Optimization-Based Artifact Correction for Electron Microscopy Image Stacks*

Samaneh Azadi, Jeremy Maitin-Shepard, and Pieter Abbeel

EECS Department, University of California at Berkeley, USA

Abstract. Investigations of biological ultrastructure, such as comprehensive mapping of connections within a nervous system, increasingly rely on large, high-resolution electron microscopy (EM) image volumes. However, discontinuities between the registered section images from which these volumes are assembled, due to variations in imaging conditions and section thickness, among other artifacts, impede truly 3-D analysis of these volumes. We propose an optimization procedure, called EMISAC (EM Image Stack Artifact Correction), to correct these discontinuities. EMISAC optimizes the parameters of spatially varying linear transformations of the data in order to minimize the squared norm of the gradient along the section axis, subject to detail-preserving regularization.

Assessment on a mouse cortex dataset demonstrates the effectiveness of our approach. Relative to the original data, EMISAC produces a large improvement both in NIQE score, a measure of statistical similarity between orthogonal cross-sections and the original image sections, as well as in accuracy of neurite segmentation, a critical task for this type of data. Compared to a recent independently-developed gradient-domain algorithm, EMISAC achieves significantly better NIQE image quality scores, and equivalent segmentation accuracy; future segmentation algorithms may be able to take advantage of the higher image quality.

In addition, on several time-lapse videos, EMISAC significantly reduces lighting artifacts, resulting in greatly improved video quality.

A software release is available at `http://rll.berkeley.edu/2014_ECCV_EMISAC`.

Keywords: Denoising, Volume Electron Microscopy, Connectomics, Optimization, Segmentation.

1 Introduction

Recent technological developments in automated volume electron microscopy (EM) enable the acquisition of multi-terravoxel volumes at near isotropic resolution in the range of 3–30 nm [12,14,24,31,26]. These high-resolution image volumes are critical to fields such as connectomics, which aims to comprehensively map neuronal circuits by densely reconstructing neuron morphology and identifying synaptic connections between neurons [16,15,6].
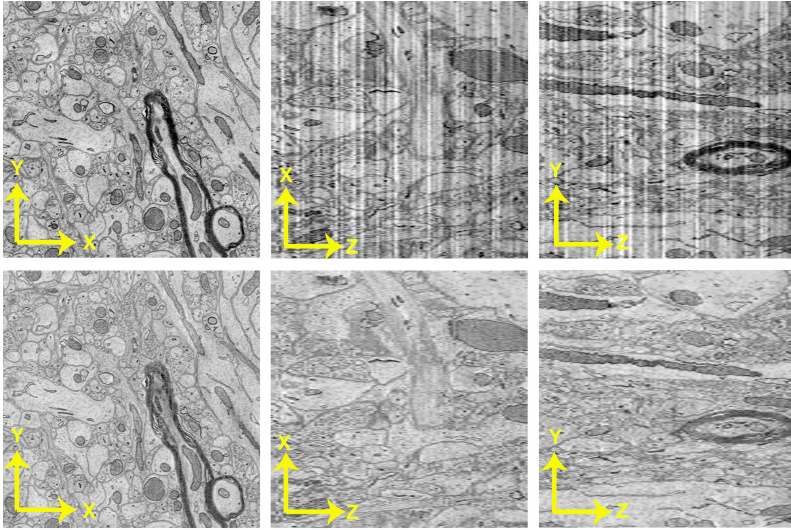

Fig. 1. Representative cross-sectional views of a $6 \times 6 \times 30$ nm ATUM-based SEM [31] image volume of mouse cortex. *First row:* The original, registered dataset clearly showing discontinuities along the Z (section) axis. *Second row:* Corrected data using our EMISAC algorithm.

All methods for volume electron microscopy, aside from tomography, which is only suitable for samples less than 1 micron in thickness, assemble the volume by spatially aligning a stack of two-dimensional images. Consequently, there can be substantial artifacts in the image volume, most notably serious discontinuities along the section axis. These are the result of variations in section thickness, sample deformations, and variations in imaging conditions. These artifacts are particularly a problem with ATUM-based SEM (Automated Tape Collecting Ultramicrotome-based Scanning Electron Microscopy) [31], a method that currently achieves the highest throughput and also has the advantage of preserving tissue sections, in contrast to the one-shot destructive imaging process of Serial Block Electron Microscopy (SBEM) [12] and Focused Ion Beam Scanning Electron Microscopy (FIBSEM) [24].

Figure 1 shows the artifacts typical in ATUM-based SEM volumes. In other imaging domains, improved imaging techniques and mathematical corrections have been devised for reducing artifacts in MRI and echo-planar images [13,2,1,11], and in 2-D electron microscopy images [21], but there has been less focus on three-dimensional EM volumes. [23]

When present, these artifacts prevent automated and manual analysis of volumes except as a series of 2-D images along the original sectioning axis, preventing in particular the extraction of truly 3-D image features. Given that structures, such as neurites in cortex, may have arbitrary 3-D orientations relative to the original sectioning axis, this limitation is a serious impediment. On SBEM and FIBSEM volumes without these artifacts, the ability to view cross sections along

arbitrary axes has aided humans tasked with manually tracing neurites and detecting synapses [16,15], and automated algorithms for reconstructing neurite morphology (via segmentation) have depended on 3-D features. [18,33,17,19,4,3].

