

# Image processing for precise three-dimensional registration and stitching of thick high-resolution laser-scanning microscopy image stacks

Chloé Murtin<sup>1,2,3</sup>, Carole Frindel<sup>3</sup>, David Rousseau<sup>3</sup>, Kei Ito<sup>1,2</sup>

*1 Institute of Molecular and Cellular Biosciences, The University of Tokyo, Yayoi, Bunkyo-ku, 113-0032 Tokyo, Japan.*

*2 Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwanoha, Kashiwa, 277-0882 Chiba, Japan.*

*3 CREATIS, Institut National des Sciences Appliquées de Lyon (INSA Lyon), 7 avenue J Capelle, bat. Blaise Pascal, F-69621 Villeurbanne cedex, France.*

*4 Institut für Zoologie, Universität zu Köln, Zùlpicher Str. 47b, 50674 Köln, Germany.*

---

## Abstract

The possible depth of imaging of laser-scanning microscopy is limited not only by the working distances of objective lenses but also by image degradation caused by attenuation and diffraction of light passing through the specimen. To tackle this problem, one can either flip the sample to record images from both sides of the specimen or consecutively cut off shallow parts of the sample after taking serial images of certain thickness. Multiple image substacks acquired in these ways should be combined afterwards to generate a single stack. However, subtle movements of samples during image acquisition cause mismatch not only in the translation along x-, y-, and z-axes and rotation around z-axis but also tilting around x- and y-axes, making it difficult to register the substacks precisely. In this work, we developed a novel approach called 2D-SIFT-in-3D-Space using Scale Invariant Feature Transform (SIFT) to achieve robust three-dimensional matching of image substacks. Our method registers the substacks by separately fixing translation and rotation along x-, y-, and z-axes, through extraction and matching of stable features across two-dimensional sections of the 3D stacks. To

---

\*Corresponding author

validate the quality of registration, we developed a simulator of laser-scanning microscopy images to generate a virtual stack in which noise levels and rotation angles are controlled with known parameters. We illustrate quantitatively the performance of our approach by registering an entire brain of *Drosophila melanogaster* consisting of 800 sections. Our approach is also demonstrated to be extendable to other types of data that share large dimensions and need of fine registration of multiple image substacks. This method is implemented in Java and distributed as ImageJ/Fiji plugin. The source code is available via Github (<http://www.creatis.insa-lyon.fr/site7/fr/MicroTools>).

*Keywords:* stitching; registration; SIFT; laser-scanning microscopy; brain; connectomics; *Drosophila*.

---

## 1. Introduction

Laser-scanning fluorescent microscopy is a powerful tool for analyzing three-dimensional (3D) complex structures found in life sciences such as neuronal structures, which can be visualized using fluorophore-conjugated antibody labeling or targeted-expression of fluorescent proteins [1]. Combinations of multiple fluorescent markers and excitation filters can be used to highlight various neuronal objects, e.g., axons as well as pre- and postsynaptic sites, in the same specimen [2]. Such imaging techniques promise to give access, for instance in neurology, to connectomics which aims to produce a comprehensive and systematic analysis of the connections between brain regions and between numerous neurons within them [3]. The success of such analysis depends on the capacity to acquire in great detail, i.e., at the scale of the synapsis, the entire volume of the brain specimen.

Although laser-scanning microscopy can acquire images of thin optical sections from thick tissues [4], the possible depth of imaging is limited by three factors. First, the working distance of the microscope objective limits the depth for which images can be recorded. Second, signals become darker in deeper regions of the samples, because both excitation laser beam and emitted fluo-

rescence are attenuated by the sample tissue that is optically not completely  
20 translucent [5]. This causes diminution of signal intensity and signal-to-noise  
ratio. And third, image quality degrades in deeper regions even when it is still  
within the working distance of the lens. Although new techniques such as clear-  
ing agents (e.g., CLARITY and Scale) can make sample tissues transparent  
[6, 7], light rays are deflected and scattered when they pass through tissues.  
25 Deep objects therefore appear blurry and lose contrast.

Thus, spatial resolution decreases with the depth from the sample surface.  
Even with the objectives with high numerical aperture and long working dis-  
tance (more than 200  $\mu\text{m}$ ), image quality decreases considerably when the focal  
plane becomes as deep as 100  $\mu\text{m}$  [8]. Two workaround techniques have been  
30 employed to overcome this problem. For the specimen that is thinner than 300  
 $\mu\text{m}$  along the optical axis ( $z$ -axis), it is possible to record the image from both  
sides of the sample that is embedded between thin cover slips. The first im-  
age substack covering the frontal half of the sample is taken from one side of  
the cover slip. The sample is then flipped, and the second image substack is  
35 recorded from the other side. For a thicker specimen, the sample can be em-  
bedded in a soft medium and using tissue sectioning (e.g., vibratome), a sample  
section is cut off from the top of the sample block after taking image stacks  
of this section. By repeating this procedure, called two-photon tomography, in  
principle any thick specimen can be imaged [9, 10].

40 In both approaches, image substacks should be acquired in an overlapping  
manner: the overlapping sections will serve as a guide indicating how neighbor-  
ing substacks can be concatenated, or stitched. However, such concatenation  
is not straightforward, because small rotation and translation can occur when  
the sample is flipped or when the block surface is cut off. Rotation can occur  
45 not only around the  $z$ -axis of the specimen but also around  $x$ - and  $y$ -axes (tilt-  
ing). Moreover, because of the photobleaching that occurs during image stack  
acquisition and because of the different depths from the sample surface, the  
intensity of the corresponding optical sections in two overlapping stacks often  
appear different.

50 This is why registration is crucial for properly stitching image substacks of  
the same sample. Each neuronal fiber from neighboring stacks must be perfectly  
connected to each other after stitching. A discontinuity in the final image stack  
would strongly affect further analysis such as neuron tracing [11, 2, 12]. Regis-  
tration is one of the most important general problems of image processing [13].  
55 Therefore, there exists a wide offer of commercial or free, manual or automated,  
2D or 3D, software platforms that address this issue. Application specific solu-  
tions, however, that are relevant for common types of samples and acquisition  
protocols, are often difficult to access for life scientist. Such a perspective could  
be judged as relatively narrow in a general computer vision context. However,  
60 as recently illustrated in this journal [14, 15, 16, 17], it is actually specifically  
meaningful in biomedical imaging where important communities of life scientists  
work on the same types of samples.

In this study, we have developed an optimized automatic registration and  
stitching algorithm, 2D-SIFT-in-3D-Space, specifically adapted for thick high-  
65 resolution laser-scanning microscopy image stacks. A visual flow chart of the  
proposed algorithm is given in Fig. 1. We applied our method to stitch large  
3D image stacks of *Drosophila melanogaster* brain samples that can be mutu-  
ally tilted by up to 20 °. Our algorithm, made available under the open source  
Fiji software that is widely used through the international bioimaging commu-  
70 nity, combines several existing approaches into a new strategy based on reliably  
detecting features in images using scale invariant feature transform (SIFT).  
In addition, to validate quantitatively the registration quality, we developed  
an original simulator that generates artificial 3D image stacks that mimic the  
properties of noise in laser scanning microscopy. We have used 2D-SIFT-in-  
75 3D-Space algorithm to assemble 3D image stacks of neurons of the *Drosophila*  
brain at a voxel resolution of 0.2 x 0.2 x 0.2  $\mu\text{m}$  (1600 x 1600 voxels and 800  
sections). The stitched dataset serves as a starting point for characterizing fine  
architecture of such large entire brain at unprecedented resolution but the 2D-  
SIFT-in-3D-Space is also shown to be useful to other types of datasets and other  
80 fluorescent microscopy systems (see supplementary data).



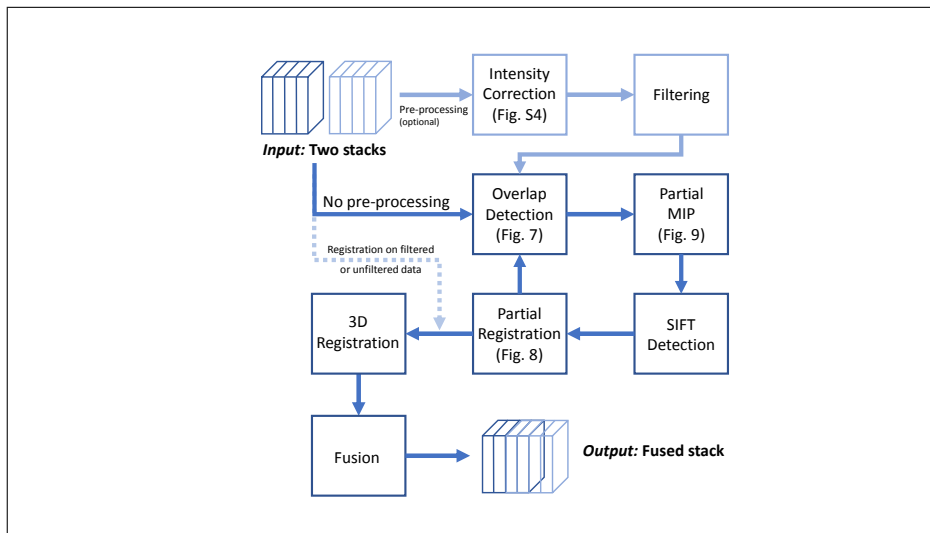


Figure 1: Flow chart of the proposed automatic registration and stitching algorithm 2D-SIFT-in-3D Space.

## 2. Related work

The primary purpose of the proposed registration method, 2D-SIFT-in-3D-Space, is to stitch two image substacks. The registration is achieved by comparing the signals of two overlapping image stacks. In this context, it shares some similarities with existing registration algorithms known to be adapted to the bioimaging community interested in the registration of 3D images. A possible classification for image registration approaches is whether the registration is based on intensity or on features [18].

Intensity-based approaches confront, with a correlation metric, the intensity patterns in images to be registered. This includes for instance software solutions such as CMTK (Computational Morphometry Toolkit), elastiX, ANTS, AMIRA). Intensity-based approaches are specifically suited when the homologous structures to be registered in the 3D stacks are well represented by the intensities through the spatial statistics of the gray levels seen as a random variable (see for instance in spatial intensity [19], or with Fourier transform of the intensities [20], or also as recently used in [21]).

Feature-based approaches, as chosen and developed in this article, realize a correspondence between homologous landmarks in the images to be registered. Feature-based approaches are specifically suited when the images are characterized by spurious small structures (e.g., vessels and blobs) such as the one highlighted in the samples of neuronal fiber images considered in this article. Feature-based approaches works in two steps: First the detection of landmarks and second the match of the homologous landmarks in the images. In some software solutions the landmarks have to be detected with another software or selected manually. Manual positioning of such landmarks can be very time consuming and can also be perturbed by human errors. To circumvent this difficulty automatic detection of homologous landmarks is preferred. This can be achieved by detecting fiducial structures of known shape and size such as fluorescence beads that are purposely added to the sample [22] or endogenous granules in the tissue [23]. Registration error, however, would occur if samples and added beads were put in fluid mounting medium such as glycerol used in our study. Whereas [23] utilized endogenous aging-related pigment granules that are distributed across the mouse neural tissue, similar approach was not applicable to our samples because such granules do not seem to be distributed ubiquitously in young fly brains. Automatic detection of homologous landmarks can also be done purely numerically based on the extraction of local image features (See [24] for a review). The great advantage of feature-based registration is that instead of using all image intensities, it is possible to register two stacks using only corresponding salient points as a statistic of the image content. Reducing the problem from full resolution image content to a relatively small number of corresponding homologous points simplifies the estimation of the transformation model and is a major computing time saver.

Feature-based and intensity-based techniques have been shown to be useful for registering image stacks of different tissue samples as well as for stitching sections and image substacks of a single tissue. Software designed for the stitching purpose tends to assume that all the sections are parallel without tilting (e.g., [20]) and do not correct rotation mismatches around all axes (see Sup-

plementary data for a comparison with our solution). Software designed for the registration purpose can generally correct rotation and translation in all 3D axes, but many of them are designed for registering image stacks that cover the same part of the specimen in different samples. To operate on the type of data considered in this article, one would require two image processing steps: first the manual crop of the overlapping part and second the stitching of the entire registered stacks. Effective manual crop may appear trivial for the samples with small tilting angles, but is actually a difficult task for large angles when dealing with self-similar samples such as the neurons in the brain. The existing techniques are, to the best of our knowledge, not natively capable of joint automatic detection of the overlap and registration and stitching of 3D image substacks that overlap only partially as found with the thick samples imaged with high-resolution laser-scanning microscopy.

To stitch image substacks with tilting errors, we propose a solution with an original use of a popular local feature SIFT introduced by [25]. Whereas we achieved 3D registration with the repeated use of 2D SIFT in three orthogonal planes, SIFT in true 3D space has also been proposed for motion recognition of video images [26], object recognition for X-ray computer tomography images [27], or in biomedical applications [28, 29, 30, 31]. Although the same approach can in principle be applied to develop an image registration software, the choice of 2D-SIFT-in-3D-Space is well adapted in our case, because the specific image acquisition sequence considered in this article stitching of microscope image substacks obtained from the same sample induces more important rotations around z-axis than the tilting around other axes; the first iteration of registration in x-y plane is thus very likely to bring a strong improvement. This prior would not be used with 3D SIFT registration, which would blindly look for solutions in any direction of the 3D space.

### 155 **3. Materials and Methods**

#### *3.1. Definition of the biological task*

As a model system, we used the brain of the adult fruit fly *Drosophila melanogaster*. It contains about 100,000 neurons in the volume that is approximately 600- $\mu\text{m}$  wide, 300- $\mu\text{m}$  tall, and 160- $\mu\text{m}$  thick. The fly brain has been  
160 an intense focus for brain-wide analysis of neural networks and their functions [32, 3, 33]. Projection patterns of neuronal fibers and distribution of synaptic connection sites can be visualized by expressing proteins that are spread along cytoplasm and those that are transported to presynaptic sites [32]. A wide variety of neuron types can be visualized using cell-specific expression driver  
165 strains [1]. Among them the dopaminergic and octopaminergic neurons – which are known to be involved in diverse brain functions – form extensive projections in almost all the brain regions [34, 35]. Because those neurons feature complex arborizations that are much denser and finer than most other neuron types (Fig. 2), high-resolution microscopy images are required. To this aim, im-  
170 ages of the fluorescent antibody-labeled samples were recorded using confocal laser-scanning microscopes (Olympus FV1000 and FV1200) with a 40x silicon immersion objective (NA=1.25) at an image resolution of 0.2 x 0.2 x 0.2  $\mu\text{m}$  each, in total 1600 x 1600 pixels and approximately 800 serial sections.

