# As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets

Stephan Saalfeld[1], Albert Cardona[2], Volker Hartenstein[3] and Pavel Tomančák[1,*]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, [2]Institute of Neuroinformatics, University and ETH Zürich, Switzerland and [3]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095-1606, USA

## ABSTRACT

**Motivation:** Tiled serial section Transmission Electron Microscopy (ssTEM) is increasingly used to describe high-resolution anatomy of large biological specimens. In particular in neurobiology, TEM is indispensable for analysis of synaptic connectivity in the brain. Registration of ssTEM image mosaics has to recover the 3D continuity and geometrical properties of the specimen in presence of various distortions that are applied to the tissue during sectioning, staining and imaging. These include staining artifacts, mechanical deformation, missing sections and the fact that structures may appear dissimilar in consecutive sections.

**Results:** We developed a fully automatic, non-rigid but as-rigid-as-possible registration method for large tiled serial section microscopy stacks. We use the Scale Invariant Feature Transform (SIFT) to identify corresponding landmarks within and across sections and globally optimize the pose of all tiles in terms of least square displacement of these landmark correspondences. We evaluate the precision of the approach using an artificially generated dataset designed to mimic the properties of TEM data. We demonstrate the performance of our method by registering an ssTEM dataset of the first instar larval brain of *Drosophila melanogaster* consisting of 6885 images.

**Availability:** This method is implemented as part of the open source software TrakEM2 (http://www.ini.uzh.ch/~acardona/trakem2.html) and distributed through the Fiji project (http://pacific.mpi-cbg.de).

**Contact:** tomancak@mpi-cbg.de

## 1 INTRODUCTION

*Transmission electron microscopy* (TEM) offers currently the highest available resolution to investigate the structure of biological samples. The superior resolution of TEM is classically used to reveal subcellular structures such as organelles and some of the largest macromolecular assemblies inside cells. For anatomical studies on the tissue level, TEM finds most applications in neurobiology to reveal neuronal connectivity maps that underlie the function of the nervous system. The reconstruction of the complete neuronal connectivity map in *Caenorhabditis elegans* remains the pinnacle of these studies (White *et al.*, 1986). On the nanometer scales, brains are gigantic structures and many overlapping images must be assembled as a mosaic to cover a substantial portion of the tissue. Moreover, serial sectioning is required to capture the volume of the specimen (Fig. 1a–d).

Digital imaging and high-precision beam or specimen shifting equipment (Suloway *et al.*, 2005) simplifies the reconstruction of

mosaics within one section by providing a good initial estimate of the tile configuration (Fig. 1b). However, due to the minute distances involved the precision of the instruments is insufficient for seamless stitching of the tiles (Fig. 1c). Adjacent sections are rotated and translated arbitrarily with respect to each other (Fig. 1d). Physical sectioning, the subsequent manipulation of the sections by contrast gaining staining procedures and imaging with the electron beam introduce significant distortions and artifacts into the captured images (Fig. 1e–h). The goal in reconstruction of serial section TEM (ssTEM) data is to identify the configuration of all tiles that best preserves both the continuity and geometry of 3D structures in the specimen making it amenable for subsequent segmentation and analysis. Particularly in neurobiology, it is important to preserve the continuity of axons across large distances and the ability to distinguish and localize synaptic connections between neurons.

Anderson *et al.* (2009) propose a registration pipeline for large ssTEM datasets based on correlation of image intensities within and across sections. They are able to identify overlapping tiles in unordered sets automatically, mosaic them by propagating pairwise translation and refine this initial layout by warping each tile to its preceding partners. The resulting section mosaics are registered relative to each other by searching a rigid transformation initially and subsequently warping one section to the other. This solution guarantees continuity of 3D structures by maximizing the overlap of similar image content in adjacent sections. On the other hand, it cannot preserve geometrical properties because the unavoidable registration error is accumulated with each consecutive registration step.

Instead of using all image intensities, it is possible to register two or more images using corresponding salient points as a statistic about the image content as proposed by Brown and Lowe (2003) for panorama stitching. Reducing the problem from full resolution image content to a relatively small number of corresponding points simplifies the estimation of the globally optimal configuration for a large number of tiles.

State of the art techniques (Bay *et al.*, 2008; Lowe, 2004; Matas *et al.*, 2002; Mikolajczyk and Schmid, 2004; Tuytelaars and Van Gool, 2004) allow both automatic detection of interest points and extraction of affine invariant local descriptors for these points. Affine invariant matching becomes nearest neighbor search in a local feature descriptor space. It was shown by Mikolajczyk and Schmid (2003) that the local descriptor as used in the *Scale Invariant Feature Transform* (SIFT; Lowe, 2004) outperforms competing techniques in both distinctiveness and robustness with respect to significant image deformation and other disturbing artifacts.

