.youtube.com/watch?v=2QVy0n_rdBI

30. Julio C Mateo. 2008. Gaze Beats Mouse : Hands-free Selection by Combining Gaze and EMG. *Chi Ea*: 3039–3044. http://doi.org/10.1145/1358628.1358804

31. Atsuo Murata. 2006. Eye-Gaze Input Versus Mouse: Cursor Control as a Function of Age. *International Journal of Human-Computer Interaction* 21, 1: 1–14. http://doi.org/10.1207/s15327590ijhc2101_1

32. Marta Pallotto, Paul V Watkins, Boma Fubara, Joshua H Singer, and Kevin L Briggman. 2015. Extracellular space preservation aids the connectomic analysis of neural circuits. *eLife*. http://doi.org/10.7554/eLife.08206

33. Stephen M Plaza, Louis K Scheffer, and Dmitri B Chklovskii. 2014. Toward large-scale connectome reconstructions. *Current opinion in neurobiology* 25C: 201–210. http://doi.org/10.1016/j.conb.2014.01.019

34. M I Posner, C R Snyder, and B J Davidson. 1980. Attention and the detection of signals. *Journal of experimental psychology* 109, 2: 160–174.

35. H S Raffle, A Wong, and R Geiss. 2012. Unlocking a screen using eye tracking information.

36. Stephan Saalfeld, Albert Cardona, Volker Hartenstein, and Pavel Tomančák. 2009. CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* 25, 15: 1984–1986. http://doi.org/10.1093/bioinformatics/btp266

37. M. Sadeghi, G. Tien, G. Hamarneh, and Ms Atkins. 2009. Hands-free interactive image segmentation using eyegaze. *Proc. of SPIE Vol* 7260: 72601H–1. http://doi.org/10.1117/12.813452

38. Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9, 7: 676–682. http://doi.org/10.1038/nmeth.2019

39. Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of eye gaze interaction. *Proc. CHI* 2, 1: 281–288. http://doi.org/10.1145/332040.332445

40. Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch : Gaze-supported Target Acquisition. 2981–2990. http://doi.org/10.1145/2207676.2208709

41. Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. 2011. Designing gaze-supported multimodal interactions for the exploration of large image collections. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications - NGCA '11*: 1–8. http://doi.org/10.1145/1983302.1983303

42. Shin-ya Takemura, Arjun Bharioke, Zhiyuan Lu, et al. 2013. A visual motion detection circuit suggested by Drosophila connectomics. *Nature* 500, 7461: 175–181. Retrieved from http://dx.doi.org/10.1038/nature12450

43. Shin-ya Takemura, Arjun Bharioke, Zhiyuan Lu, et al. 2013. A visual motion detection circuit suggested by Drosophila connectomics. *Nature* 500, 7461: 175–81. http://doi.org/10.1038/nature12450

44. Srinivas C Turaga, Joseph F Murray, Viren Jain, et al. 2010. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* 22, 2: 511–538. http://doi.org/http://dx.doi.org/10.1162/neco.2009.10-08-881

45. Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit CHI 99*: 246–253. http://doi.org/10.1145/302979.303053

46. Eyewire A Game To Map The Brain. Retrieved from https://eyewire.org