Image quality degrades as the plane of the scanning optical section goes  
175 deeper into the specimen (Fig. 3) even with high-resolution objective lenses that are designed to match the refraction index of the mounting medium. Thus, although the total thickness of the samples (Ca. 160  $\mu\text{m}$ ) is well within the working distance of the objective lens (Ca. 280  $\mu\text{m}$ ), deeper half of the samples cannot be recorded with optimum resolution if they are imaged only from one  
180 side. To address this issue, we mounted the specimen between thin cover slips on both sides with a space of 200- $\mu\text{m}$  thickness, and the deeper half of the sample is recorded from the other side by flipping the preparation (Fig. 4). The two substacks, taken from the front and back sides of the brain, were then merged after flipping the volume data of the back substack.

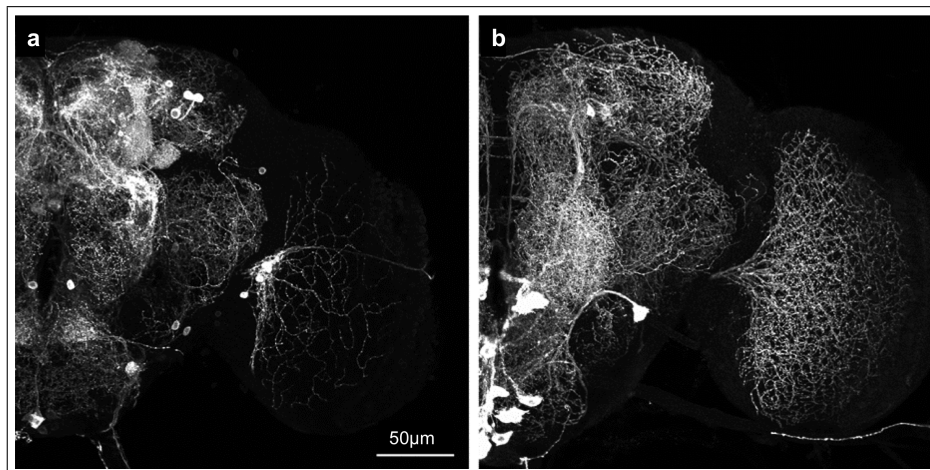


Figure 2: Spatial arborization of dopaminergic neurons (a) and octopaminergic neurons (b) of the adult *Drosophila melanogaster* brain. 3D reconstruction of the entire stack after front and back substacks are registered and merged. 1600 x 1600 x 800 voxels, voxel resolution =  $0.2 \times 0.2 \times 0.2 \mu\text{m}$ . Scale bar =  $50 \mu\text{m}$ .

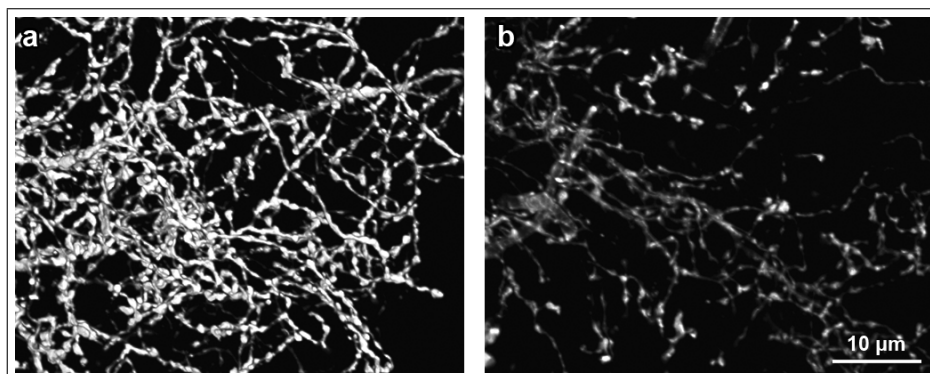


Figure 3: Difference in image quality along z-axis for the labeled fibers located in the area close to the sample surface (a: depth  $30\text{-}40 \mu\text{m}$ ) and rather deep in the specimen (b:  $80\text{-}90 \mu\text{m}$ ). Note that not only brightness but also sharpness is decreased. Scale bar =  $10 \mu\text{m}$ .

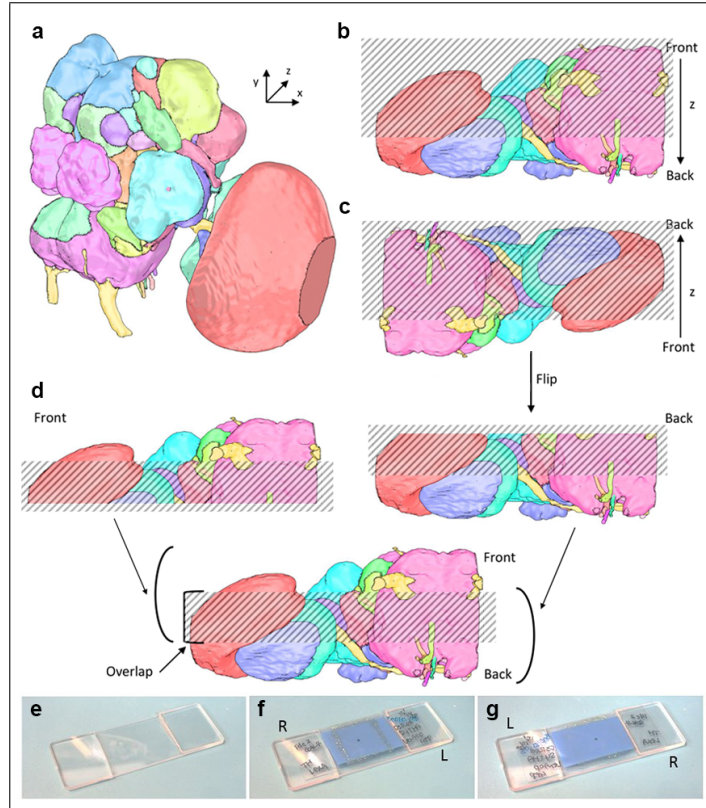


Figure 4: Image stacks acquisition. (a) A schematic view of the fly brain seen anterior-obliquely. (b) Optical sections of the front substack are acquired from the frontal surface of the brain toward its mid-level. (c) After flipping the sample on the microscope stage, images of the back substack are acquired from the back surface of the brain toward mid-level. (d) Image registration: the back substack is flipped and fused with the front one. (e-g) The sample is mounted between two thin cover slips. (e) Support plate for image acquisition. Glass spacers are put on both sides of the slide glass. (f) Front substack acquisition. The sample sandwiched between coverslips is put on the spacers. The space between the back coverslip and the slide glass avoids the occurrence of Newton ring that occurs when two glass material contact each other. (g) The sample is flipped for back substack acquisition.

185 This approach works well if the samples are mounted in rigid medium and if  
the sample after flipping can be placed exactly at  $180^\circ$  from the original direc-  
tion. However, fluorescent samples are often mounted in a fluid substrate such  
as 80% glycerol, in which the specimen are not completely fixed but stay afloat.  
Slight rotation may therefore occur when the samples are flipped. In addition,  
190 because of the instrumental error, the two cover slips may not completely be in  
parallel, causing the flip not to be exactly at  $180^\circ$ . Such error would not cause  
severe problems if the images were taken at relatively low resolution. However,  
to reconstruct fine neuronal fibers which are often thinner than  $0.5 \mu\text{m}$ , even  
subtle misalignment results in discontinuity such as gaps, if the substacks are  
195 concatenated without fine three-dimensional registration (Fig. 5).

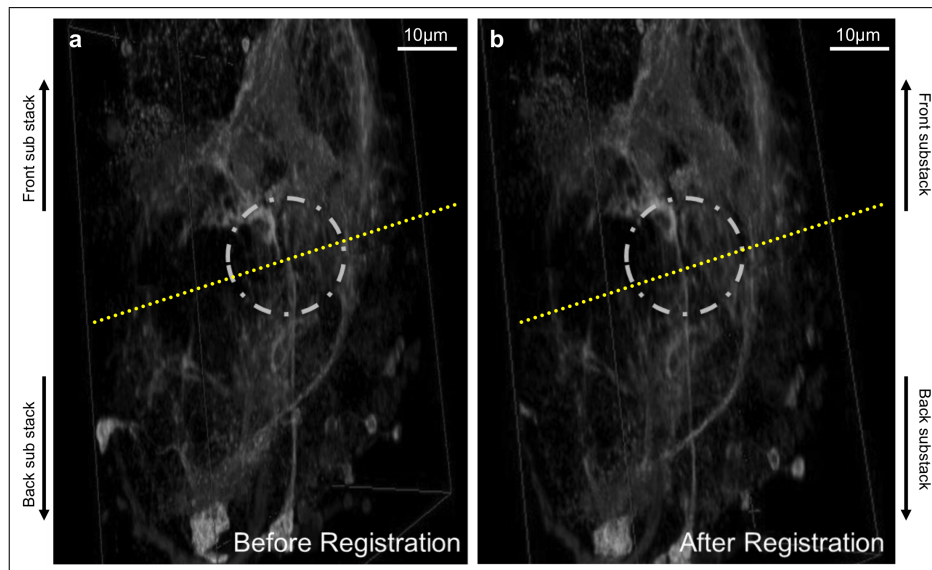


Figure 5: Importance of registration to avoid discontinuity in the final 3D image stack. Oblique view of the 3D reconstruction after concatenation of substacks. (a) Concatenation without fine registration. Gaps and seems (dotted circle) are observed in the neural fibers that should be continuous at the boundary between the front and back substacks (dotted line). (b) Concatenation after fine registration. Fibers appear continuous.

### 3.2. 2D-SIFT-in-3D-Space Algorithm

Our registration and stitching algorithm consists of several steps (Fig. 6):  
 (1) the overlap detection in the two data substacks, (2) partial maximum intensity projection (MIP), (3) filter selection, (4) SIFT features extraction and  
 200 correspondence detection, (5) affine transformation parameters estimation and  
 (6) registered overlapping portions fusion. The details of all the steps will be described in the following sections.

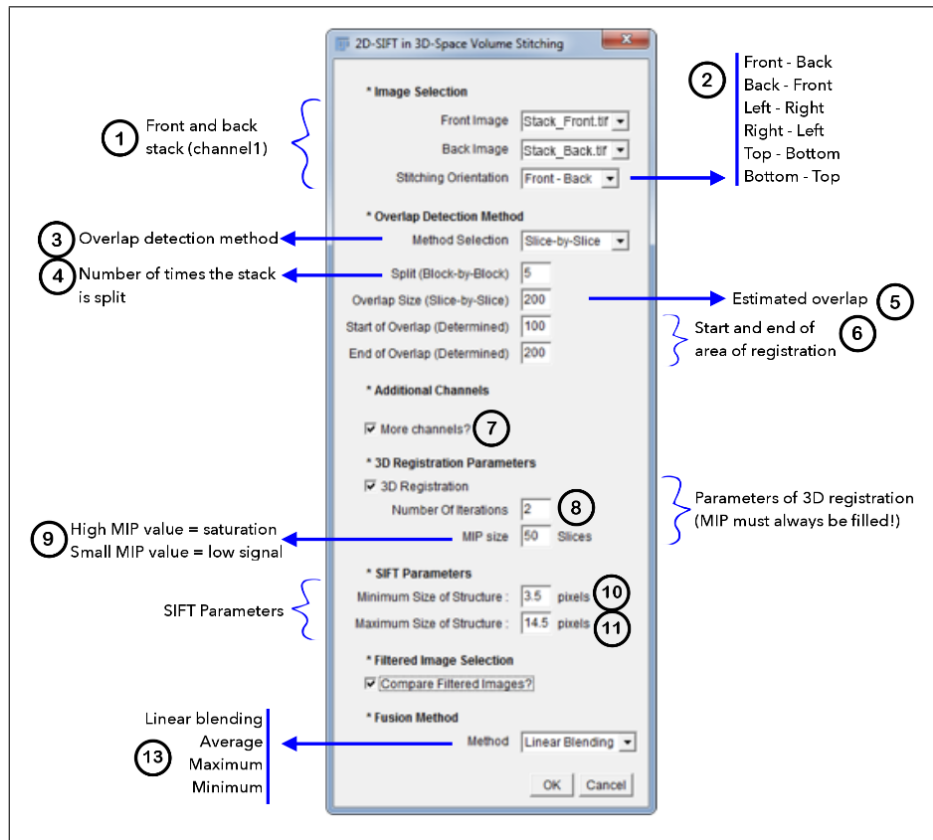


Figure 6: Control panel of the 2D-SIFT-in-3D-space Volume Stitching plugin, implemented for ImageJ/Fiji. See main text for detail.

*Overlap detection.* This step aims to find the portions of the two image substacks containing overlapping image data. To do so, we compare the image



205 content across sections of the two substacks using SIFT features (see Fig. 7a1-  
2). For each section comparison, the number of corresponding SIFT features  
is stored which allows the computation of the correspondence curve where the  
overlap positions are detected with the highest peak of the correspondence curve  
(see Fig. 7a3-4). To detect the overlap, we proposed two alternative approaches.  
210 The first approach performs a slice-by-slice comparison: the last section of the  
front substack is compared to each section of the back substack along the z-axis  
(Fig. 7a). The overlap positions are detected with the highest peak of the cor-  
respondence curve, in other words the section of the back substack with most  
common information with the last section of the front substack. This approach  
215 is based on the assumption that tilting between the planes of the two substacks  
is usually small enough to find important similarities between substack sections  
of the corresponding depths within the overlap. However, slice-by-slice compar-  
ison requires long computation time on substacks made of numerous sections.  
Moreover, when the tilting between the two substacks is big, similarity between  
220 sections is often not sufficient for identifying corresponding sections. For those  
reasons, we developed an alternative approach, block-by-block comparison (Fig.  
7b) using partial MIP presented in the following section.

