Mosaicing of Single Plane Illumination Microscopy Images Using Groupwise Registration and Fast Content-Based Image Fusion

Stephan Preibisch[a], Torsten Rohlfing[b], Michael P. Hasak[b], and Pavel Tomancak[a]

[a]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[b]SRI International, Neuroscience Program, Menlo Park, CA 94025, USA

ABSTRACT

Single Plane Illumination Microscopy (SPIM; Huisken *et al.*, Nature 305(5686):1007–1009, 2004) is an emerging microscopic technique that enables live imaging of large biological specimens in their entirety. By imaging the living biological sample from multiple angles SPIM has the potential to achieve isotropic resolution throughout even relatively large biological specimens. For every angle, however, only a relatively shallow section of the specimen is imaged with high resolution, whereas deeper regions appear increasingly blurred. In order to produce a single, uniformly high resolution image, we propose here an image mosaicing algorithm that combines state of the art groupwise image registration for alignment with content-based image fusion to prevent degrading of the fused image due to regional blurring of the input images. For the registration stage, we introduce an application-specific groupwise transformation model that incorporates per-image as well as groupwise transformation parameters. We also propose a new fusion algorithm based on Gaussian filters, which is substantially faster than fusion based on local image entropy. We demonstrate the performance of our mosaicing method on data acquired from living embryos of the fruit fly, *Drosophila*, using four and eight angle acquisitions.

Keywords: Groupwise image registration, single plane illumination microscopy, image restoration and enhancement, three-dimensional mosaicing

1. INTRODUCTION

1.1 Biological Background

Microscopic imaging of developing biological samples, such as embryos, with sufficient spatial resolution to distinguish individual cells throughout the sample volume and with enough temporal resolution to follow morphogenetic events occurring during development is important for understanding of developmental processes. Single Plane Illumination Microscopy (SPIM)[1] is an emerging technique that enables live imaging of large specimens in their entirety. By imaging the living sample from multiple angles (views) SPIM has the potential to achieve isotropic resolution throughout even relatively large specimens. It also offers a number of other advantages over traditional confocal microscopy approaches, namely reduced photo damage to the sample and high acquisition frame rates of a CCD-based detection system. The necessary prerequisite for realizing the full potential of SPIM for biological imaging, however, is an image mosaicing stage that achieves precise registration of the data acquired from multiple angles, followed by proper fusion into a single isotropic 3D volume.

We use SPIM recordings of *Drosophila* embryogenesis to test the performance of the algorithms we have developed for this purpose. *Drosophila* embryo is a premium model system for studying animal embryonic processes. An extensive reverse genetic toolkit allows us to specifically mark general anatomical structures such as cell nuclei in the embryo as well as generate specific gene expression reporters for any gene in the fruit fly genome. We intend to map, using SPIM, the dynamic spatial and temporal changes in gene expression specificity for all important developmental genes with cellular resolution in the context of the entire intact developing embryo. Our algorithms for SPIM mosaicing will be crucial to efficiently process these systematically acquired *Drosophila* embryo recordings, but will be also applicable for processing of any multi-view SPIM image data of other biological specimens.

Medical Imaging 2008: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6914, 69140E, (2008)
1605-7422/08/$18 · doi: 10.1117/12.770893