In this article, we present a fully automatic registration method for large ssTEM image mosaics that globally minimizes the

---

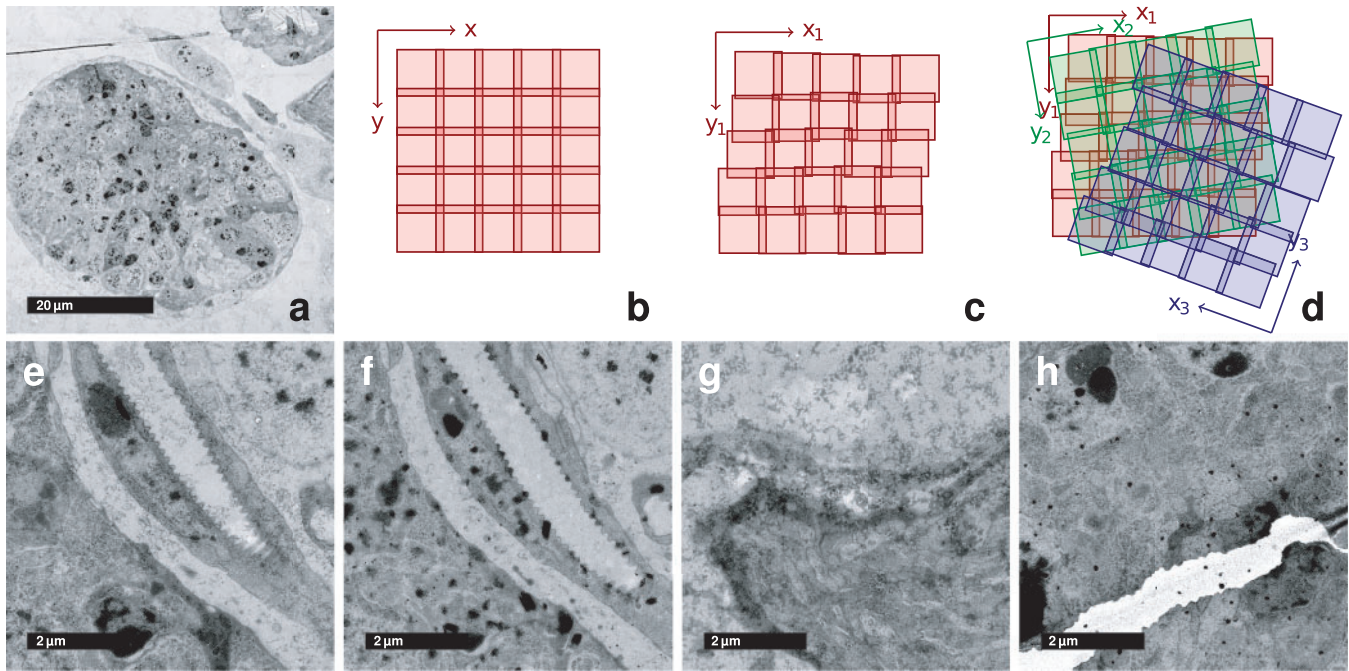*To whom correspondence should be addressed.

**Fig. 1.** Properties of ssTEM image mosaics. (**a**) Low magnification overview of a single section mosaic consisting of $9 \times 9$ registered tiles. (**b**) Each section is imaged as a sequence of overlapping tiles that is assumed to be a regular grid. (**c**) Each tile's true pose in the local section coordinate frame is affected by odometry errors of the microscope, that are propagated over consecutive tiles. (**d**) The relative pose of two consecutive sections is arbitrarily shifted and rotated. Clean sections are relatively rare (**e**); typically sections contain artifacts such as dirt (**f**), staining precipitate (**g**) and folds or cracks (**h**).

registration error. We identify overlapping images by corresponding image content within and across sections using SIFT features and a geometric consistency constraint. We approximate the non-linear transformation between adjacent sections by an independent rigid-body transformation (rotation and translation) for each image tile of the section mosaics. In a TEM section series, none of the sections can be considered non-deformed and thus, no section can serve as a template for registration. Therefore, we define the registration problem to be template free and, instead, explicitly minimize the non-rigid deformation applied to all sections. This is achieved by globally minimizing the square displacement of all corresponding SIFT landmarks within and across sections. By this means, the resulting tile configuration is non-rigid but *as-rigid-as-possible*. In order to evaluate the performance of this approach, we developed a synthetic ground truth dataset that is designed to mimic the properties of ssTEM data. We were able to register 6885 images from 85 serial sections through the *Drosophila* first instar larval brain; each section consisting of $9 \times 9$ tiles of $2048 \times 2048$ px. The registered dataset serves as a starting point for characterizing the fine architecture of this large brain at unprecedented resolution.

## 2 METHODS

### 2.1 Definition of the task

Specimens prepared for serial sectioning are typically embedded in a block of rigid medium such as resin, plastic or ice. Using a *microtome*, ultrathin consecutive slices are cut from the rigidly mounted block. The slices float on the surface of a liquid in the 'knife boat' and are manually picked up onto a TEM grid. While the section thickness may vary slightly within and across sections, in this work, we consider sections to be planar and of constant

thickness. In reality, the cross-section variation can be minimized by careful operation of the microtome, and, as both the specimen itself and the knife of the microtome are rigidly mounted, intrasection variation is compensated throughout the section series by complementary variation in other sections.

Each section is imaged as a set of overlapping tiles. For each tile $t$, the microscope registers a 2D translation vector $\mathbf{t}_t = (u, v)^\mathsf{T}$ in a relative section reference frame $\mathbb{R}_s^2$, whereas the section index $s$ is given by the human operator. These coordinates are inaccurate due to measurement errors $\theta_t = (u, v)^\mathsf{T}$ (Fig. 1b and c). A section's pose relative to its adjacent sections is an unknown rigid transformation $\mathbf{R}_s$ (Fig. 1d). Moreover, sectioning, preparation after sectioning and even the imaging process itself induce noticeable deformations of both sections as a whole ($\mathbf{D}_s$) as well as on the scale of single tiles ($\mathbf{D}_t$) (Fig. 1a and e–h).