*Partial MIP.* In this approach, we split the two entire substacks into several  
blocks and make the maximum intensity projection of the sections within each  
225 block (partial MIP). SIFT comparison is then performed between the partial  
MIP of the last block of the front substack with the partial MIP of each block  
in the back substack (Fig. 7b1). The number of blocks is determined by the  
parameter Split set in the control panel (Fig. 6). For example, if the splitting  
parameter is set at 2, the substacks will be divided into two blocks. The partial  
230 MIP of the last block of the front substack will then be compared with the MIPs  
of the two back-substack blocks and the best match is selected. The selected  
back-substack block is further split in two incrementally and compared with the  
equally split last block of the front substack (Fig. 7b2-3), until the successively  
split blocks contain only one section (Fig. 7b4). The splitting parameter can

235 be set by the user, and the effect of its choice will be discussed in the results section.

*Filter selection.* This step allows using preprocessed (filtered) images instead of original stack images to calculate registration parameters. The transformation model will be computed by comparing the SIFT features of the filtered images, 240 but the resulting model will be applied to the original data. Applying various filters to enhance biological structures to be registered may boost feature extraction and thus improve the final stitching on real images. For example, for our specific application, one can use a vesselness filter to enhance tubular structures of neuronal fiber images. For more generic applications, one can choose local 245 contrast filter to enhance faint signals or denoising filter to eliminate misleading signals. Such filters are available in ImageJ/Fiji software. Another effective way of using preprocessed images is to perform registration in 8-bit images instead of original images with larger bit depth (e.g., 12-bit or 16-bit), which will significantly reduce computation time. However, stitching of such preprocessed 250 images may not be scientifically pertinent, because it may affect signal intensity and distribution as well as bit-depth resolution. To provide users full flexibility, final concatenation can be performed either with the preprocessed data or unprocessed raw image stacks.

*SIFT features extraction and correspondence detection.* Comparison of sections 255 of the two substacks is performed using SIFT [25]. SIFT is a local descriptor that allow both automatic identification of salient points in a section – by detecting blobs within a specified size range using the Difference of Gaussian detector [36] – and extraction of features for these points. A feature consists in an invariant descriptor to scaling, orientation, and partially invariant to affine distortion and 260 illumination changes. For this step, we use the version of 2D SIFT algorithm developed by Stephan Saalfeld ([http://fiji.sc/Feature\\_Extraction](http://fiji.sc/Feature_Extraction)). Following this implementation, we provide two key parameters: (i) minimum and (ii) maximum size of biological structures (in pixels) to be detected, as depicted in Fig. 6. SIFT features are extracted at all scales between maximum and

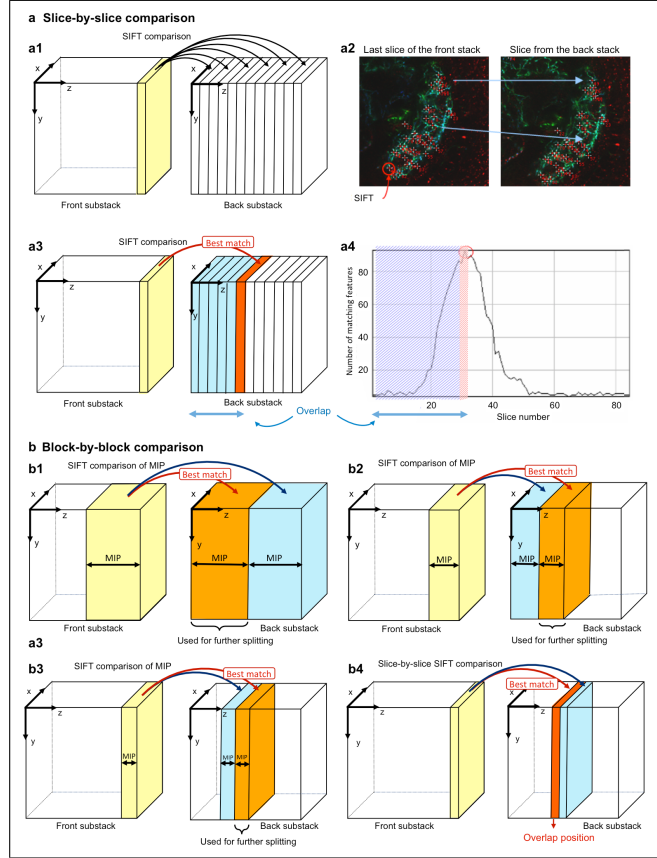


Figure 7: Overlap detection. (a) Slice-by-slice method: (a1) the last section of the front substack is compared to each section of the back substack along the z-axis using (a2) SIFT comparison until (a3) the best match is found. (a4) Overlap correspondence curve, which shows the number of corresponding SIFT features identified between the last section of the front substack and each slice of the back substack. The highest peak of the curve corresponds to the position of overlap (red section in a3). (b) Block-by-block detection: (b1) the back substack is split into several (here 2) blocks, and a MIP image is computed for each block. Each MIP of the blocks of the back substack is compared using SIFT with the MIP of the block of similar size (yellow) of the front substack. The best matching block (orange) is selected. (a2) The selected block is split again into several blocks, whose MIPs are compared using SIFT with the MIP of the corresponding thickness from the front substack to select the best-matching block (orange). (b3) This process is repeated with decreasing thickness until each block contains only one section (b4). The position of overlap is identified as the best matching section of the back substack (red section) compared to the last section of the front substack (yellow).

265 minimum size. Correspondence between SIFT features is identified by nearest  
neighbor matching in the local descriptor space. However, it results in a sig-  
nificant number of false correspondences. In our registration context, the set  
of SIFT correspondences are related by an unknown 2-d affine transformation  
 $T$ . The tool we used to separate true and false correspondences is the random  
270 sample consensus (RANSAC) implementation proposed by [37, 38]. In short,  
it works as follows: for a fixed number of iterations, it randomly selects a set  
of correspondence candidates and estimates  $T$  for them. The residual error of  
all candidates in terms of  $T$  is calculated and candidates with a residual error  
lower than some maximum displacement are collected as true correspondences.  
275 The largest set of true correspondences found is used to estimate the optimal  
 $T$ . Number of iterations and maximal displacement error are parameters of the  
implementation and were left to default values proposed in the implementation  
of [37, 38].

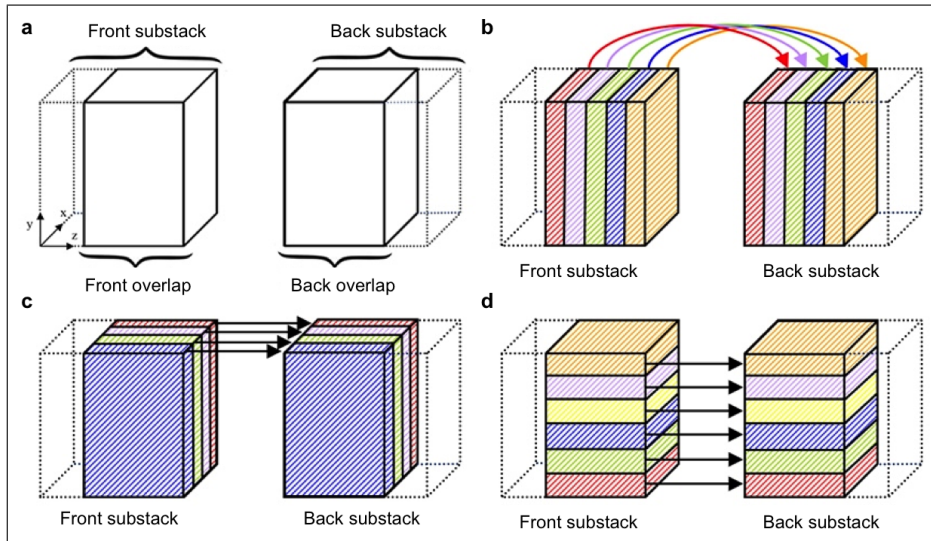


Figure 8: 3D registration by three consecutive 2D comparisons: (a) Overlapping volume between substacks is identified as shown in Fig. 7. (b-d) SIFT comparison is then performed between spatially corresponding pairs of partial-MIP slices, first in the x-y plane (b) and then y-z plane (c) and finally x-z plane (d).

*Affine transformation parameters estimation in three planes.* Registration is  
280 performed using only the data of the overlapping portions (Fig. 8a). Three  
successive rounds of registration are performed in three orthogonal planes (Fig.  
8b-d), where SIFT comparison is performed between spatially corresponding  
pairs of slices. First, the overlapping data portions of the front and back sub-  
stacks are re-sliced along the z-axis to obtain cross-section slices in the x-y plane  
285 (Fig. 8b). Slices of similar depth are compared using SIFT features between  
front and back overlapping portions to calculate translation parameter in the  
x-y plane and rotation parameter around z-axis. Second, the re-slicing step is  
performed along x-axis to obtain cross sections in the y-z plane, and the gener-  
ated slices are compared to calculate translation in the y-z plane and rotation  
290 around x-axis (Fig. 8c). Finally, the re-slicing step is performed along y-axis  
to calculate in the same way translation in the x-z plane and rotation around  
y-axis (Fig. 8d). This combination of steps allows for a 3D registration starting  
with the plane containing the most information and the potentially more dis-  
placement (which is linked to the acquisition protocol which rotates the sample  
295 according to the x-y plane). In theory, iterative uses of 2D SIFT might increase  
the chance that the whole registration would fail. It is not the case in our ap-  
proach, because we initiate the iterative process in the direction of the most  
likely important transformation.

In this processing step, comparison of single sections within the overlapping  
300 portion often do not contain enough amount of corresponding signals, so that  
extracted SIFT features in the section of the front substack may match with  
only a tiny number of features in a section of the back substack. To overcome  
this problem, instead of comparing single sections, we again proposed the use  
of partial MIP (Fig. 9). Besides its advantage of increasing the number of  
305 SIFT correspondences due to denser signals in each partial-MIP slice compared  
to single slices of the original stack, it also reduces the impact of rotation in  
the directions other than the one studied. The thickness of the partial MIP is  
determined by the parameter MIP size (in slices) in Fig. 6. This parameter  
is taken into account by performing the re-slicing step slice by slice and then

310 computing partial MIP on all the blocks of the specified size parameter from  
 resliced front and back overlapping data portions. The effect of using partial  
 MIP in this specific step will be discussed in the results section.

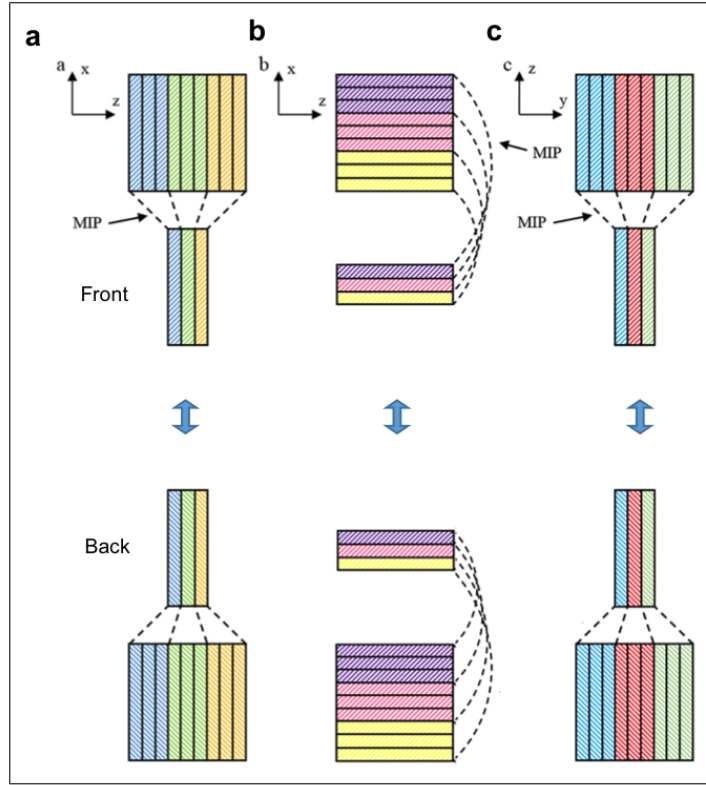


Figure 9: Creation of partial-MIP slices along  $z$ ,  $x$  and  $y$ . The image stack volume is re-sliced into three directions, and MIP is calculated from subsets of slices. (a) Partial MIP parallel to  $x$ - $y$  plane. (b) Partial MIP parallel to  $y$ - $z$  plane. (c) Partial MIP parallel to  $x$ - $z$  plane. Each partial-MIP slice contains more signals for feature extraction compared to thin single sections. Reduction of the total number of slices for comparison also reduces computation time.

The set of three registration rounds can be repeated as many times as necessary to further improve the precision of registration. Multiple application of  
 315 translation and rotation would result in the accumulation of image degradation caused by recalculation and interpolation of voxels at each registration step. To avoid this we combined the translation and rotation parameters calculated by all the previous steps and applied them to the original substack data to generate

the final registered substack data.

320 *Registered overlapping portions fusion.* Finally a fusion step enables to combine  
information from the two registered overlapping portions into a single image  
stack. It allows fixing the difference of brightness within the overlap due to light  
attenuation by tissue and photobleaching of signals during image acquisition.  
For this step, we use the stitching plugin developed by Stephan Preibisch ([http :](http://imagej.net/ImageStitching)  
325 [//imagej.net/ImageStitching](http://imagej.net/ImageStitching)). Following this implementation, different fusion  
methods such as Average, Maximum, Minimum and Linear blending of the two  
substacks can be selected.