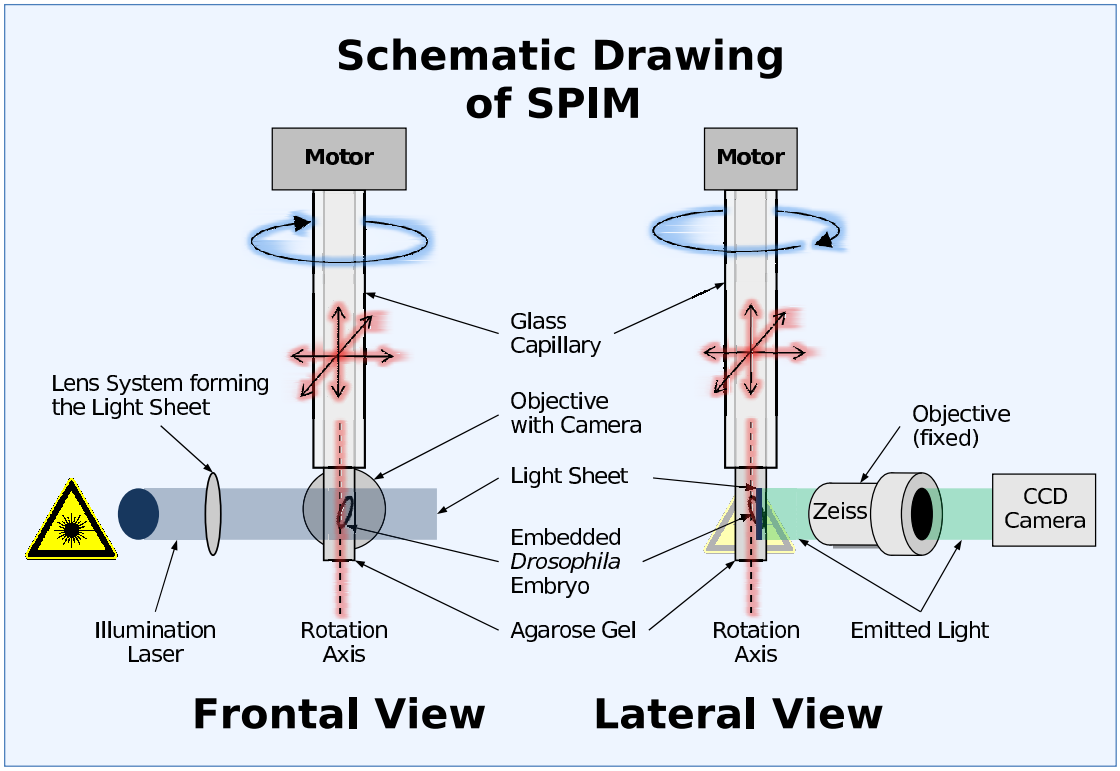

Figure 1. Schematic drawing of SPIM acquisition setup. *Left:* Frontal view, *right:* lateral view. The specimen (*Drosophila* embryo) is embedded in agarose gel protruding from a glass capillary. The capillary is rotated so the light sheet emitted by a laser illuminates the specimen at different angles. Emitted fluorescent light is detected by a CCD camera. The rotation angle of the specimen is known, but the precise orientation of the (invariant) rotation axis is not. Likewise, the translational position of the capillary is shifted for each angle to bring the specimen into the camera field of view.

2. METHODS

2.1 Imaging

SPIM, schematically illustrated in Fig. 1, achieves optical sectioning by focusing the excitation laser into a thin light sheet that penetrates the living sample embedded in an agarose gel and suspended by gravity in a water filled chamber. The objective lens is arranged perpendicular to the axis of illumination and so the laser illuminates only the imaged plane of the specimen, which the lens is focused on, thus minimizing photo-bleaching and laser damage of the living samples and allowing for very long time-lapse recordings. Two-dimensional images of emitted fluorescent light are captured by a CCD camera and discretized with a dynamic range of 12 bits. The use of a CCD camera enables a very fast acquisition rate that is important for capturing dynamic developmental events in living embryos. In order to acquire 3D image stacks, the sample is moved perpendicular to the light sheet, in our case in increments of 2 µm.

The SPIM instrument can, in principle, achieve an isotropic, high resolution along x, y and z axis allowing for the first time imaging of large three-dimensional specimens in their entirety. In order to achieve an isotropic resolution uniformly across the sample's 3D volume in all three dimensions, it is necessary to rotate the sample and record image stacks for the same specimen from different angles (usually 8 to 12).

We used a Zeiss 20×/0.5 Achroplan objective to image a developing *Drosophila* embryo expressing NLS-GFP* that marks all nuclei in the specimen. We acquired 4 and 8 angles for each time-point in the extended time-lapse recording, spaced by 90° and 45° respectively. To minimize the morphological changes during multi-view acquisition of a single time-point we slowed down the development by reducing temperature to 18°C using a Peltier-cooled sample chamber. For