Therefore, the task is to reconstruct a 3D image volume from serial sections that have to be reconstructed from single tiles while compensating for noticeable deformation within and across sections. Particularly, the deformation–compensation poses the question how to invert the largely unknown transforming process that mapped a 3D coordinate $(x, y, z)^\mathsf{T}$ in scene space $\mathbb{R}_S^3$ into a 2D coordinate $(u, v)^\mathsf{T}$ in tile space $\mathbb{R}_t^2$. Assuming planar sections of constant thickness as mentioned above, we attach scene and section spaces by a permutation $\mathbf{P}_{zs}$ between the scene dimension $z$ and section index $s$. Now this mapping shows as

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{R}_s \mathbf{D}_t \mathbf{D}_s \mathbf{P}_{zs} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbf{t}_t + \theta_t \tag{1}$$

with $\mathbf{R}_s \mathbf{D}_t \mathbf{D}_s$ and $(x, y, z)^\mathsf{T}$ being unknown. Without precise knowledge about the deforming process $\mathbf{D}_t \mathbf{D}_s$, the appropriate solution is that with minimal non-rigid deformation for all sections, a model being as-rigid-as-possible. This coincides with the researcher's request not to introduce unmotivated artificial deformation to the image data that would adulterate
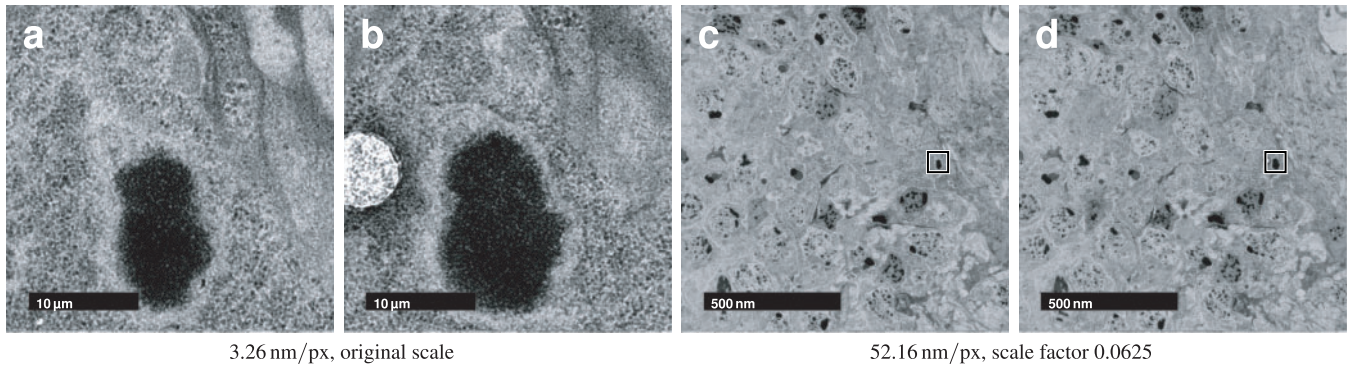
3.26 nm/px, original scale                    52.16 nm/px, scale factor 0.0625

**Fig. 2.** Similarity of serial sections depends on the scale of observation. Two consecutive 60 nm sections from the registered *Drosophila* first instar larval brain dataset showing a part of the cortex and the neuropile. The same region is shown with a resolution of 3.26 nm/px (**a,b**) and 52.16 nm/px (**c,d**). It is clearly visible, that the latter, having approximately isotropic resolution, shows significant similarity across sections while it is hard to identify corresponding pixels in the higher resolution example.

his observations. Declaring a tile being the unit of rigidity, we approximate $R_sD_tD_s+t_t$ by a rigid-per-tile transformation $R_t$ such that

$$\begin{pmatrix} u \\ v \end{pmatrix} = R_t P_{zs} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \omega_t + \theta_t \tag{2}$$

with $\omega_t$ being the transformation error introduced by the approximation. Maximal consistency relative to the scene is guaranteed by minimizing the remaining error $\epsilon_t = \omega_t + \theta_t$.

Without knowing the scene, it is impossible to estimate $\epsilon_t$ directly. However, consistency with respect to the scene implies consistency relative to overlapping tiles. Let $T$ be the set of all tiles $t \in T$ and $R$ be the set of all rigid transformations $R_t \in R$, then the best configuration $R$ is that minimizing the sum of all relative transfer errors of pairwise overlapping tiles:

$$\underset{R}{\arg\min} \sum_{t \in T} \left( \sum_{o \in T \setminus \{t\}} \epsilon_{to} \right) \tag{3}$$

Using *landmark correspondences* as a statistic about corresponding image content, $\epsilon_{to}$ is the sum of all square correspondence point displacements. Let $C_{to}$ be a set of landmark correspondences $(\mathbf{x},\mathbf{y})$ between two tiles $t \in T$ and $o \in T \setminus \{t\}$ with $\mathbf{x} \in t$ and $\mathbf{y} \in o$, then the best configuration $R$ is that minimizing the sum of all square correspondence displacements:

$$\underset{R}{\arg\min} \sum_{t \in T \setminus \{f\}} \left( \sum_{o \in T \setminus \{t\}} \left( \sum_{(\mathbf{x},\mathbf{y}) \in C_{to}} \|R_t \mathbf{x} - R_o \mathbf{y}\|^2 \right) \right) \tag{4}$$

with $f \in T$ being a fixed tile that defines the scene reference frame. All tiles must represent one single interconnected graph being pairwise connected by sets of landmark correspondences.