### 3.3. *In silico simulation of laser-scanning microscopy image stacks*

To quantitatively evaluate our registration algorithm, we need a set of sample  
330 image substacks for which degrees of rotation and tilting are precisely known.  
Such quantitative comparison is difficult on real samples, because image stacks  
contain uncontrollable factors such as noise, signal attenuation and photobleach-  
ing, making them difficult to register precisely with human eyes. An alternative  
approach is to use in silico simulation to generate images for the evaluation of  
335 our registration algorithm. This approach starts by analyzing the noise in laser  
scanning- microscope images.

*Noise analysis.* The noise in 3D image stacks acquired by confocal laser-scanning  
fluorescence microscopy is known to follow a Gamma law of shape  $k$  and scale  
 $\theta$  [39]:

$$f(x; k; \theta) = x^{k-1} \frac{\exp(-x/\theta)}{\theta^k \Gamma(k)}, \quad (1)$$

340 for  $x > \theta$  and  $k > 0$  and  $\theta > 0$ . It must be emphasized that the parameters  $k$   
and  $\theta$  of the gamma law are not the same in the signal and in the background  
and are depth-dependent. Thus, we estimated those parameters for two different  
parts of the image: the signal (in our case neuronal fibers) and the background.  
For this purpose, we made a binary mask of neuronal fibers by thresholding each  
345 section of our real image stacks (Fig. 10a). The threshold was set automatically

(Fig. 10b) using Li entropy method [40]. The fiber signal is then extracted by multiplying the corresponding binary mask with the original stacks (Fig. 10c); whereas the background signal is obtained by multiplying the inverse of the binary mask with the original stacks (Fig. 10d).

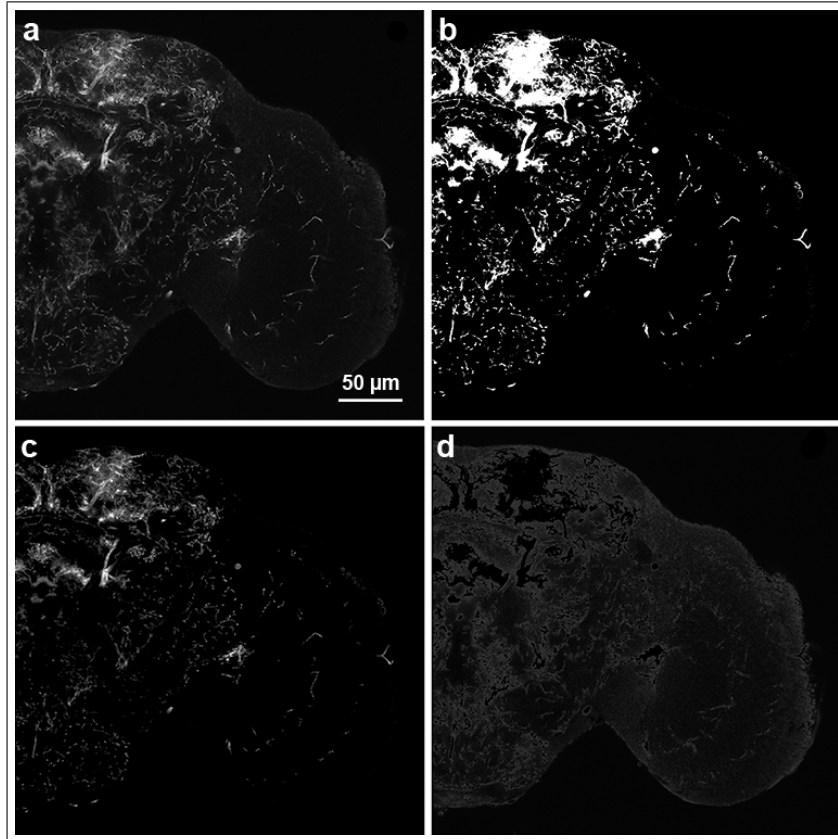


Figure 10: Noise estimation for a real image stacks of *Drosophila* brain. Image intensity in (a) and (d) is enhanced to make dark noises more visible. (a) Original image. (b) Binary image (threshold selected automatically using Li method). (c) Signal segmentation (original image within the binary image mask of b). (d) Background segmentation (original image outside the binary image mask of b).

350 To evaluate the nature of the noise observed in our datasets, the resulting fiber signal and background images were fitted to a gamma distribution to estimate the parameters  $k$  and  $\theta$  throughout the sections of the front and back



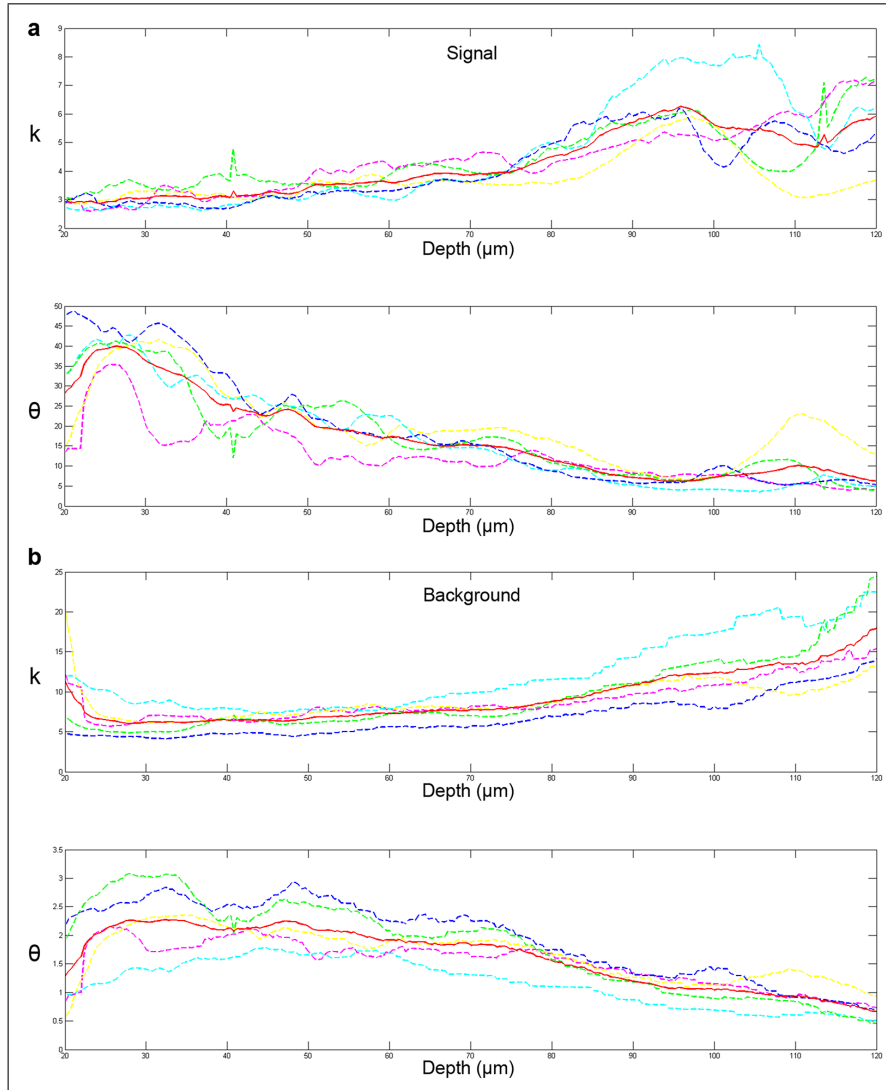


Figure 11: Evolution of the parameters  $k$  (top panel) and  $\theta$  (bottom panel) of the gamma law of the microscope noise within the signal portion (a) and background portion (b) of the front substack. Results of five different brain samples and their average (red). The first  $20 \mu\text{m}$ , which contains mostly the image of round cell bodies, are excluded from evaluation.

substacks in five different brain samples (Figs. 11-12). Noise in the fiber signal of the front substacks shows a symmetric evolution for  $k$  and  $\theta$  (Fig. 11a):  
355 while  $k$  is increasing,  $\theta$  is decreasing with the depth. In short, the increase of the shape parameter  $k$  with the depth traduces the convergence of the shape of the Gamma distribution towards a Gaussian distribution and the decrease of scale parameter  $\theta$  with the depth is coherent with the global decrease of the signal intensity observed in deep layers of the image stacks. Noise in the back-  
360 ground shows similar tendency as in the signal part (Fig. 11b). However, the  $k$  parameter is much greater in the background, thus the background noise tends to be more Gaussian. On the contrary, the  $\theta$  parameter is more than 10 times smaller in the background compared to the signal, which means that the signal distribution is much less spread out.

365 Noise in the signal and background parts of the back substacks shows similar evolution (Fig. 12): while  $k$  is increasing,  $\theta$  is decreasing with the depth in the same order of magnitude. However in the overlap (part of the sample that was already imaged from the frontal surface)  $\theta$  is reduced due to photobleaching. Using these data we estimated the noise parameters  $k$  and  $\theta$  as the average  
370 of the fitting curves of the five different brain samples in the signal and background parts for the front and back substacks (Figs. 11, 12). These averaged  $k$  and  $\theta$  parameters estimated through the depth can then be used as empirical laws for the evolution of noise in laser-scanning microscope modeled as a depth dependent gamma distribution.

375 *Simulation.* Based on the previous noise model, we created an in silico simulator of laser-scanning microscopy image stacks (Fig. 13). We first took the binary masks of the signal (i.e. fibers) and background structures (Fig. 13 left) generated for the study of the noise (Fig. 10). We then added two different types of simulated noise on the signal and background structures, respectively,  
380 depending on the depth from the surface of the virtual sample (Fig. 13 middle). Finally, signal and background images with added noises were merged to produce simulated laser-scanning microscopy images (Fig. 13 right).

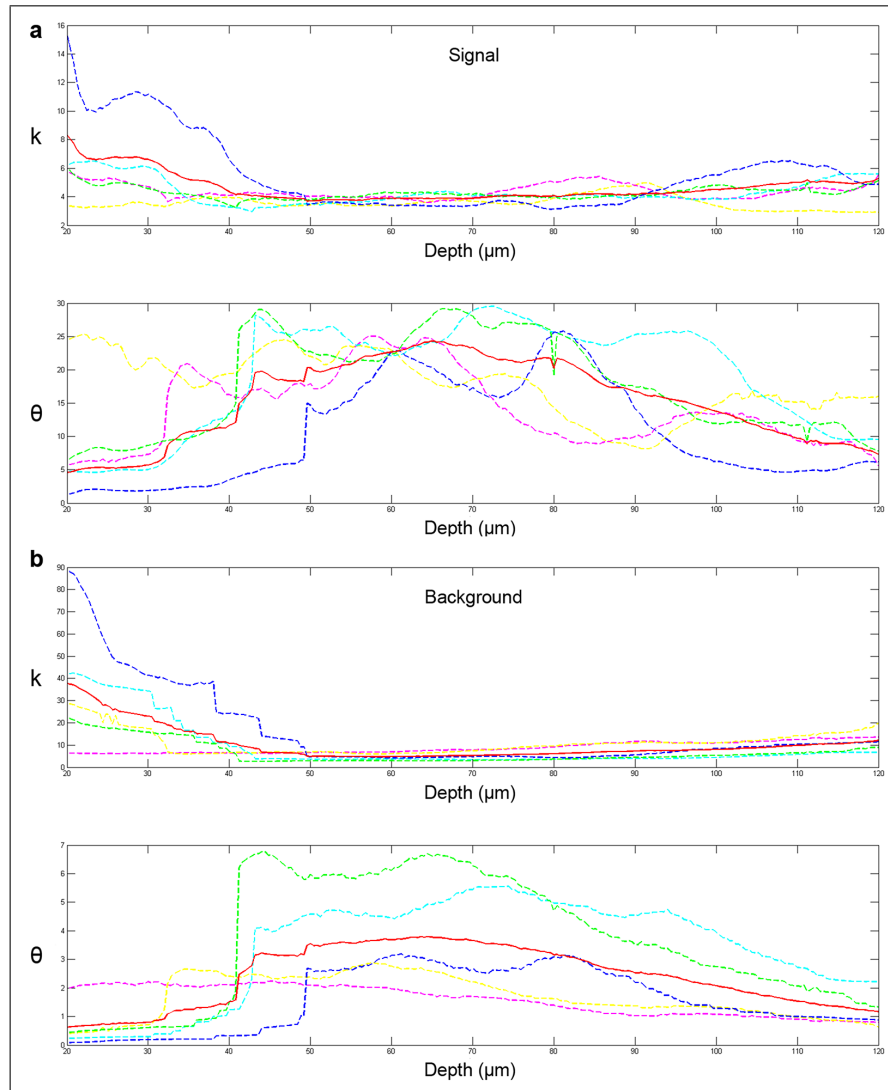


Figure 12: Evolution of the parameters  $k$  (top panel) and  $\theta$  (bottom panel) of the gamma law of the microscope noise within the signal portion (a) and background portion (b) of the back substack. Results of five different brain samples and their average (red), after removing the first 20  $\mu\text{m}$  (cell bodies).