*NLS = Nuclear Localization Signal; GFP = Green Fluorescent Protein

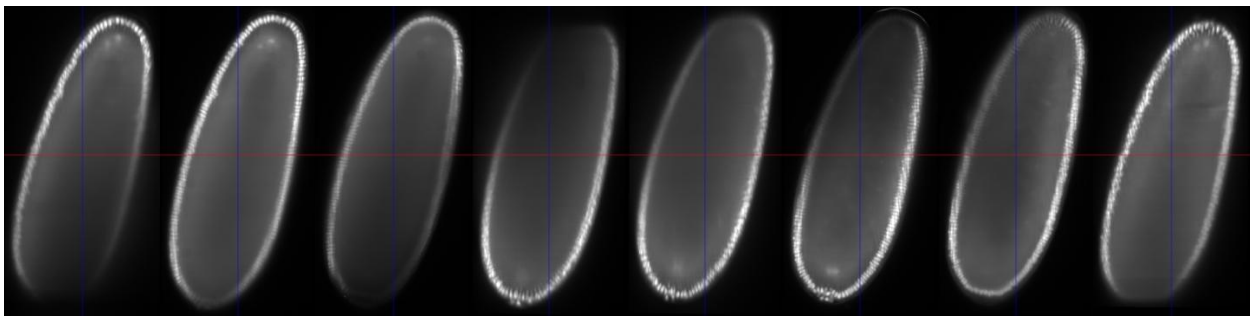

Figure 2. Slices from an 8-angle SPIM acquisition (after co-registration). The native image resolution was $0.73 \times 0.73 \times 2\,\mu$m pixel size, which was resampled to $0.73^3\,\mu$m isotropic resolution after registration. The varying regions of sharp and blurred image data are clearly visible here. The images were acquired with 45 degree increments, but note that the rotation was out of plane with respect to the slice orientation shown here.

registration and fusion shown in this paper we selected a single time-point just before gastrulation when the embryo consists only of a single closed layer of densely packed epithelial cells, surrounding the inner yolk. This stage of *Drosophila* embryogenesis therefore defines a comparatively complicated registration task during the whole development as the specimen is highly uniform and provides only minimal global structures for the registration.

Image samples from an 8-angle acquisition of a *Drosophila* embryo are shown in Fig. 2. It is apparent that each of these image stacks shows only part of the embryo in detail, largely due to scattering of light sheet and emitted light by the sample itself. This effect complicates the registration of the views and requires an advanced scheme for image fusion to create the final isotropic image.

2.2 Image Registration

2.2.1 Registration Cost Function

The registration was implemented using the stack entropy similarity measure used in the "congealing" algorithm by Learned-Miller.[2] In short, this measure calculates the entropy of pixel intensities over all views at each pixel, which is then summed over all pixels. By minimizing the total entropy, simultaneous registration of the entire group of images can be achieved.

2.2.2 Application-Specific Two-Tier Transformation Model

As the transformation model, we use a novel, application-specific model with per-image as well as groupwise transformation parameters. In general, each image relates to the image-independent reference coordinate space by a rigid coordinate transformation (Fig. 3). However, by construction of the SPIM imaging setup, the following observations can be made *a priori*:

1. There is only one rotational degree of freedom (rotation of the specimen). The precise rotation angle is known for each of the acquired views, based on the precise stepper motor that turns the capillary with the embedded specimen.
2. The precise orientation of the capillary rotation axis is not calibrated and therefore unknown, but constant for all views, thus leading to *two transformation degrees of freedom shared by all images.*
3. The specimen translation is different for each view, due to re-adjustment of the camera field of view, leading to *three independent degrees of freedom for each of the views.*

For the i-th view, its coordinate transformation $\mathbf{T}_i$ is parameterized as

$$\mathbf{T}_i(x_i, y_i, z_i; \alpha, \beta; \underbrace{\Phi_i}_{\text{const.}}) : \mathbb{R}^3 \mapsto \mathbb{R}^3, \tag{1}$$

where x_i, y_i, and z_i are the three translational parameters for image i; α and β are two angles that define the orientation of the acquisition rotation axis; and Φ_i is the known view angle for view i.