This optimization task requires a reliable landmark correspondence detector. Within a section, the appropriate interest point detector needs to be covariant to translation and robust against small amounts of non-rigid deformation as the overlapping parts of two consecutive tiles show the projection of the same tissue slightly deformed by heat during imaging. For cross-section matching, the appropriate interest point detector needs covariance to rotation as well as an automatism to detect only structures that appear similar, since two adjacent tiles show different portions of the tissue and are related by a rigid transformation and some plastic deformation. In ssTEM data, depending on section thickness, corresponding locations can be identified by local appearance of large enough structures sectioned perpendicularly. Such structures appear on different scales (for instance,

mitochondria 500 nm, nuclei 2 μm and neuropile compartments 40 μm) and thus detecting structures with large $(x,y)^{\mathsf{T}}$-scale raises the chance to detect the same structures in adjacent sections (Fig. 2).

For its scale invariant nature and robustness, we decided to use SIFT (Lowe, 2004) for both intra- and intersection landmark correspondence detection.

## 2.2 SIFT

SIFT (Lowe, 2004) detects blobs of arbitrary size as interest points using the *Difference of Gaussian* detector, estimates one or more dominant orientations for each detection and extracts a scale and orientation invariant local descriptor for each. Such blob-like structures are common in most textured images including TEM data. Typically, we detect several hundreds to thousands of interest points in a TEM micrograph of $2000 \times 2000$ px image size.

The detection's size, orientation and 2D location define a local coordinate frame for extraction of an invariant descriptor. We found empirically, that a larger descriptor built from $8 \times 8$ local histograms with eight bins each (512 dimensions) provides substantially better matching results for our data than the originally proposed $4 \times 4 \times 8$ descriptor. The smaller descriptors were proposed for natural images where viewpoint change and occlusion play a significant role. In TEM data, where two regions are related only by a rigid transformation with some deformation, a larger descriptor region necessarily increases the distinctiveness (Fig. 3).

Interest point correspondence candidates are identified by nearest neighbor matching in the local descriptor space. As suggested by Lowe (2004), good candidates are those that are significantly better than the next nearest neighbor. Significantly better means that the ratio between the Euclidean distances to the nearest and the next nearest neighbor is smaller than a given threshold. Empirically, we identified a threshold of 0.92 for TEM images to yield a reasonable number of correspondence candidates. Since real-time performance is not required, we identify the exact nearest neighbor by exhaustive search.

Local descriptor matching results in a significant number of false correspondences. We use the *Random Sample Consensus* (RANSAC) (Fischler and Bolles, 1981) to separate true correspondences that behave consistently with respect to a rigid transformation up to a maximal correspondence displacement $\epsilon_{\max}$ (see Algorithm 1). For estimation of the optimal rigid transformation by means of least square correspondence displacement, we use the closed form solution described by Schaefer *et al.* (2006). Significant deformation requires $\epsilon_{\max}$ to be relatively tolerant thus accepting some false correspondences. We filter these with a robust regression scheme that iteratively removes correspondences with
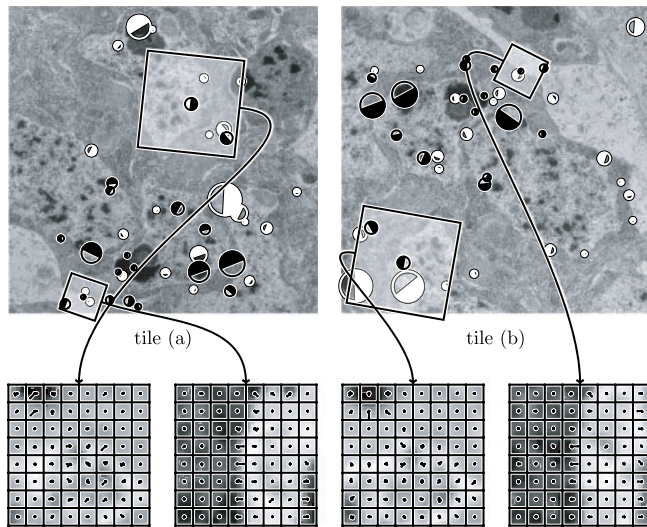
**Fig. 3.** SIFT-feature correspondences in two overlapping tiles from adjacent sections. 1003 feature candidates were extracted in tile (**a**), 971 in tile (**b**). 41 correspondence pairs were identified by local feature descriptor matching, 19 of them are true matches consistent to common transformation model. False matches are displayed in white, true matches in black. The size of the circles is proportional to the features scale, the filled part visualizes a feature's orientation. Two true correspondence pairs are selected as an example for the local SIFT-descriptor. The values of the local histogram bins are shown as combs on top of the local region the descriptor is extracted from.

---

**Algorithm 1** RANSAC

**Input:**    Set of data points $C$
Model $\mathbf{M}$
Maximal number of iterations $k$
Maximal allowed error $\epsilon_{max}$
Minimal required number of inliers $n_{min}$
**Output:** Set of inlier data points $\hat{C}_+$
Model $\hat{\mathbf{M}}$ best fitting to $\hat{C}_+$