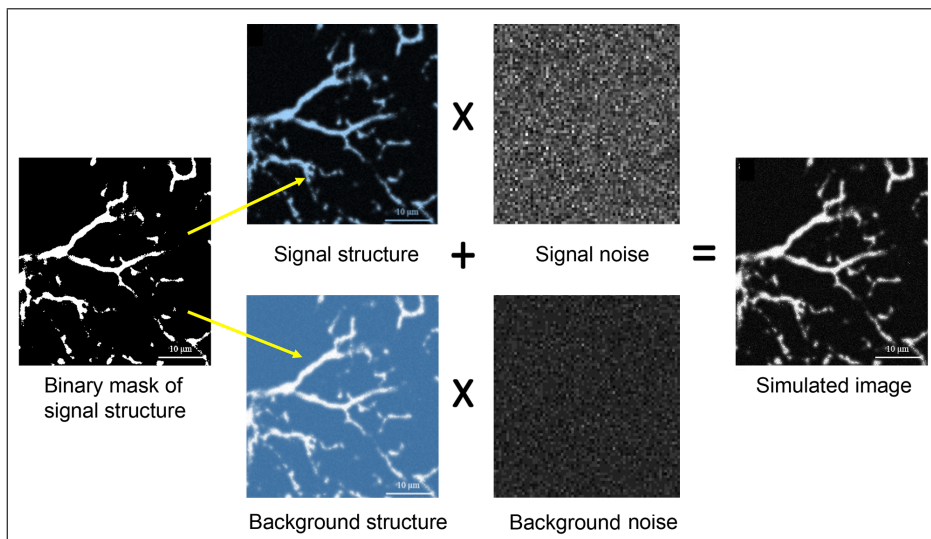


Figure 13: Overview of the in silico simulation of confocal microscopy image stacks. First, binary mask of the signal portion is generated from a real dataset of laser scanning microscopy. The mask appears sharp and bright regardless of the depth of the section in the sample. Next, simulated noise in the signal (fibers) and background according to the depth is convoluted to the binary masks. Finally, the signal and background images are merged.

In this study, we simulated a brain sample with a thickness of  $120\ \mu\text{m}$  (Fig. 14 top,  $800 \times 800 \times 300$  voxels). We divided the sample into two overlapping substacks (Fig. 14a top) and added depth-dependent noise to the sample regarding the structure parts and the nature of the substack (i.e., front or back). In these simulated substacks, the red-colored section in (Fig. 14a bottom) is respectively  $30\ \mu\text{m}$  and  $90\ \mu\text{m}$  away from the surface of the sample in the front and back substacks. Therefore it appears relatively clear in the front substack (Fig. 14d), but more blurred in the back substack (Fig. 14e). These observations are consistent with real laser-scanning images of neuronal fibers. Finally, by rotating and translating these virtual substacks, we are able to create tilted substacks with known transformation parameters to evaluate the performance and the accuracy of our registration algorithm.

Providing synthetic ground truth for binary structures (such as for the fly brain neurons in Fig. 14) is relatively straightforward. Much more effort would be required for textured samples, and, in practice this is not done (i.e. one rather resort to the sole visual inspection on real images to qualitatively assess the registration performance). We provide an illustration on such a textured sample with images of real fly leg in supplementary data. However for some biological structures, it is possible to generate ground truth numerically from scratch. This is for instance possible for spheroid i.e. cellular spherical aggregates of cells which constitute 3D in vitro models for life sciences (see supplementary data for images of real spheroid). Synthetic models of spheroid exist in the literature [41] and to further illustrate the genericity of our simulator we provided additional quantitative analysis with experiments on synthetic spheroids as illustrated in Supplementary material 5.

#### 4. Results

Detailed illustrations are given on the results of the experiments carried on simulated or real fruit fly brains in this section. Supplementary experiments successfully carried on real and simulated spheroids and real fly legs are given

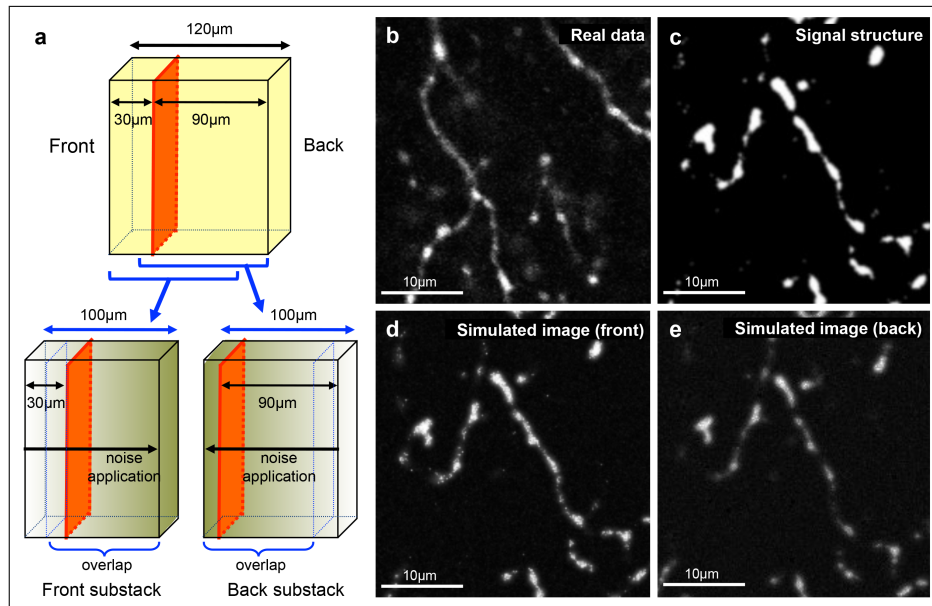


Figure 14: Generation of virtual substacks from a single image stack. (a) A single image stack (in this example the thickness of 300 sections = 120  $\mu\text{m}$ ) is divided into two overlapping substacks (with the thickness of e.g., 250 sections = 100  $\mu\text{m}$ ). Noise simulation is then applied to the front substack from front to back to simulate the imaging from the frontal side, and from back to front to the back substack to simulate the imaging from the back side. (b) Real image. (c) Binary mask of the signal structure. (d) In silico simulation of the image section if it is placed at 30  $\mu\text{m}$  from the surface. (e) The same section if it is placed at 90  $\mu\text{m}$  from the surface. Note that the signal appears darker and more blurred and background noise appears higher in (e) compared to (d), which is consistent with the image degradation in real confocal microscopy image stacks.

in the supplementary data so as to illustrate the generic value of the proposed method.

#### 4.1. Validation of registration accuracy

415 Using the simulator, we tested the quality of our registration algorithm (Figs. 15-16). We rotated the virtual back substack by  $1^\circ$  around x-,  $2^\circ$  around y-, and  $5^\circ$  around z-axes, applied the noise simulator, and subjected the data to 2D-SIFT-in-3D-Space registration. The performance of our algorithm can be qualitatively assessed in Fig. 15. The front substack in green is superposed  
420 to the back substack in magenta. The superposed image should appear white where the two substacks are perfectly registered. The two images do not match just after the overlap detection step (Fig. 15a). Subsequent registration in the x-y plane (Fig. 15b) and then x-z plane (Fig. 15c) and y-z plane (Fig. 15d) gradually improves matching and the overlapping images appear almost  
425 completely white with only a slight green or magenta blurred halo, attesting a good quality of registration. The final rotation matrix computed for this dataset is shown in Table 1a. When we convert this rotation matrix into Euler angles Rx, Ry and Rz (respectively rotations around x-, y- and z-axes), we obtain  $R_x = -1.2757^\circ$ ,  $R_y = -2.0232^\circ$ ,  $R_z = -5.0010^\circ$ , which are very similar to the opposite  
430 of the angles introduced artificially in our simulated data. Registration was very precise for z (registration error 0.019%) and y (1.160%) but worse for x (27.566%).

Registration can be improved by a second loop through the registration process. Indeed, after two iterations of registration, the rotation error around x is  
435 greatly improved (Table 1b, i.e., 0.428%). Higher number of iterations further decreases the rotation errors as well as translation errors, which are lower than 1% of the original rotation angles and less than 1.3 voxels ( $0.26 \mu\text{m}$ ) of translation. In practice, 2 iterations seem to be a good compromise between accuracy and computation time, because further registrations do not greatly improve the  
440 errors. It can be noted however from Table 1b that rotation error after multiple iterations becomes very small around z-axis, but remain a bit larger around x-

Table 1: Rotation parameters after registration of the simulated data. Initial substack tilt corresponds to  $R_x = 1^\circ$ ,  $R_y = 2^\circ$ , and  $R_z = 5^\circ$ . 1 voxel =  $0.2 \mu m$ . Total rotation error =  $\sqrt{\Delta R_x^2 + \Delta R_y^2 + \Delta R_z^2}$  and Total translation error =  $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$

a: Conversion matrix (Substack size  $800 \times 800 \times 250$ , i.e., overlap: 200, after single iteration)

```

0.9955825655501388    0.08765056866153828    -0.03262316976298832
-0.08711895143914962    0.9957863605004061    0.023091300973873324
0.03532588980227439    -0.02226238860907968    0.9997414741550107
Rx = -1.2757 Ry = -2.0232 Rz = -5.0010

```

b: Registration results after different number of iterations (Substack/overlap size are the same as above)

Iter #	Euler angles:			Rotation error:			Total rot. error (%)	Total trans. error (voxels)
	Rx	Ry	Rz	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$		
	( $^\circ$ )	( $^\circ$ )	( $^\circ$ )	(%)	(%)	(%)		
1	-1.2757	-2.0232	-5.0010	27.566	1.160	0.019	27.590	1.898
2	-0.9957	-1.9853	-4.9998	0.428	0.734	0.004	0.850	1.292
3	-0.9914	-1.9908	-5.0043	0.855	0.461	0.085	0.975	1.243
4	-0.9902	-1.9865	-5.0015	0.976	0.674	0.030	1.186	1.221
5	-0.9927	-1.9888	-5.0004	0.733	0.560	-0.009	0.922	0.928

c: Registration results for different amount of overlap (Results after 2 iterations)

Overlap (voxels)	Euler angles:			Rotation error:			Total rot. error (%)	Total trans. error (voxels)
	Rx	Ry	Rz	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$		
	( $^\circ$ )	( $^\circ$ )	( $^\circ$ )	(%)	(%)	(%)		
200	-0.9957	-1.9853	-4.9998	0.428	0.734	0.004	0.850	1.292
150	-1.0029	-1.9854	-4.9982	0.292	0.730	0.036	0.787	1.867
100	-1.0259	-1.9925	-5.0006	2.594	0.377	0.012	2.621	2.895
50	-1.0122	-2.0001	-5.0106	1.218	0.007	0.213	1.237	3.884



and y-axes. The better registration accuracy around z-axis compared to x and y can be explained by the size of the image slices. Indeed, the x-y slices are much larger (800 x 800 pixels) than the x-z and y-z slices (800 x 200 pixels),  
445 leading to a greater number of matching features. Registration performance can also be examined using the overlap correspondence curve, which shows the number of matching features between the last section of the front substack and each section of the back substack (Fig. 16a). After the overlap detection (zero iteration), the corresponding curve appears very broad with a small number of  
450 matching features (blue line in Fig. 16a). Indeed, the front and back substacks are not registered yet and thus the similarity between them is still rather low. On the contrary, the correspondence curve after the first round of registration is much thinner and sharper, and the number of features is three times larger, traducing a higher similarity between the front and back substacks (see Fig. 16b)  
455 and thus the efficiency of the registration. The correspondence curves after the second iteration onwards are essentially superposition of the image stacks, without showing further improvement (Figs. 16c-d). Finally, because image size is likely to affect registration quality, and because image sizes in x-z and y-z slices are determined by the overlap size, we also investigated its effect. When we  
460 reduced the overlap by taking fewer sections from the front and back substacks, registration error increased (Table 1c). If the overlap size was too small (e.g., 50 voxels), the algorithm struggled to find rotation around x and y with the default parameters. However, modification of those parameters (here reducing the minimum size of structures to 1) fixed the problem (Table 1c, bottom).  
465 These results show that multiple iterations and large enough overlap size are both important for precise registration.

#### *4.2. Performance comparison between block-by-block and slice-by-slice overlap detection*

We evaluated the accuracy of the block-by-block and slice-by-slice overlap  
470 detection approaches. Block-by-block approach is meant to allow registration of substacks with large tilting angles. To determine the tilting limit of this

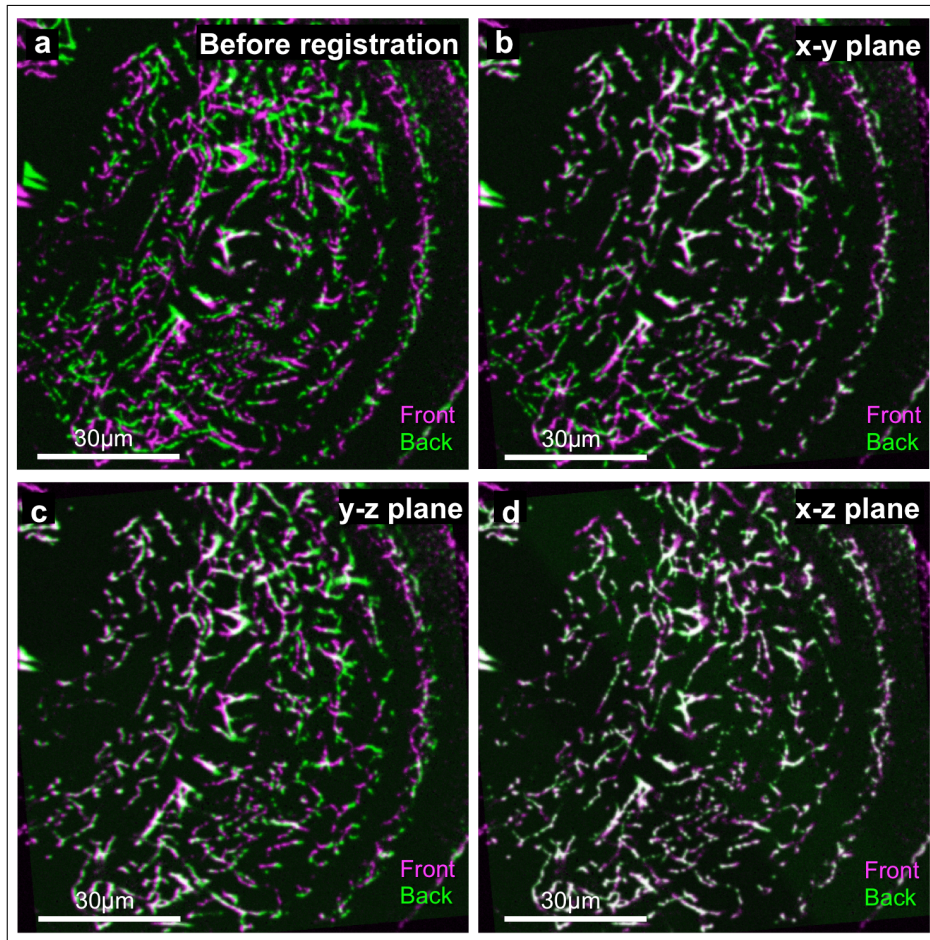


Figure 15: Superposition of the front (magenta) and back (green) substacks at each step of registration. (a) Before registration. (b) After registration in x-y plane (around z). (c) After registration in y-z plane (around x). (d) After registration in x-z plane (around y).