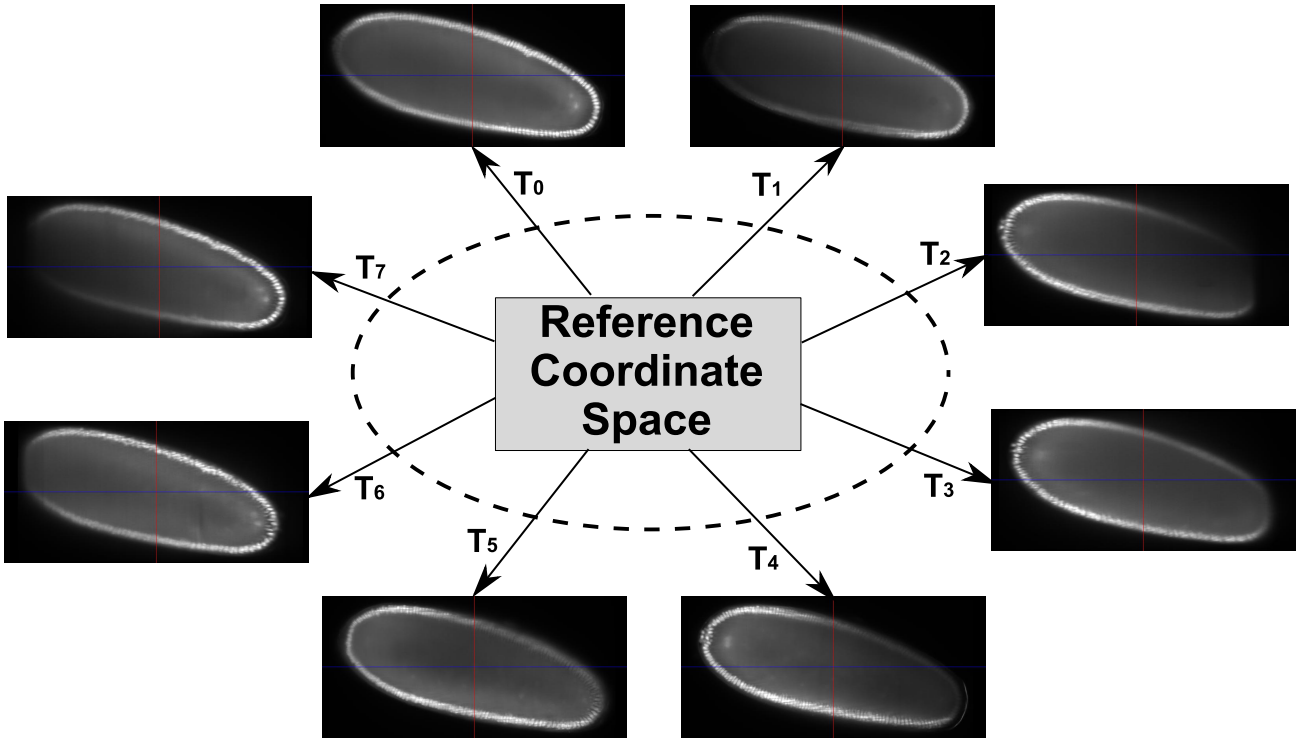

Figure 3. Illustration of the groupwise transformation model. An image-independent reference coordinate space relates to each of the input images (here: eight) by a separate coordinate transformation, $\mathbf{T}_0$ through $\mathbf{T}_7$. The transformations are parameterized and constrained to leave the reference space effectively unmoved. See text for details.

The resulting transformation model for eight views, therefore, has a total of 26 degrees of freedom, compared with 48 degrees of freedom for independent rigid transformations. For four views, the number of parameters in the application-specific transformation model is 14, compared with 24 for independent transformations.

2.2.3 Optimization Strategy

The parameters of the combined transformation model that minimize the stack entropy image similarity measure are determined by an alternating iterative optimization scheme. The per-image translational parameters are optimized using a gradient-based line search algorithm.[3] The orientational parameters of the rotation axis are determined using a hill climbing scheme similar to the algorithm proposed by Studholme *et al.*[4] At each level of a multi-resolution schedule, both optimization stages are performed alternatingly until no further improvement can be achieved. In parallel, the resolution of the image data is incrementally increased in a multi-scale hierarchy, starting at 8 times the original pixel size, down to the original image resolution.

2.3 Content-Based Image Fusion

An effective fusion algorithm of the aligned image stacks must ensure that each angle contributes to the final fused volume only useful sharp image data acquired from the area of the sample close to either the entry point of the light sheet or the detection lens, ideally both. The influence of the blurred parts of each single view caused by illumination and detection light scattering through the sample must be minimized, which makes simple image averaging unsuitable (see Figs. 4 and 5).