1: $\hat{C}_+ \leftarrow \emptyset$                  ▷ Initialize best set of inliers.
2: **for** $k$ iterations **do**
3:     $C_{min} \leftarrow$ choose random minimal subset of $C$ to solve $\mathbf{M}$
4:     $\mathbf{M} \leftarrow$ fit to $C_{min}$
5:     $C_+ \leftarrow \emptyset$                  ▷ Initialize set of inliers.
6:     **for all** $c \in C$ **do**
7:         **if** $\mathbf{M}(c) < \epsilon_{max}$ **then**
8:             $C_+ \leftarrow C_+ \cup \{c\}$
9:         **end if**
10:     **end for**
11:     **if** $(|C_+| \geq n_{min}) \wedge (|C_+| > |\hat{C}_+|)$ **then**
12:         $\hat{C}_+ \leftarrow C_+$
13:     **end if**
14: **end for**
15: **if** $|\hat{C}_+| \geq n_{min}$ **then**
16:     $\hat{\mathbf{M}} \leftarrow$ fit to $\hat{C}_+$        ▷ Refine $\hat{\mathbf{M}}$ with respect to all inliers.
17: **end if**

---

**Algorithm 2** Iterative square displacement minimization

**Input:**    Set of tiles $T$
Set of fixed tiles $T_f \subset T$
Set of landmark correspondences $C$
**Output:** Optimal configuration $R = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{|T|}\}$

1: $\bar{\epsilon}_R \leftarrow \infty$                  ▷ Initialize the average displacement.
2: **repeat**
3:     **for all** $t \in T \setminus T_f$ **do**
4:         $\mathbf{R} \leftarrow$ fit rigid transformation to $C_t$
5:         **for all** $(\mathbf{x}, \mathbf{y}) \in C_t$ **do**
6:             $\mathbf{x} \leftarrow \mathbf{Rx}$                  ▷ Update landmarks in $C_t$.
7:         **end for**
8:         $\mathbf{R}_t \leftarrow \mathbf{R}\mathbf{R}_t$
9:     **end for**
10:     $\bar{\epsilon}_R^2 \leftarrow \frac{1}{|C|} \sum_{(\mathbf{x},\mathbf{y}) \in C} \|\mathbf{y} - \mathbf{x}\|^2$
11: **until** $\bar{\epsilon}_R$ converges

---

a displacement larger than $3\times$ the median displacement. For details, see Saalfeld and Tomančák (2008).

### 2.3 Global optimization of the tile configuration

Having the set of true landmark correspondences identified successfully, the optimal configuration $R$ of all tiles has to be estimated. As shown in Equation (4), this is the configuration with minimal square correspondence displacements.

We minimize this term using an iterative optimization scheme (see Algorithm 2). In case that prior knowledge about the gross configuration of all tiles is available, the minimization is initialized with this configuration. Otherwise all tiles start from $\mathbf{R}_t = \mathbf{I}$. In each iteration $i$, the optimal $\mathbf{R}_t$ for each single tile $t \in T \setminus \{f\}$ relative to the current configuration is estimated using the closed form least square solution by Schaefer *et al.* (2006) and applied to all landmark coordinates in this tile. The scheme terminates on convergence of the average correspondence displacement $\bar{\epsilon}_{Ri}$ that is estimated after each iteration. As convergence criteria, we require $\bar{\epsilon}_{Ri}$ to have a low absolute slope $|\nabla_h \bar{\epsilon}_{Ri}|$ and an absolute value below some threshold given by the user. It is crucial to prevent being trapped in local minima or at long plateaus. We address this effectively by requiring $|\nabla_h \bar{\epsilon}_{Ri}|$ to be very low for all $h$ from a set of decreasing lengths. The maximal length $h$ of a plateau or local valley, the maximal total number of iterations and the maximal accepted remaining error are user-defined parameters with empirically estimated defaults suggested.

### 2.4 Notes on implementation

The described registration algorithm including SIFT, RANSAC, closed form least square error solutions for a variety of transformation models, the robust regression scheme and the iterative optimizer are implemented using the **?** programming language and provided as an Open Source library through the ImageJ distribution Fiji (Schindelin, 2008). The registration procedure is included in the software toolkit TrakEM2 (Cardona, 2006) for management, registration and analysis of massive serial section microscopy datasets and, by that, in daily use in a large number of laboratories allover the world.

## 3 RESULTS

### 3.1 Overview of the registration algorithm

Principally, the registration procedure consists of extracting landmarks from all tiles, identifying landmark correspondences between tile pairs and estimating the tile configuration that minimizes the sum of all square correspondence displacements as shown in Algorithm 3.
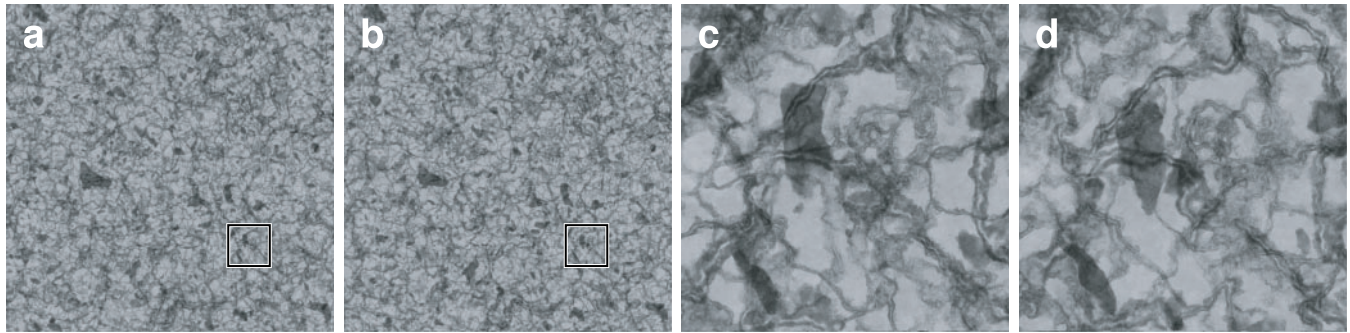
**Fig. 4.** Artificially generated evaluation dataset. The dataset simulates thin transilluminated volumes of 20 px thickness with membrane- and blob-like structures at various scales. Structures are defined by volumetric density functions. Locations with higher density scatter more 'light' than those with lower density and appear darker in the shadow projection. Two adjacent sections are shown to visualize the cross-section change of visible structures. In (**a,b**), the full dataset (4096 × 4096 px) is shown at low magnification, an area of 512 × 512 px is marked and shown in (**c,d**) at high magnification.