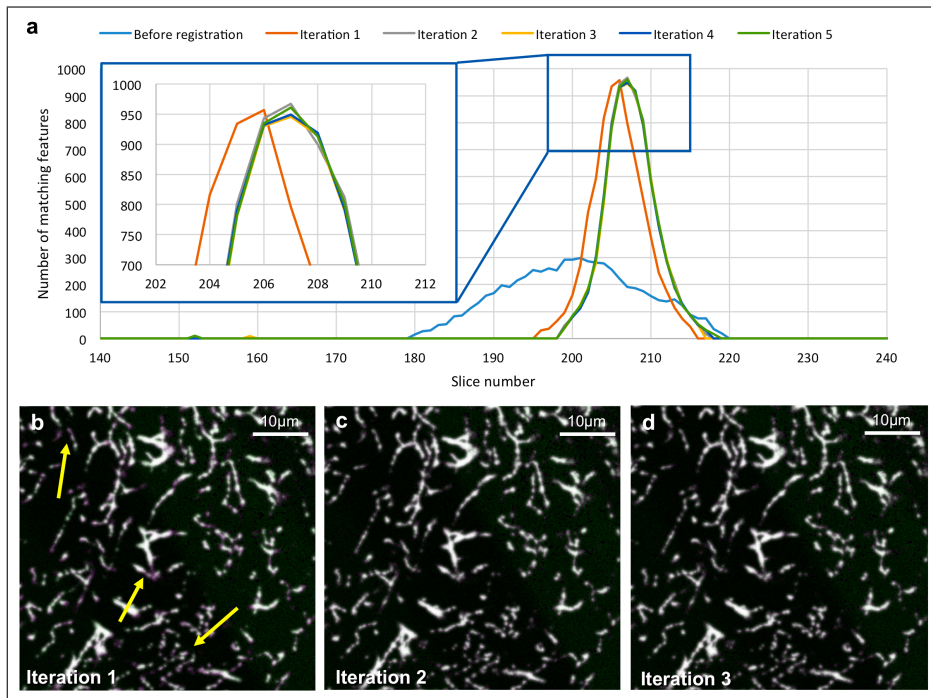


Figure 16: Registration of simulated data with small tilting ( $R_x = 1^\circ$ ,  $R_y = 2^\circ$ ,  $R_z = 5^\circ$ ). (a) Overlap correspondence curve for zero to five iterations. Inset shows the expanded view of the graph around the peak to better show the overlapping lines. (b-d) Superposition of the corresponding sections of the front substack (green) and registered back substack (magenta) after (b) one, (c) two and (d) three iterations.

approach, we fixed the tilting angles of  $R_z = 5^\circ$ ,  $R_y = 2^\circ$  and gradually increased the value of  $R_x$  by  $5^\circ$  steps and submitted the data to slice-by-slice (Table 2a) and block-by-block (Table 2b) approaches. The splitting parameter
   
 475 was set at 5 for block-by-block approach. Please note that both approaches used partial MIP in the registration step and that its size was set to 20 slices for both versions. Both overlap detection approaches performed similarly for small tilting angles ( $< 15^\circ$ ) leading to good results. However, stitching with
   
 480 than slice-by-slice approach. The difficulty of overlap detection increases with tilting because the similarity of sections perpendicular to z-axis decreases. Overlap detection with slice-by-slice approach becomes difficult as  $R_x$  reaches  $20^\circ$ , where the correspondence curve does not present the typical bell-shape but is extremely noisy with very few matching features (Fig. 17a). Thanks to the registration step around the three axes, overlap detection was drastically improved
   
 485 after a few registration iterations in terms of matching features (orange curve in Fig. 17b). However, the computed rotation parameters were not as accurate as expected, with high rotation error rate around x- and y-axes (6.35% and 19.1% respectively, Table 2a bottom).

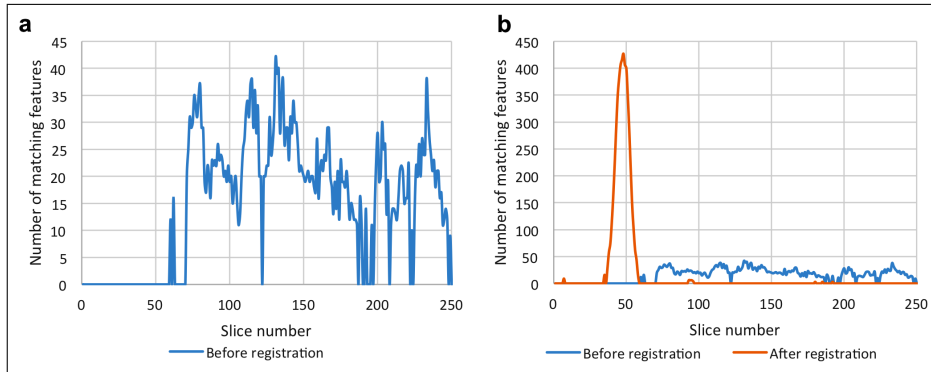


Figure 17: Registration of simulated data with big tilting ( $R_x = 20^\circ$ ,  $R_y = 2^\circ$ ,  $R_z = 5^\circ$ ) with the slice-by-slice overlap detection. (a) Overlap correspondence curve before registration. Highest peak was at the 131st section. (b) Comparison of the curves before and after registration. Actual matching peak after registration was at the 49th section.

Table 2: Comparison of registration errors between block-by-block and slide-by-slide overlap detection for increasing tilting angles. (Rx = 5 to 20°, with Ry = -2° and Rz = 5°, substack size: 800 x 800 x 250 . Partial MIP size is set at 20 slices. Computation time was measured on a computer with an Intel Core i7-4930k processor (6-core) running at 3.4GHz.)

a: Slice-by slice overlap detection

Tilt	Euler angles:			Rotation error:			Total	Total	Comput.
	Rx	Ry	Rz	$\Delta$ Rx	$\Delta$ Ry	$\Delta$ Rz	rot. error	trans. error	Time
(°)	(°)	(°)	(°)	(%)	(%)	(%)	(%)	(voxels)	(min)
5	-4.99	2.00	-4.99	0.036	0.180	0.007	0.18	2.38	71
10	-9.99	2.01	-5.00	0.084	0.566	0.017	0.57	4.50	53
15	-14.98	2.00	-5.00	0.081	0.086	0.046	0.12	6.54	45
20	-18.72	-1.61	-4.96	6.352	19.05	0.616	20.09	12.13	37

b: Block-by-block overlap detection

Tilt	Split	Euler angles:			Rotation error:			Total	Total	Comput.
		Rx	Ry	Rz	$\Delta$ Rx	$\Delta$ Ry	$\Delta$ Rz	rot. error	trans. error	Time
(°)		(°)	(°)	(°)	(%)	(%)	(%)	(%)	(voxels)	(min)
5	5	-4.99	2.00	-4.99	0.012	0.126	0.049	0.13	2.38	40
10	5	-9.99	2.00	-4.99	0.090	0.356	0.095	0.48	4.46	26
15	5	-14.97	2.00	-5.00	0.162	0.125	0.079	0.22	6.65	23
20	5	-168.2	-1.24	-5.44	741.2	37.8	8.8	742.3	16124	10
20	2	88.93	58.6	-88.4	544.6	2832.8	1669.8	3333.1	368420	11
20	1	-19.93	2.00	-5.01	0.319	0.126	0.242	0.42	8.78	33

490       Stitching with block-by-block approach at  $R_x = 20^\circ$  also caused erroneous  
 computed rotation angle, with an error rate as high as 741% and 79.9% around  
 x- and y-axes (Table 2b, 4th row). Because the thickness of the detected initial  
 overlap was small (77-section thick), the matching features were concentrated  
 only in a small portion of the data (Fig. 18a, green circle). To address this we  
 495 then decreased the splitting parameter to two, but the detected overlap thickness  
 was again as low as 76 sections (Fig. 18b, green circle), causing similarly high  
 error rates (Table 2b, 5th row). In our algorithm, we first detect the overlapping  
 portion of the substacks and then perform subsequent registration only using  
 this portion of the data. With high tilting angle between substacks, only a small  
 500 part of the overlapping volume might be detected. However, we can effectively  
 skip overlap detection by setting the splitting parameter to one (i.e., no splitting)  
 so that the program uses the data of the entire substacks to compute registration  
 (Fig. 18c, green circle). This strategy resulted in much improved rotation error  
 as low as 0.32% (Table 2b, last row).

505 *4.3. Influence of the splitting parameter in block-by-block approach*

We next analyzed the effect of the splitting parameter for block-by-block  
 overlap detection for a moderate tilting angle of  $R_z = 5^\circ$ ,  $R_y = 2^\circ$  and  $R_x =$   
 $5^\circ$  (Table 3). We performed registration for splitting parameter ranging from 1  
 (no split) to the number of the sections in the substack (corresponding to the  
 510 slice-by-slice approach). After 5 iterations, registration accuracy was essentially  
 the same for all splitting parameters (Table 3), but it should be noted that  
 computation time was long when the split value = 1 (58 min) because the entire  
 substack data are subjected to subsequent registration. Because the overlap  
 does not cover all slices in case of moderate tilting angles, many SIFT features  
 515 will not match and their consideration does not bring any improvement to the  
 registration result. Because of this, a split value of 1 must be used only in case  
 of high tilting between the two stacks as discussed in the previous subsection.

When splitting was activated (split value  $\geq 1$ ), computation time depended on  
 the total number of SIFT comparisons for overlap detection (Table 3, second

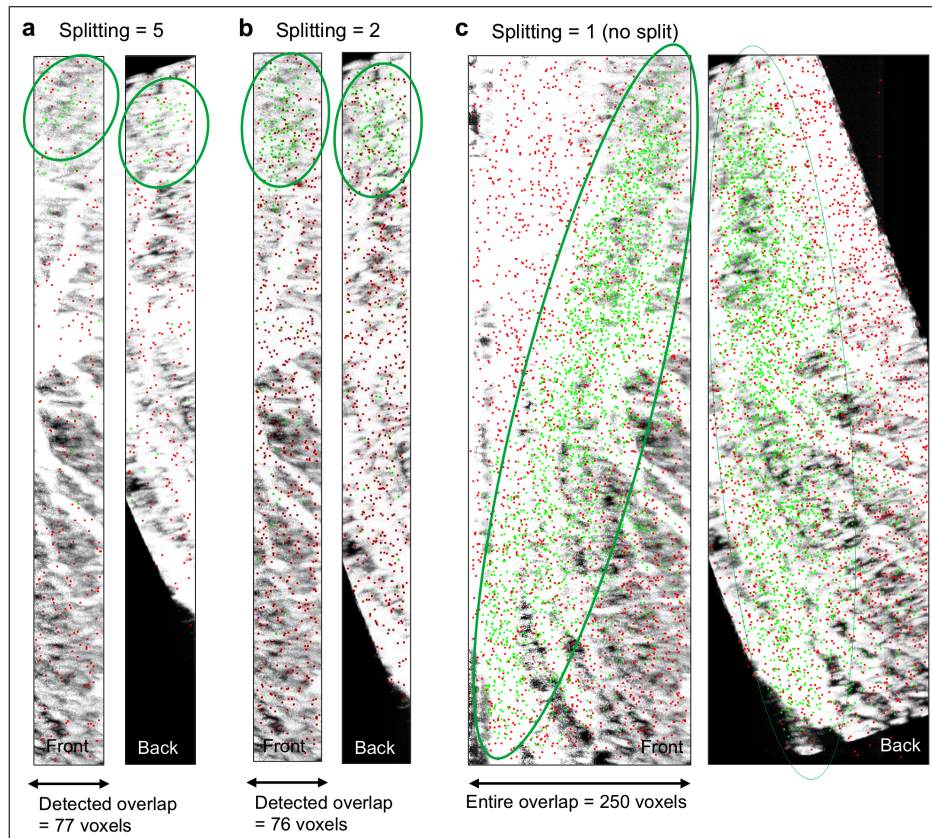


Figure 18: Repartition of matching features (green) of artificial data with big tilting ( $R_x = 20^\circ$ ,  $R_y = 2^\circ$ ,  $R_z = 5^\circ$ ) using the block-by-block overlap detection algorithm with the splitting parameter of 5 (a), 2 (b), and no splitting (c). Detected overlap was 77, 76 voxels for (a) and (b). In (c), the algorithm uses the data of the entire substacks (250 voxels).

Table 3: Registration errors and computation time of the block-by-block algorithm for different splitting parameters with  $R_x = 5^\circ$ ,  $R_y = -2^\circ$  and  $R_z = 5^\circ$ . Partial MIP size is set at 20 slices.