Following an idea developed by Goshtasby,[5] we fuse the co-registered images according to their local information content, and thus aim to overweight the "sharp" versus the "blurred" parts of the individual images. One method to approximate the local information is by evaluating the entropy in the local environment of each pixel in each view α and use those values as weighing factors for the averaging of every individual pixel over the registered volume. For every pixel of image I_α the regional entropy $H_\alpha(x, y, z)$ is computed in a local pixel neighborhood. The resulting fused pixel is then computed as the weighted average of the pixels in the original images, each weighted with the regional entropy. As even

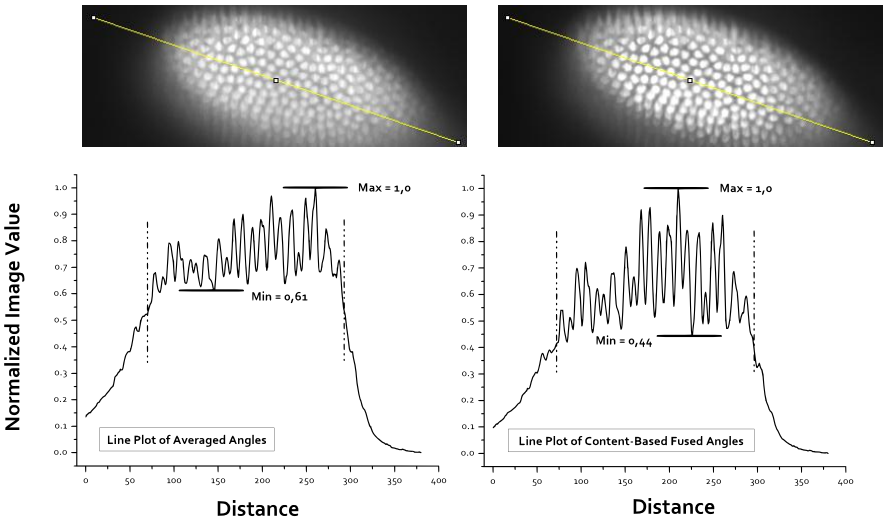

Figure 4. Two identical planes near the surface of the embryo in the images fused by simple averaging (*top left*) and weighted by local image entropy (*top right*). The entropy-weighted fusion yields noticeably higher contrast compared with the unweighted average image, so that the nuclei are more easily distinguished. Higher contrast is also confirmed by the line profile plots (*bottom*) along the yellow lines in both images.