---

**Algorithm 3** Register TEM-dataset: principal procedure

---

**Input:**   Set of tiles $T$
**Output:** Tile configuration $R$

$\qquad\qquad\qquad\qquad$ ▷ Extract interest points from all tiles.
1:  **for all** $t \in T$ **do**
2:  $\qquad F_t \leftarrow$ extract interest points
3:  **end for**

$\qquad\qquad$ ▷ Identify interest point correspondences for pairs of tiles.
4:  $C$ : empty set of correspondences
5:  **for all** $t \in T$ **do**
6:  $\qquad C_t$ : empty set of correspondences
7:  $\qquad$ **for all** $o \in T \setminus \{t\}$ **do**
8:  $\qquad\qquad C_{to} \leftarrow$ identify correspondence pairs from $F_t$ and $F_o$
9:  $\qquad\qquad C_t \leftarrow C_t \cup C_{to}$
10: $\qquad$ **end for**
11: $\qquad C \leftarrow C \cup C_t$
12: **end for**

$\qquad\qquad\qquad\qquad$ ▷ Estimate the optimal tile configuration $R$.
13: $R \leftarrow \arg\min \epsilon_R$

---

Practically, since adjacent sections are related by a rigid transformation $\mathbf{R}_s$ plus some non-rigid deformation $\mathbf{D}_s$ that is to be compensated, we would like to use the transformation $\mathbf{R}_s$ as initialization for the sought after configuration $R$ and as a hint which tiles to exclude from pairwise feature comparison. Unfortunately, single sections do not always represent a single landmark-interconnected graph of tiles, while the whole volume eventually does. This is a typical scenario when imaging tree-like structures and the operator following only the branches ignoring the rest of the volume. That is, instead of the section as a whole, all graphs of a section have to be tested against all graphs present in the previous section. Two graphs that can be registered to each other are then joined into a single one by matching and filtering the features of overlapping tiles. Eventually, this will result in one single graph representing the whole volume. After joining two graphs, the configuration of the resulting graph is optimized thus sequentially building up the optimal global configuration from optimal intermediate configurations.

We applied this registration approach to test sets of TEM data of the *Drosophila* first instar larval brain sectioned into 60 nm sections imaged at 3.26 nm/px resolution. We developed a visualization of the global optimization progress where for each iteration the corresponding landmarks and the residual transfer error between them are highlighted by green dots and red lines, respectively. Movie 1 available as supplementary on-line material shows the progress of the optimization for a single section dataset consisting of 6 × 6 tiles.[1] It highlights the ability of the method to correctly place even tiles that contain mostly background (lower left corner tile) and gain very little SIFT feature detections, as long as these are attached to the graph. If a tile lacks SIFT correspondences altogether, it represents an independent graph on its own. The user can choose to drop or hide all but the largest graph from the dataset. Movie 2 shows the progress of the optimization on three serial sections imaged as 4 × 4 tile mosaics demonstrating that the procedure for multiple section is essentially the same as for a single section.[2]

## 3.2 Quantitative assessment of registration results on synthetic evaluation data

We created an artificial dataset with the ray-tracing program POV-ray (Cason *et al.*, 2007) that allows to define the interior of a volume as a density pattern based on an arbitrary volumetric function, where high density scatters more light than low density (Fig. 4).[3] This simulates the imaging process in the Transmission Electron Microscope where structures with a higher density of heavy-metal atoms scatter more electrons than structures with lower density and appear darker at the screen. The interior of the evaluation dataset contains filamentous and 'blob'-like structures, both over three scale octaves with multiscale turbulence. In this way, it mimics typical structures in the stained tissue, like membranes, nucleoli and mitochondria.

To emulate the ssTEM data we generated 16 projections of 4096 × 4096 px each through consecutive volumes of 20 px

---

[1]Movie 1: http://fly.mpi-cbg.de/saalfeld/36.avi
[2]Movie 2: http://fly.mpi-cbg.de/saalfeld/48.avi
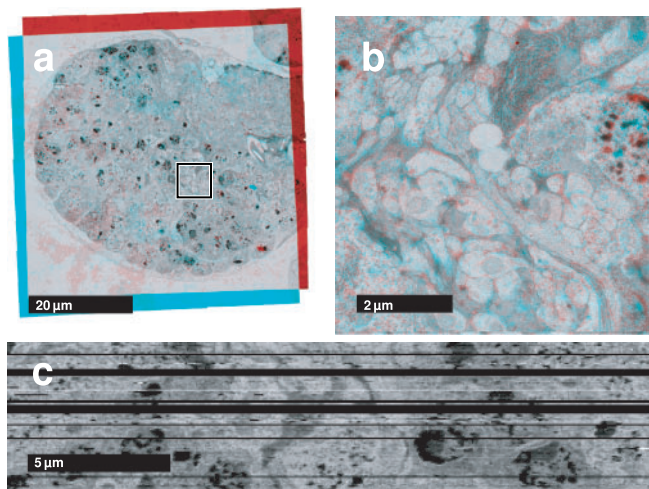[3]Evaluation-dataset: http://fly.mpi-cbg.de/saalfeld/tem-evaluation