Split	Compa- rison #	Euler angles:			Rotation error:			Total rot. error (%)	Total trans. error (voxels)	Comput. Time (min)
		Rx ( $^\circ$ )	Ry ( $^\circ$ )	Rz ( $^\circ$ )	$\Delta R_x$ (%)	$\Delta R_y$ (%)	$\Delta R_z$ (%)			
1*	1	-4.99	2.003	-4.9977	0.061	0.162	0.046	0.180	2.38	58
2	15	-4.997	2.000	-5.001	0.061	0.031	0.034	0.076	2.44	32
5	17	-4.999	2.002	-4.997	0.012	0.126	0.049	0.136	2.38	40
10	22	-4.998	2.000	-4.998	0.028	0.023	0.024	0.042	2.40	44
15	31	-4.991	1.999	-5.001	0.164	0.028	0.030	0.169	2.38	46
20	32	-4.999	2.005	-5.001	0.013	0.271	0.033	0.273	2.49	48
50	55	-4.998	2.003	-4.998	0.032	0.160	0.035	0.167	2.37	60
125	76	-4.998	2.004	-4.998	0.040	0.199	0.029	0.205	2.42	125
250**	71	-4.998	2.003	-4.999	0.036	0.180	0.007	0.184	2.38	250

\* No split, \*\*Effectively the same as slice-by-slice comparison

column). This comparison number is linked to the split value. For our simulated  
substacks (250 sections thick) and a split value of 5, the program first splits the  
substacks into 5 blocks of 50 sections and does 5 SIFT comparisons (between  
the last block of the front substack and the 5 blocks of the back substack) to  
find the best match. The selected block is then split into 5 blocks of 10 sections  
and compared with a block of the same size from the front substack (5 compar-  
isons). The best-matching block is further split into 5 blocks of 2 sections  
and once again compared five times. Because the new selected block contains  
only 2 sections, it cannot be further split and a slice-by-slice comparison is used  
to determine the exact position of overlap (two comparisons). Thus, we need  
in total 17 comparisons to detect the overlap size. This number of compar-  
isons is much smaller than that of the slice-by-slice comparison (250). However,  
the computation time was only about two times faster (40 min versus 71 min,  
Table 3). This is likely because thick partial MIP images present more SIFT  
candidates. The number of comparison increases as we set higher splitting pa-  
rameter, which results in longer computation time (Table 3). Thus, for practical  
purposes splitting parameter between 2 and 5 should be preferred. However, if  
the sections contain very dense signals, setting a too small split value may cause  
saturation or too many unmatching features. The running time then increases  
a lot but not the quality of registration.



540 *4.4. Impact of image filtering on registration accuracy*

As previously discussed in introduction, signal quality tends to decrease with the depth in the samples. Unless adequately compensated, images of the sections that are deep in the specimen tend to appear darker than those that are close to the sample surface because of the attenuation of light passing through the specimen and photobleaching of fluorophores during image acquisition [5].  
545 Despite the relative robustness of SIFT matches toward brightness difference between the two images to be registered, it can be helpful to perform a filtering step of the image substacks to compensate the brightness mismatch of comparable sections before registration. To address this problem we developed a Fiji/ImageJ plugin Progressive Intensity and Gamma Correction (see Figs. S4, S5 of the Supplementary data). Using this plugin we tested the effect of intensity/gamma correction on the precision of registration. Fig. 19 compares the results of overlap extraction for the same pair of substacks with or without intensity/gamma correction. Compared to the original substack (Fig. 19a),  
555 intensity was corrected between 100% and 300% linearly and gamma between 1.0 and 1.4 linearly from the first (shallowest) to the last (deepest) sections of substacks (Fig. 19b). After the overlap detection (before registration), the overlap correspondence curve appears much thinner and sharper, and the number of features is five times larger (Fig. 19c) after intensity/gamma correction (blue line) than with the original data (green line). Indeed, corresponding curve of the original data presents a plateau, making the overlap size determination confusing, whereas the curve of the corrected data shows a sharp peak from which we can precisely deduce the overlap size. This difference can be noticed as a slight shift along the z-axis when the front and back substacks are merged.  
560 In the cross section view of the stack (Fig. 19d, e), we can see that the superposition works better (white color without green or magenta fringes) for the intensity-corrected data (Fig. 19e) than for the original data (arrows in Fig. 19d).

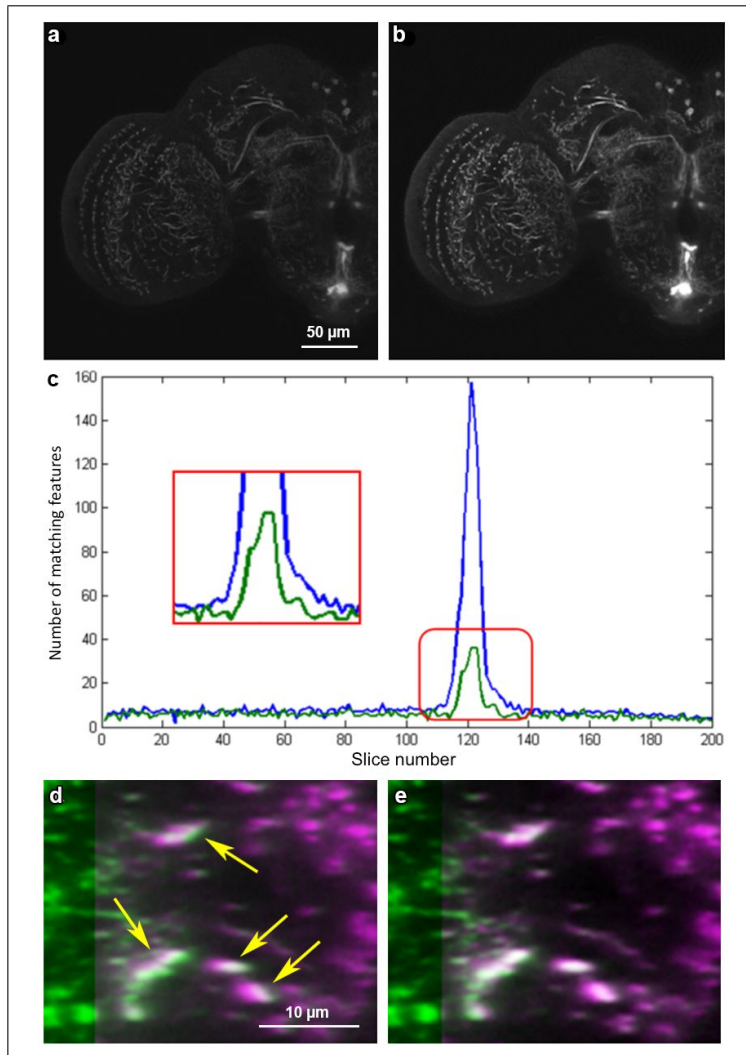


Figure 19: Improved registration after intensity and gamma correction. (a) Original data. (b) Corrected data (Intensity 100% to 300% linear, gamma 1 to 1.4 linear). (c) Number of SIFT features in the overlap detection for original data (green) and intensity/gamma corrected data (blue). Rectangle shows the plateau of the matching peak without intensity and gamma correction. (d, e) The x-z cross-sections of the stack, showing superposition of the front (green, left side of the panels) and back substack images (magenta, right side of the panels). Registration on original data (d) and corrected data (e). Brightness is enhanced for better visualization. Arrows in (d) indicate green and magenta fringe caused by the mismatch between the front and back substacks, which are hardly seen in (e). Scale bars = 50  $\mu\text{m}$  (a, b), 10  $\mu\text{m}$  (d, e).

#### 4.5. Performance on real data

570 Our plugin can efficiently perform very fine adjustments and precisely register small objects such as neuronal fibers and presynaptic sites. Fig. 20 shows the performance of our registration algorithm for the real image stacks of the *Drosophila* brain at a voxel resolution of  $0.2 \times 0.2 \times 0.2 \mu\text{m}$  (1600 x 1600 voxels and 800 sections). The front substack (Fig. 20a, shown in magenta in Fig. 20c) was registered and superposed to the back substack (Fig. 20b, shown in green in Fig. 20c). In the overlaid image (Fig. 20c), green and magenta cast should appear on the opposite sides of the objects if two stacks are out of alignment. The lack of such cast attests perfect registration. (Note: green or magenta cast that appears on the entire object is caused by the voxel intensity differences between front and back substacks.) The stitched dataset serves as a starting point for characterizing the fine architecture of the visualized neurons at very high resolution.

#### 4.6. Effect of optical aberrations

2D-SIFT-in-3D-Space Volume Stitching performs only affine transformation. This is because the section images of the same sample obtained from different sides should have same morphology. However, we found that this assumption may not always be true. In a few cases, we noticed that peripheral areas of the image could not be registered completely even when complete match was achieved in the central parts of the image. There is apparently a slight non-linear distortion between the section images of the same part of the sample obtained from the different substacks (Fig. 21). This is most likely because of the optical aberrations caused by the microscope objectives and scan optics. Among Seidels five aberrations, spherical aberration, coma, and astigmatism mainly affect the sharpness of the image without affecting the location of the signal. Those aberrations may cause uneven sharpness in the image field. Then sharper parts of the image with the highest feature correspondences concentration will have a stronger influence on the model computation, leading to perfect registration in those parts but a shift between the front and back substacks in other parts of

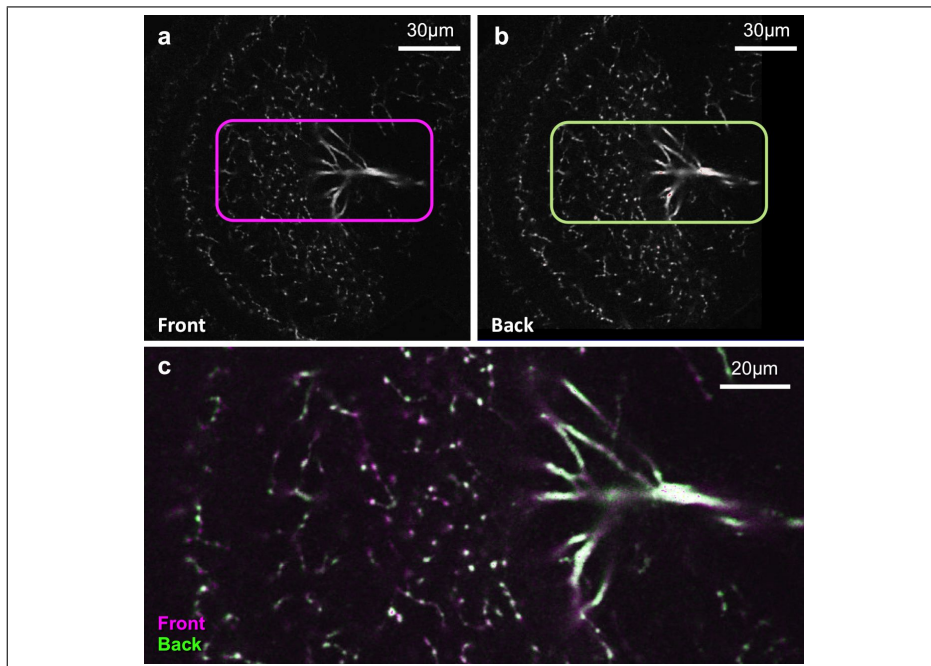


Figure 20: Registration and superposition of the substacks. Example of a real image stacks of the *Drosophila* brain (a) A section from the front substack, and (b) its corresponding section from the back substack after registration. (c) Close-up view of the superposition of the two sections (rectangles in upper panels). Magenta (section of the front substack) and green (section of the back substack) images are superposed. Scale bars = 30  $\mu\text{m}$  (a, b), 20  $\mu\text{m}$  (c).

the image. Moreover, the other two aberrations field curvature and distortion  
600 cause distorted projection from the signal distribution in the sample to the ob-  
tained 3D image space, making straight lines in the sample appear curved along  
z-axis (field curvature) or in the x-y plane (distortion). An objective lens tuned  
for the best performance at certain focal depth in the sample may not perform  
optimally at shallower or deeper levels. Thus, the optical sections of the same  
605 part of the specimen but obtained from different sides (i.e., at different depth)  
may not be identical. Adaptive optics and post-imaging distortion correction  
should be helpful to reduce the effect of optical aberrations [42, 43, 44].

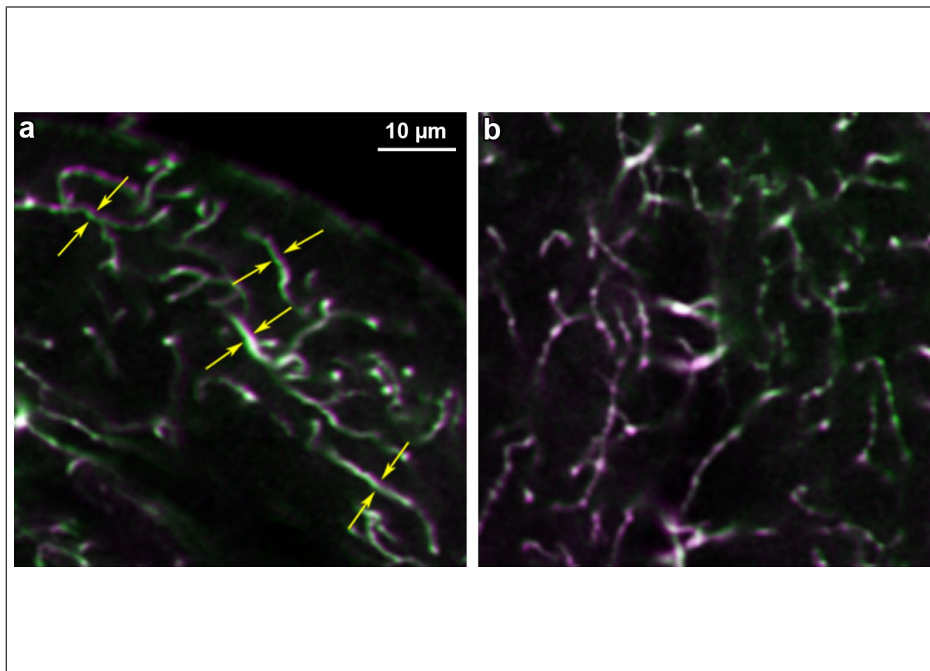


Figure 21: Illustration of registration errors caused by optical aberration. Superposition of the front substack (green) and back substack (magenta). (a) The two substacks are registered nicely in the central part of the image. (b) In the peripheral areas of the image, however, small mismatch tends to remain regardless of the registration around any axes (arrows).