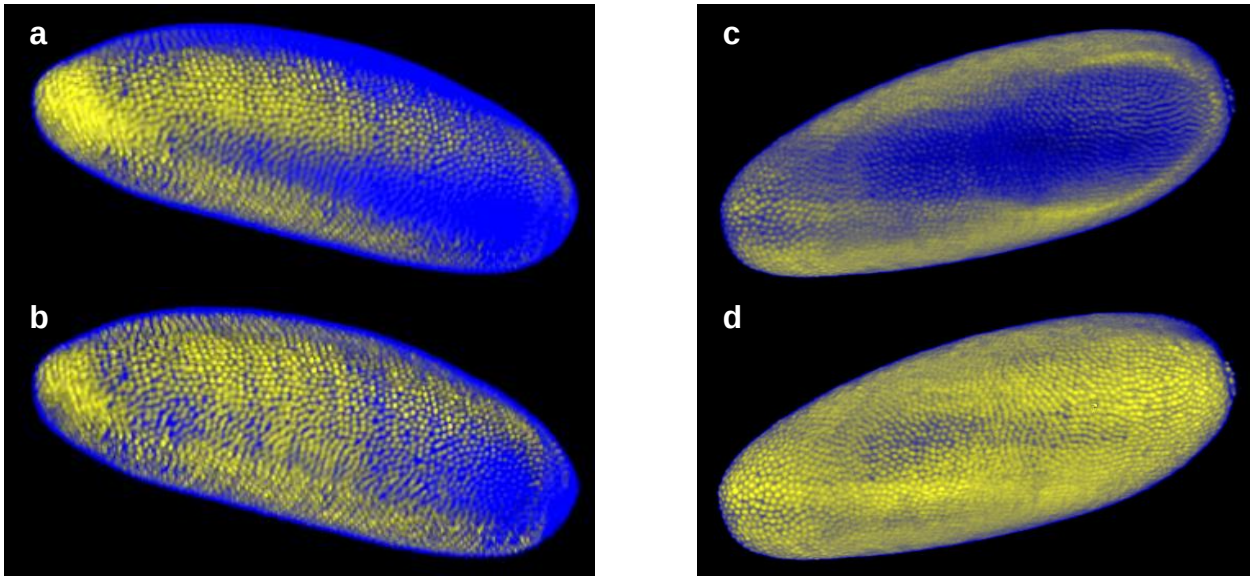

Figure 5. Volume renderings of fused SPIM data from 4 angles (a,b) and 8 angles (c,d) using unweighted averaging and entropy-based fusion. High image intensities are rendered in yellow, whereas lower intensities are rendered in blue. The 4-angle renderings reveal that in areas where none of the views has perfect data the averaging (a) looses even more contrast (blue stripes) while the entropy fusion (b) is able to preserve an almost constant contrast. For the the eight angle data, unweighted averaging suffers from a high amount of blurred data in each pixel whereas the entropy-based fusion is able to reduce this effect substantially

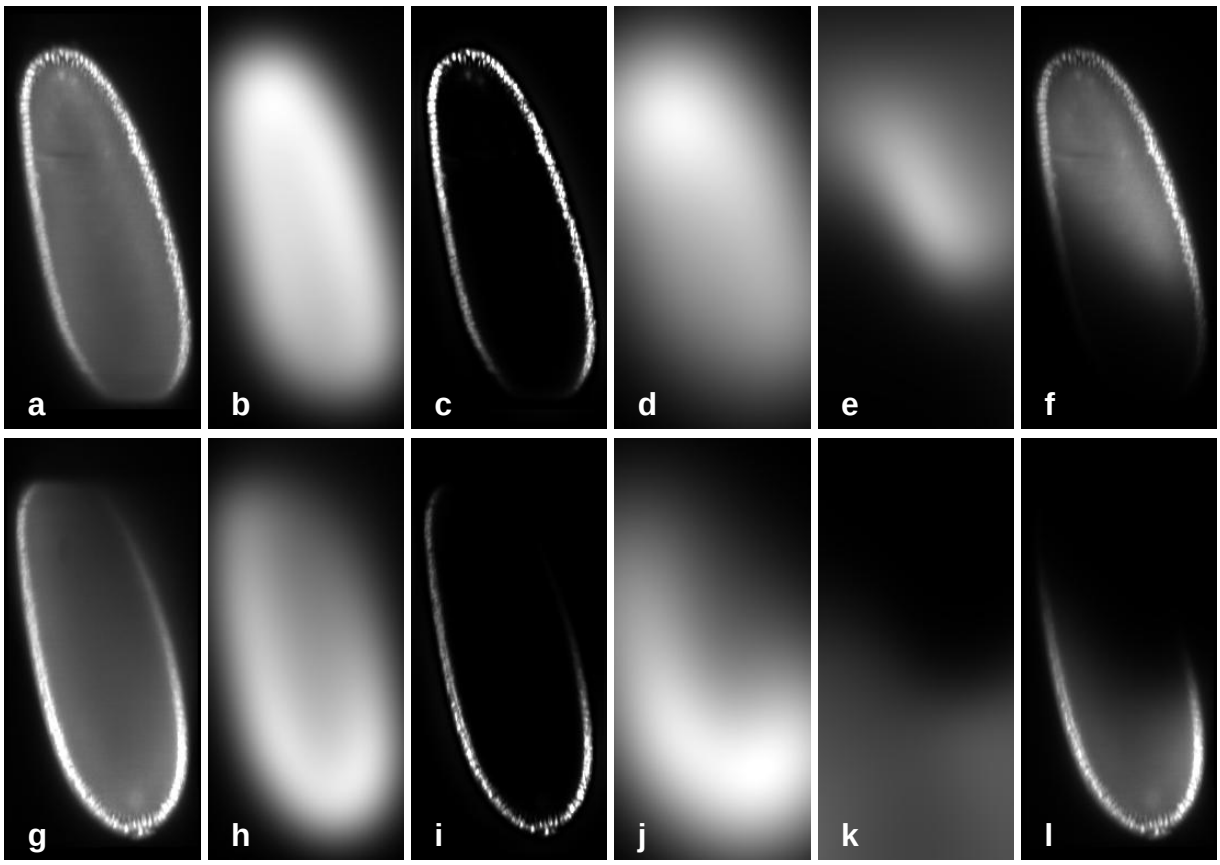

Figure 6. Steps of the fast content-based fusion algorithm illustrated on two different angles (top row: 0°, bottom row: 180°). a) One of the original SPIM images. b) Image after first Gaussian smoothing. c) Squared difference between original and smoothed image. d) Smoothed difference image. e) Smoothed difference image for this angle, divided at each pixel by the sum of the smoothed differences over all angles. This image quantifies the local image information relative to the set of images from all angles. f) Original image weighted at each pixel with the relative information in e). Images g) through l) show the corresponding results for a second angle of the same acquisition.