**Fig. 5.** *Drosophila* first instar larval brain ssTEM dataset. Two consecutive registered sections from the dataset as red-cyan color merge. The diameter of the brain is ∼60 μm. (**a**) The section mosaics as a whole with the areas zoomed in (**b**) marked. (**c**) A single perpendicular section through the entire registered volume. Several sections were lost during sectioning and collecting onto the electron microscopy grid and shown here as black rows.

thickness, which is roughly equivalent to the ratio of $x, y$ versus $z$ resolution in typical TEM data. We split the synthetic ground truth data into $1024 \times 1024$ px tiles with 102 px overlap resulting in 25 images per section. We ran our registration program on the resulting 400 images as if there was nothing known about the tiles poses other than the section index. The registration results in a stack that is anchored to a randomly selected tile in the world reference frame. Thus, missing comparable world coordinates, we evaluate the quality of the registration by estimating the coordinate translation that gives the minimal average displacement when applied to the registered stack. This identifies the world reference frame of the registered stack and gives an average registration error of a pixel relative to the original scene.

We select a random sample of 1000 locations in each tile and transfer all locations into the world reference frame using the true and the registered transformation model of the tile. The translation between registration and ground truth domain is that between the centroids of both point clouds. The average residual transfer error serves as a quantitative measure of the registration process. We estimated an average displacement $\bar{\epsilon} = 4.14$ px, SD $\sigma_\epsilon = 3.63$ px and maximal displacement $\epsilon_{max} = 15.71$ px for the registration result. These results demonstrate that, with the proposed method, we can both identify overlap in an unknown configuration of images and register them.

### 3.3 Registration of the *Drosophila* first instar larval brain ssTEM dataset

As an example of real biological data, we registered the *Drosophila* first instar larval brain ssTEM dataset consisting of 85 sections of 60 nm thickness covering lateral neuronal layers and part of the neuropile of the left hemisphere. Each section was imaged with TEM using a moving stage operated by the Leginon software as $9 \times 9$ tiles overlapping by ∼6%. Tiles have a size of $2048 \times 2048$ px and a resolution of 4.68 nm/px. All 6885 tiles were registered fully

automatically. Intrasection configurations were initialized with the odometry data of the microscope. Correspondence estimation using SIFT descriptors and robust geometric consensus filters performed very well even in presence of significant changes in illumination and sharpness, dirty sections and significant gaps of up to six sections in the stack. Visual inspection of the registered data shows reliable continuity of biologically relevant structures such as axon bundles within and across sections both at low and high scales (Fig. 5a and b).[4] Moreover, when the registered dataset is cut perpendicularly, the resulting image, whose lateral resolution is limited by section thickness, resembles electron microscopy data without major discontinuities suggesting that the registration was successful (Fig. 5c).[5] For visualization and collaborative annotation, we present the registered dataset on-line through the CATMAID web interface (Saalfeld *et al.*, 2009).[6]

## 4 DISCUSSION

We presented a fully automated method for registration of large ssTEM image mosaics. The method identifies the optimal configuration of image tiles regardless whether or not an initial guess of the configuration is available. It is capable of correctly placing tiles with only minimal image content provided that this content is connected to the rest of the tile graph. Disconnected graphs of tiles are registered independently. Thus, the approach is ideally suited for alignment of ssTEM data generated either manually or using a robotic setup.

We use SIFT to identify corresponding landmarks and use them as a statistical sample for similar image content in the TEM images. The scale invariance of SIFT is well suited to capture the changing appearance of small structures at low scale levels and the presence of meaningful similar features at multiple scales. Larger, more distinctive local descriptors together with the requirement that correspondence candidates must be geometrically consistent yields reliably large sets of true matches.

The reduction of the registration task to a compact set of corresponding landmarks enables global minimization of square correspondence displacements for very large tile systems. Registering the 6885 tiles *Drosophila* larval brain dataset took ∼24 h on an Intel® Xeon® computer with two 2.66 GHz dual-core-CPUs and 8 GB of RAM, with the most time-consuming operation in the procedure being the exhaustive nearest neighbor search in the 512 D local descriptor space. Principally, there is no upper limit to the size of the dataset other than the available storage space.

The described methods for robust outlier removal and global optimization are applicable to a wide variety of microscopy image mosaicking tasks. We used it successfully to register light microscopy image mosaics including sets of overlapping 3D confocal image stacks (Preibisch *et al.*, 2009b) and applied it to bead-based registration of Selective Plane Illumination Microscopy (SPIM; Huisken *et al.*, 2004) datasets consisting of 3D stacks of the same specimen taken from different angles (Preibisch *et al.*, 2009a). In Preibisch *et al.* (2009b), we identify the pairwise 3D translation between overlapping confocal image stacks by normalized cross-correlation and globally minimize the translational offset of large

---

[4]Section series as a movie: http://fly.mpi-cbg.de/saalfeld/series.avi
[5]Resliced series as a movie: http://fly.mpi-cbg.de/saalfeld/resliced.avi
[6]Registered dataset on-line: http://fly.mpi-cbg.de/first-instar-brain

sets of overlapping stacks. In Preibisch *et al.* (2009a), we proposed the usage of fluorescent beads embedded in the rigid mounting medium as fiduciary markers. We use the local constellations of neighboring beads to automatically identify correspondences invariantly to 3D rotation and scale. The model for the geometric consistency constraint and global minimization of the square transfer errors is a 3D affine transformation with one of the 3D stacks serving as a template.