## Conclusion

Although confocal laser-scanning fluorescent microscopy is used extensively  
610 in many biological laboratories to visualize a large variety of cells and specimens,  
its depth limitation problem makes it difficult to study thick samples. Even  
though microscope manufacturers try to provide objective lenses with longer  
working distances, diffraction and scattering of lights within the optically un-  
even specimen still causes practical limitations. Two-photon microscopy does  
615 not solve this problem; although long-wavelength infrared light of the excita-  
tion laser is less prone to diffraction and scattering, emission light of short-  
wavelength from the labeled signals suffers from the same optical problem as in  
single-photon confocal microscopy. A promising alternative solution to obtain a  
thick high-resolution image stack is to stitch multiple substacks that are taken  
620 either from both sides of the specimen by sample flipping or from the same  
side by cutting off the sample surface progressively with microtomes. This al-  
ternative solution raises the problem of registration and stitching of partially  
overlapping 3D stacks. In this article, we proposed 2D-SIFT-in-3D-Space as an  
efficient method to address the registration and stitching of a variety of 3D im-  
625 age stacks. It specially allows tight adjustment of stacks of thin structures such  
as neurons to avoid seams between stitched substacks, or tiles, reducing errors in  
further 3D image processing and analyses. The registration is realized with an  
iterative combination of existing 2D local features in the 3D space. 2D-SIFT-in-  
3D-Space algorithm is not limited to neuron images of front and back substacks;  
630 it can be used on a variety of different datasets that require precise registration  
(See Figs. S2, S3 of the Supplementary data for additional examples). Also  
to help advanced analyses of 3D image stack in laser-scanning microscopy and  
to improve SIFT registration of dark samples, we provide a tool that improves  
intensity and contrast of serial section image stacks to compensate the loss in-  
635 duced by photobleaching and attenuation of the fluorescence (see Figs. S4, S5  
of the Supplementary data). Quantitative validation of 2D-SIFT-in-3D-Space  
is realized with the help of an in silico virtual test data specially created for

this study to simulate the noise in fluorescent laser-scanning microscopy. Given the huge literature and software libraries dedicated to image registration, other programs, as discussed in the related work section, would in principle have the potential to also perform the registration problem illustrated here on *Drosophila* brain. Different image datasets possess diverse characteristic features that affect registration, and each registration method has to be tuned properly for each purpose for a fair comparison. A quantitative comparison with all registration solutions and exhaustive investigation of the whole parameter spaces of each of all these solutions would be an important milestone. Such an exhaustive work has recently been proposed for instance for particle tracking [45]. To target a similar milestone for 3D image registration of fluorescent microscopy, annotated image datasets with ground-truth registration solution would be required. Manual annotation and registration of such datasets would be very difficult and time consuming. The simulator proposed in our manuscript therefore opens the way to the quantitative benchmark of various registration methods of 3D images acquired with fluorescent microscopy.

Other interesting perspectives include evolution of our algorithm. We considered situations where it was possible to assume that the samples were not deforming during two imaging steps. In this framework, we logically considered rigid registration. In case where this assumption would not hold it would then be necessary to realize non rigid transformation. Also, the proposed algorithm allow to register multiple channels in fluorescence stack of images. However, the registration is based on one single channel (the best contrasted one), and the computed transformation is then applied to the other channels. SIFT can be applied to multiple component 3D images in principle and a joint registration of the channel could be considered as a further improvement.

### Information Sharing Statement

The described algorithm 2D-SIFT-in-3D-Space Volume Stitching is implemented using the Java programming language and provided as Open Source

plugin module of Fiji/ImageJ. It depends in Fiji on mpicbg package for SIFT and RANSAC and stitching package to fuse image stacks. The source code of the algorithm is available via Github (<https://github.com/rosahuaman/2D-SIFT-in-3D-Space-master>). The simulator of laser-scanning microscopy image stacks was performed using Matlab (Natick, Massachusetts). The script code and artificially generated datasets from simulation are available via <https://www.creatis.insa-lyon.fr/site7/fr/MicroTools>.

### Acknowledgements

We thank Hanchuan Peng and Henrik Skibbe for their comments and suggestions, and Haruka Hirose, Youko Ishida, Ryoko Tatsumi, Jun Tanimura, Tomoko Yano, Takaaki Miyazaki and Asako Tsubouchi for their technical and scientific assistance for sample data preparation. We thank Bloomington Drosophila Stock Centre for fly strains and Hybridoma Bank for antibodies. This study is supported by the Strategic Research Program for Brain Sciences and CREST Program by the Japan Science and Technology Agency (JST) / Japan Agency for Medical Research and Development (AMED) to KI.

### References

- [1] A. H. Brand, N. Perrimon, Targeted gene expression as a means of altering cell fates and generating dominant phenotypes, *development* 118 (2) (1993) 401–415.
- [2] K. M. Brown, G. Barrionuevo, A. J. Canty, V. De Paola, J. A. Hirsch, G. S. Jefferis, J. Lu, M. Snippe, I. Sugihara, G. A. Ascoli, The diadem data sets: representative light microscopy images of neuronal morphology to advance automation of digital reconstructions, *Neuroinformatics* 9 (2-3) (2011) 143–157.
- [3] C.-T. Shih, O. Sporns, S.-L. Yuan, T.-S. Su, Y.-J. Lin, C.-C. Chuang, T.-Y. Wang, C.-C. Lo, R. J. Greenspan, A.-S. Chiang, Connectomics-based anal-



- ysis of information flow in the *Drosophila* brain, *Current Biology* 25 (10)  
695 (2015) 1249–1258.
- [4] L. Fritzky, D. Lagunoff, *Advanced methods in fluorescence microscopy*,  
*Analytical Cellular Pathology* 36 (1-2) (2013) 5–17.
- [5] K. Ito, R. Okada, N. K. Tanaka, T. Awasaki, Cautionary observations  
on preparing and interpreting brain images using molecular biology-based  
700 staining techniques, *Microscopy research and technique* 62 (2) (2003) 170–  
186.
- [6] K. Chung, J. Wallace, S.-Y. Kim, S. Kalyanasundaram, A. S. Andalman,  
T. J. Davidson, J. J. Mirzabekov, K. A. Zalocusky, J. Mattis, A. K. Denisin,  
et al., Structural and molecular interrogation of intact biological systems,  
705 *Nature* 497 (7449) (2013) 332–337.
- [7] H. Hama, H. Kurokawa, H. Kawano, R. Ando, T. Shimogori, H. Noda,  
K. Fukami, A. Sakaue-Sawano, A. Miyawaki, Scale: a chemical approach  
for fluorescence imaging and reconstruction of transparent mouse brain,  
*Nature neuroscience* 14 (11) (2011) 1481–1488.
- 710 [8] B. W. Graf, S. A. Boppart, *Imaging and analysis of three-dimensional cell  
culture models*, *Live Cell Imaging: Methods and Protocols* (2010) 211–227.
- [9] T. Ragan, L. R. Kadiri, K. U. Venkataraju, K. Bahlmann, J. Sutin,  
J. Taranda, I. Arganda-Carreras, Y. Kim, H. S. Seung, P. Osten, Serial  
two-photon tomography for automated ex vivo mouse brain imaging, *Nature  
715 methods* 9 (3) (2012) 255–258.
- [10] T. Zheng, Z. Yang, A. Li, X. Lv, Z. Zhou, X. Wang, X. Qi, S. Li, Q. Luo,  
H. Gong, et al., Visualization of brain circuits using two-photon fluorescence  
micro-optical sectioning tomography, *Optics express* 21 (8) (2013) 9839–  
9850.
- 720 [11] A. Cardona, *Towards semi-automatic reconstruction of neural circuits*,  
*Neuroinformatics* 11 (1) (2013) 31–33.

- [12] E. Meijering, Neuron tracing in perspective, *Cytometry Part A* 77 (7) (2010) 693–704.
- [13] B. Zitova, J. Flusser, Image registration methods: a survey, *Image and vision computing* 21 (11) (2003) 977–1000.
- 725
- [14] I. N. Figueiredo, S. Moura, J. S. Neves, L. Pinto, S. Kumar, C. M. Oliveira, J. D. Ramos, Automated retina identification based on multiscale elastic registration, *Computers in Biology and Medicine* 79 (2016) 130–143.
- [15] M. Yang, H. Ding, L. Zhu, G. Wang, Ultrasound fusion image error correction using subject-specific liver motion model and automatic image registration, *Computers in Biology and Medicine* 79 (2016) 99–109.
- 730
- [16] Y. Song, H. Lee, H. C. Kang, J. Shin, G.-S. Hong, S. H. Park, J. Lee, Y.-G. Shin, Interactive registration between supine and prone scans in computed tomography colonography using band-height images, *Computers in Biology and Medicine* 80 (2017) 124–136.
- 735
- [17] J. Kim, J. Lee, J. W. Chung, Y.-G. Shin, Locally adaptive 2D–3D registration using vascular structure model for liver catheterization, *Computers in Biology and Medicine* 70 (2016) 119 – 130.
- [18] A. D. Savva, T. L. Economopoulos, G. K. Matsopoulos, Geometry-based vs. intensity-based medical image registration: A comparative study on 3D CT data, *Computers in Biology and Medicine* 69 (2016) 120–133.
- 740
- [19] T. Rohlfing, User guide to the computational morphometry toolkit, Neuroscience Program, SRI International, Menlo Park, CA (2011).
- [20] S. Preibisch, S. Saalfeld, P. Tomancak, Globally optimal stitching of tiled 3D microscopic image acquisitions, *Bioinformatics* 25 (11) (2009) 1463–1465.
- 745
- [21] M. N. Economo, N. G. Clack, L. D. Lavis, C. R. Gerfen, K. Svoboda, E. W. Myers, J. Chandrashekar, A platform for brain-wide imaging and reconstruction of individual neurons, *Elife* 5 (2016) e10566.

- 750 [22] S. Preibisch, S. Saalfeld, J. Schindelin, P. Tomancak, Software for bead-based registration of selective plane illumination microscopy data, *Nature methods* 7 (6) (2010) 418–419.
- [23] M. N. Economo, N. G. Clack, B. J. Arthur, C. Bruns, E. Bas, J. Chandrashekar, Registration and resampling of large-scale 3D mosaic images, in: *Proceedings of BioImage Informatics Conference*, 2015.
- 755 [24] J. Li, N. M. Allinson, A comprehensive review of current local features for computer vision, *Neurocomputing* 71 (10) (2008) 1771–1787.
- [25] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision* 60 (2) (2004) 91–110.
- 760 [26] P. Scovanner, S. Ali, M. Shah, A 3-dimensional SIFT descriptor and its application to action recognition, in: *Proceedings of the 15th ACM international conference on Multimedia*, ACM, 2007, pp. 357–360.
- [27] G. T. Flitton, T. P. Breckon, N. M. Bouallagu, Object recognition using 3D SIFT in complex ct volumes., in: *BMVC*, 2010, pp. 1–12.
- 765 [28] W. Cheung, G. Hamarneh, N-sift: N-dimensional scale invariant feature transform for matching medical images, in: *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on, IEEE*, 2007, pp. 720–723.
- 770 [29] C. Paganelli, M. Peroni, F. Pennati, G. Baroni, P. Summers, M. Bellomi, M. Riboldi, Scale invariant feature transform as feature tracking method in 4d imaging: a feasibility study, in: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE*, 2012, pp. 6543–6546.
- 775 [30] S. Allaire, J. J. Kim, S. L. Breen, D. A. Jaffray, V. Pekar, Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis, in: *Computer Vision and Pattern Recog-*

- 780 [31] D. Ni, Y. Qu, X. Yang, Y. P. Chui, T.-T. Wong, S. S. Ho, P. A. Heng, Volumetric ultrasound panorama based on 3d sift, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2008, pp. 52–60.
- [32] K. Ito, K. Shinomiya, M. Ito, J. D. Armstrong, G. Boyan, V. Hartenstein, S. Harzsch, M. Heisenberg, U. Homberg, A. Jenett, et al., A systematic nomenclature for the insect brain, *Neuron* 81 (4) (2014) 755–765.
- [33] S.-y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, et al., A visual motion detection circuit suggested by *Drosophila* connectomics, *Nature* 500 (7461) (2013) 175–181.
- 790 [34] S. Busch, M. Selcho, K. Ito, H. Tanimoto, A map of octopaminergic neurons in the *Drosophila* brain, *Journal of Comparative Neurology* 513 (6) (2009) 643–667.
- [35] K. E. White, D. M. Humphrey, F. Hirth, The dopaminergic system in the aging brain of *Drosophila*, *Frontiers in neuroscience* 4 (2010) 205.
- 795 [36] T. Lindeberg, Feature detection with automatic scale selection, *International journal of computer vision* 30 (2) (1998) 79–116.
- [37] S. Saalfeld, P. Tomancák, Automatic landmark correspondence detection for imagej, in: Proceedings of the ImageJ User and Developer Conference, 2008, pp. 128–133.
- 800 [38] S. Saalfeld, A. Cardona, V. Hartenstein, P. Tomancák, As-rigid-as-possible mosaicking and serial section registration of large sstem datasets, *Bioinformatics* 26 (12) (2010) i57–i63.

- [39] G. Herberich, R. Windoffer, R. Leube, T. Aach, Signal and noise modeling in confocal laser scanning fluorescence microscopy, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012* (2012) 381–388.
- 805
- [40] C. H. Li, C. Lee, Minimum cross entropy thresholding, *Pattern recognition* 26 (4) (1993) 617–625.
- [41] S. Preibisch, F. Amat, E. Stamataki, M. Sarov, R. H. Singer, E. Myers, P. Tomancak, Efficient bayesian-based multiview deconvolution, *nature methods* 11 (6) (2014) 645–648.
- 810
- [42] S. Saalfeld, R. Fetter, A. Cardona, P. Tomancak, Elastic volume reconstruction from series of ultra-thin microscopy sections, *Nature methods* 9 (7) (2012) 717–720.
- [43] C. J. Sheppard, M. Gu, Aberration compensation in confocal microscopy, *Applied optics* 30 (25) (1991) 3563–3568.
- 815
- [44] M. J. Booth, M. A. Neil, R. Juškaitis, T. Wilson, Adaptive aberration correction in a confocal microscope, *Proceedings of the National Academy of Sciences* 99 (9) (2002) 5788–5792.
- [45] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, et al., Objective comparison of particle tracking methods, *Nature methods* 11 (3) (2014) 281–289.
- 820