the blurred image regions are not completely uniform (e.g., due to image noise and illumination gradient), they, too, have entropies larger than zero. To further suppress these regions, we emphasize the regional entropy weight by taking H into the exponent:

$$I_{\text{fused}}(x,y,z) = \frac{\sum_\alpha 100^{H_\alpha(x,y,z)} I_\alpha(x,y,z)}{\sum_\alpha 100^{H_\alpha(x,y,z)}}. \tag{2}$$

The regional image entropy is implemented using a discrete intensity histogram. The number of histogram bins and the size of the spatial window in which the regional entropy is computed are adjustable parameters of the entropy computation. In order to determine the optimal parameters, Goshtasby proposed to apply an optimization scheme that maximizes the global image entropy. Due to the staggering amount of data in SPIM acquisitions, such an optimization approach is not feasible here. Instead, we applied Goshtasby's approach to a small subvolume of the data once to determine acceptable parameters tuned for our type of images. This yielded a window size of 19^3 pixels and 256 histogram bins, which may be suboptimal for other types of image data and should be established independently for other applications.

2.4 Fast Implementation of Content-Based Fusion

Compared with the two-dimensional 640×480 pixel images fused by Goshtasby, our images are rather large in size: the cropped, registered, and reformatted images from the 8-angle acquisition used in this paper have a size of 359×379×773 pixels, which is about two orders of magnitude (and one dimension) larger. Computational performance therefore becomes

Figure 7. Illustration of truncation artifact that is removed by content-based image fusion. (a) Due to the truncated acquisition of one of the eight views, volume rendering the result of simple averaging shows a clear artifact in the shape of an ellipse. (b) The artifact is removed by the content-based image fusion, due to the low information content of the image region in the truncated view near the truncation surface.

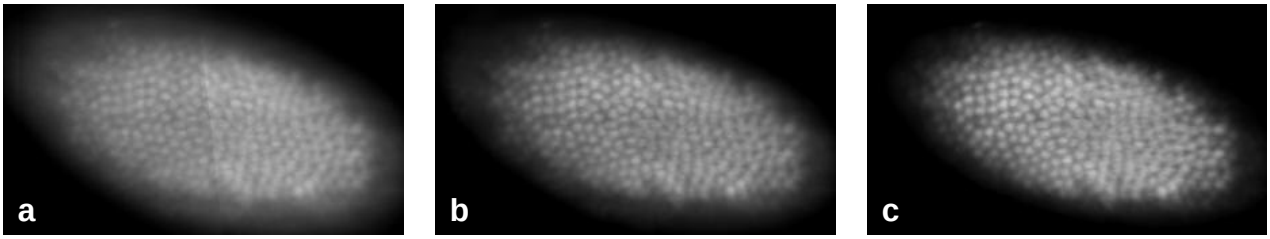

Figure 8. Comparison of image contrast using (a) unweighted averaging, (b) entropy-based fusion, and (c) fast fusion using Gaussian filters. Note the improved sharpness of the nuclei in (b) and (c). Note also again in (b) and (c) the effective removal of the view truncation artifact that is visible as a near-vertical boundary in (a).

an issue, and we now propose a computationally efficient approximation algorithm to perform the content-based image fusion. In a first simplification step, we note that the local entropy in an image region is monotonically related (via the log function) to the variance of the pixel values in that region, if they are normally distributed. Unfortunately, computing the local regional variance is still computationally expensive. Instead, we implement an approximation to the regional variance by convolution with a Gaussian kernel, followed by a pixel-wise squared difference operator, followed by a second Gaussian kernel convolution:

$$W_\alpha = G_{\sigma_2} * (I_\alpha - (G_{\sigma_1} * I_\alpha))^2 \tag{3}$$

Using a recursive implementation of the Gaussian filter[6] in the InsightToolkit (`http://www.itk.org`), we arrive at an implementation that runs in constant time, regardless of the filter kernel size. The fused image is then easily computed as

$$I_{\text{fused}}(x,y,z) = \frac{\sum_\alpha W_\alpha(x,y,z) I_\alpha(x,y,z)}{\sum_\alpha W_\alpha(x,y,z)}. \tag{4}$$

The processing steps of the fast fusion algorithm as illustrated in Fig. 6 for two views of the 8-view SPIM data.