We generated a unique evaluation dataset designed to resemble the appearance of biological tissue in ssTEM data for quantitative comparison of the registration results. The evaluation framework can be used to assess the performance of any registration scheme that records its transformation model and could become the standard for objectively measuring the performance of various registration approaches. In future work, we will induce artificial deformation and variance in section thickness to the dataset in order to examine the limits of our registration method. As an alternative more realistic ground truth evaluation dataset, we propose to use *Serial Block-Face Scanning Electron Microscopy* data (Denk and Horstmann, 2004) where consecutive sections are aligned per definition.

The presented registration method is non-rigid but as-rigid-as-possible with a tile being the unit of rigidity for intrasection alignment. This delivers an ssTEM image dataset amenable to *quantitative* analysis, such as double dissector estimation of synaptic density (Geinisman *et al.*, 1996), which require volumes to be as reliable as possible. On the other hand, rigid-per-tile registration cannot compensate large-scale deformation by tile displacement without introducing noticeable discontinuities at the tile borders. The solution to this problem lies in extending the approach to include arbitrary non-rigid deformation at scales below the tile level preserving the regularization in terms of local rigidity. That means deforming all images minimally. We are currently exploring various ideas of how to express this regularization for a completely non-rigid global registration. The evaluation dataset will be instrumental in assessing the performance of such approaches with respect to reflecting faithfully the ground truth data.

The globally optimal reconstruction of entire brains on TEM level will enable registration of 3D light microscopy data onto electron microscopy volumes. By that it will be possible to establish the connection between brain macro (neuronal lineages) and micro (synaptic connectivity) circuitry (Cardona *et al.*, 2009).

## ACKNOWLEDGEMENTS

## REFERENCES

Anderson,J.R. *et al.* (2009) A computational framework for ultrastructural mapping of neural circuitry. *PLoS Biol.*, **7**, e1000074.
Bay,H. *et al.* (2008) SURF: Speeded up robust features. *Comput. Vis. Image Understand.*, **110**, 346–359.
Brown,M. and Lowe,D.G. (2003) Recognising panoramas. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, IEEE Computer Society. Washington, DC, pp. 1218–1225.
Cardona,A. (2006) TrakEM2: an ImageJ-based program for morphological data mining and 3D modeling. In *Proceedings of the 1st ImageJ User and Developer Conference*, Luxembourg, May 18–19.
Cardona,A. *et al.* (2009) Drosophila brain development: Closing the gap between a macroarchitectural and microarchitectural approach. *Cold Spring Harb. Symp. Quant. Biol.* [Epub ahead of print, doi:10.1101/sqb.2009.74.037].
Cason,C. *et al.* (2003–2007) POV-ray – the persistance of vision raytracer [release version 3.6.1]. Availabe at http://www.povray.org. (last accessed date August 8, 2007).
Denk,W. and Horstmann,H. (2004) Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.*, **2**, 1900–1909.
Fischler,M.A. and Bolles,R.C. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, **24**, 381–395.
Geinisman,Y. *et al.* (1996) Unbiased stereological estimation of the total number of synapses in a brain region. *J. Neurocytol.*, **25**, 805–819.
Huisken,J. *et al.* (2004) Optical sectioning deep inside live embryos by Selective Plane Illumination Microscopy. *Science*, **305**, 1007–1009.
Lowe,D.G. (2004) Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, **60**, 91–110.
Matas,J. *et al.* (2002) Robust wide baseline stereo from maximally stable extremal regions. In Rosin,P.L. and Marshall,D. (ed), *Proceedings of the British Machine Vision Conference*, Vol. 1, BMVA, London, pp. 384–393.
Mikolajczyk,K. and Schmid,C. (2003) A performance evaluation of local descriptors. In *International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 257–263.
Mikolajczyk,K. and Schmid,C. (2004) Scale and affine invariant interest point detectors. *Int. J. Comput. Vis.*, **60**, 63–86.
Preibisch,S. *et al.* (2009a) Bead-based mosaicing of single plane illumination microscopy images using geometric local descriptor matching. In Pluim,J.P.W. and Dawant,B.M. (ed), *Proceedings of The International Society for Optical Engineering*, Vol. 7259.
Preibisch,S. *et al.* (2009b) Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics*, **25**, 1463–1465.
Rasband,W. (1997–2009) ImageJ: Image processing and analysis in Java [version 1.43l], Available at http://rsb.info.nih.gov/ij/ (last accessed date November 24, 2009).
Saalfeld,S. and Tomančák,P. (2008) Automatic landmark correspondence detection for ImageJ. In *Proceedings of the ImageJ User and Developer Conference*, Luxembourg, pp. 128–133.
Saalfeld,S. *et al.* (2009) CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics*, **25**, 1984–1986.
Schaefer,S. *et al.* (2006) Image deformation using moving least squares. *ACM Trans. on Graph.*, **25**, 533–540.
Schindelin,J.E. (2008) Fiji is just ImageJ—batteries included. In *Proceedings of the ImageJ User and Developer Conference*, Luxembourg, pp. 99–104.
Suloway,C. *et al.* (2005) Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.*, **151**, 41–60.
Tuytelaars,T. and Van Gool,L. (2004) Matching widely separated views based on affine invariant regions. *Int. J. Comput. Vis.*, **59**, 61–85.
White,J.G. *et al.* (1986) The structure of the nervous system of the nematode Caenorhabditis elegans. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.*, **314**, 1–340.