3. RESULTS

The effectiveness of the image registration algorithm proposed herein is visually apparent from the volume-rendered composite image stacks in Fig. 5. This figure and Fig. 4 also clearly demonstrate the need for content-based image fusion, as opposed to simple image averaging. As a beneficial side effect, content-based fusion also eliminates artifacts in the combined image that would otherwise arise from views that were acquired with a reduced scanning depth, thus truncating the embryo (Fig. 7).

The results of the fast fusion approach using Gaussian filters are compared with those of the entropy-based fusion and simple averaging in Fig. 8. The two content-based fusion techniques produced visually similar results, and they both clearly outperformed simple averaging. However, entropy-based fusion using 19×19×19 pixel neighborhoods took about 17 hours of computation time on a single CPU (single core on an Intel Core2 Quad Xeon at 1.8 GHz). Compared with that, the fast fusion algorithm based on Gaussian filters ($\sigma_1 = 42\,\text{pixels}$, $\sigma_2 = 88\,\text{pixels}$), took only about 15 minutes on the same CPU, albeit at the cost of either larger memory consumption or more disk I/O.

4. DISCUSSION

To the best of our knowledge, this paper makes the following contributions: (1) We have described the first integrated algorithm for mosaicing of SPIM images acquired without calibrated rotation axis. Our method is also the first application of a groupwise registration method to volumetric image mosaicing in general. (2) We have introduced an application-specific transformation model for the groupwise alignment of SPIM data, which separates groupwise from per-image transformation parameters, thus constraining the search space and reducing the number of degrees of freedom for optimization. The application-specific transformation model also restricts the overall alignment of the views to transformations that are physically possible (e.g., such that rotate all views around the same, constant rotation axis) and effectively prohibits alignments that are not physically plausible. (3) We have introduced a fast algorithm for image fusion based on local information content, which is implemented using efficient Gaussian filtering and is about two orders of magnitude faster than a comparable algorithm based on local image entropy.

In summary, we have developed an image mosaicing technique for SPIM images that uses groupwise image registration with an application-specific transformation model and content-based image fusion. Groupwise registration successfully aligned the images acquired from different angles. Content-based fusion effectively identified the information-carrying areas in each of the partial images, substantially reduced blurring as compared with simply averaging of the co-registered images, and removed volume truncation artifacts. The resulting images appear suitable for further analysis, such as segmentation and tracking of single cell nuclei. In future acquisitions, the successful suppression of truncation artifacts by the content-based fusion will enable us to purposely reduce the image acquisition depth of each view, which in turn will enable increased numbers of views and/or increased temporal resolution with a larger number of time points.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Carl Zeiss Microimaging GmbH (Jena, Germany), in particular Christian Dietrich, Helmut Lippert and Chris Power, for providing the SPIM imaging equipment and for their support. We thank Radoslaw Ejsmont at the MPI-CBG for creating the flies and acquiring the SPIM data. SP and PT were supported by the Max-Planck-Gesellschaft. TR and MPH were supported by the National Institute on Alcohol Abuse and Alcoholism through Grant AA005965. SP and TR thank the organizers of the 2007 Janelia Hackerthon, held at the HHMI Janelia Farm Research Campus, for facilitating this collaboration. SP acknowledges a travel grant from Boehringer Ingelheim Fonds to support an International Fellowship visit to SRI International.

REFERENCES

1. J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science* **305**, pp. 1007–1009, August 2004.
2. E. G. Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, pp. 236–250, February 2006.
3. T. Rohlfing and C. R. Maurer, Jr., "Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees," *IEEE Transactions on Information Technology in Biomedicine* **7**(1), pp. 16–25, 2003.
4. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Medical Physics* **24**, pp. 25–35, January 1997.
5. A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing* **23**, pp. 611–618, June 2005.
6. R. Deriche, "Fast algorithms for low-level vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, pp. 78–87, January 1990.