To eliminate these artifacts and enable truly 3-D analysis of such image volumes, we propose a coarse-to-fine optimization-based procedure EMISAC (EM Image Stack Artifact Correction). We note that a single per-section brightness and contrast adjustment (i.e. linear transform of intensity values) is inadequate on typical datasets for correcting discontinuities except very locally. EMISAC optimizes the parameters of spatially varying linear transformations of the data in order to minimize the squared norm of the gradient along the section axis, subject to detail-preserving regularization, as described in Section 2.

We applied EMISAC to a publicly available ATUM-based SEM volume of mouse cortex as well as to a serial section Transmission Electron Microscopy (ssTEM) volume of Drosophila larva ventral nerve cord. Figure 1 shows several cross-sectional views of the output of our algorithm. Qualitatively, EMISAC appears to completely eliminate all discontinuity artifacts while preserving all of the original detail.[1] (We did not attempt to address the problem of lower Z resolution than X-Y resolution.) Quantitatively, an evaluation based on the NIQE blind image quality metric [28] confirms the qualitative results. More importantly, we evaluated the effect of EMISAC as a preprocessing step on the accuracy of automated segmentation of neurites, a key challenge for this type of data; consistent with the NIQE scores, EMISAC dramatically improved segmentation accuracy.

After developing our approach independently, we became aware of a recent method [23] for addressing the same problem. Like our approach, this alternative method involves a quadratic optimization to minimize the squared norm of the gradient along the section axis, but uses a different parameterization and a different form of regularization and post-processing to preserve the intra-slice detail. We included this alternate method in our evaluations, and found EMISAC matches or outperforms it. In addition, we evaluated several other generic correction methods on the raw datasets including histogram equalization, contrast-limited adaptive histogram equalization [38], and local normalization, and found that EMISAC significantly outperforms all of them.

Furthermore, while designed primarily for electron microscopy image stacks, EMISAC is also applicable to lighting correction in time-lapse photography. Raw time-lapse image sequences typically have serious inter-frame lighting discontinuities. Although a few tools such as LRTimelapse[2] are designed to edit time-lapse sequences and correct these artifacts, our method does not require manual editing and can make local corrections to lighting as well. We evaluated EMISAC on several time-lapse videos exhibiting lighting problems; compared to the original videos, EMISAC produced a large qualitative improvement, and eliminated essentially all lighting discontinuities

---

[1] Our qualitative assessment was based on only a random subset of the cross-sections; we did not scrutinize every single cross-sectional view. Our quantitative assessment is more comprehensive.
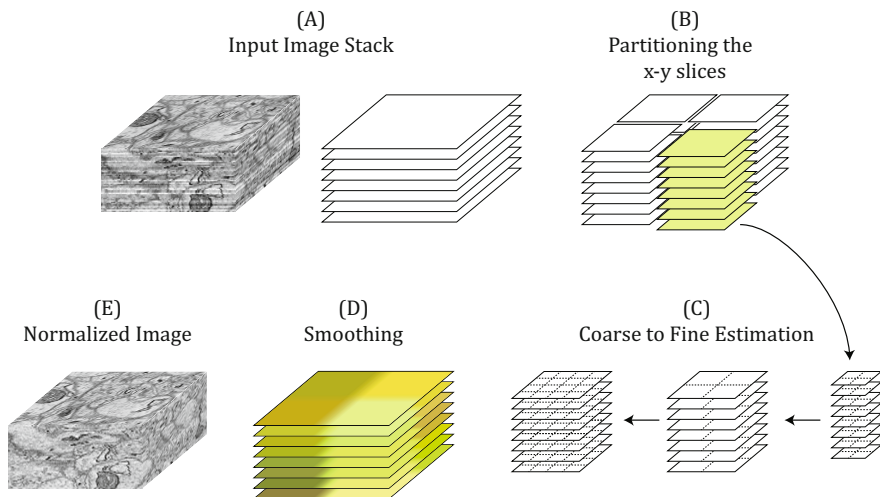
[2] http://lrtimelapse.com

Fig. 2. Schematic of our approach. (A) An aligned EM image stack of 2-D slices. (B) The volume is optionally partitioned into a grid, which can be distributed across multiple processors; only limited communication is required between neighboring processors. (C) A coarse-to-fine procedure, in which both the data as well as the parameters are initially downsampled, speeds up the optimization of the objective in Eq. (3). From right to left: downsampled x-y slices and downsampled parameters (larger block size); downsampled parameters only; full parameters (small block size). (D) The parameters are spatially smoothed within each x-y slice to remove the blocking effect. Four blocks are shown for illustration purposes, but in practice there are many more blocks. (E) Corrected output image.

2   Artifact Correction Algorithm

In order to maximize the applicability of our approach, and avoid introducing a model bias[3] that could harm the accuracy of later stages of processing, we designed our approach to make as few and as simple assumptions as possible:

1. the true (undistorted) image volume is mostly continuous along the section axis;
2. the distortions in the volume can be expressed as *local* linear transformations of the intensity values, where the parameters of the linear transforms vary smoothly within each section (but are not smooth between sections).

The detailed formulation of our method is explained in the following sections. A summary of our approach is illustrated in Fig. 2.

---

[3] A denoising method based on a learned sparse-coding dictionary, for instance, could potentially introduce patterns that were not present in the original data.
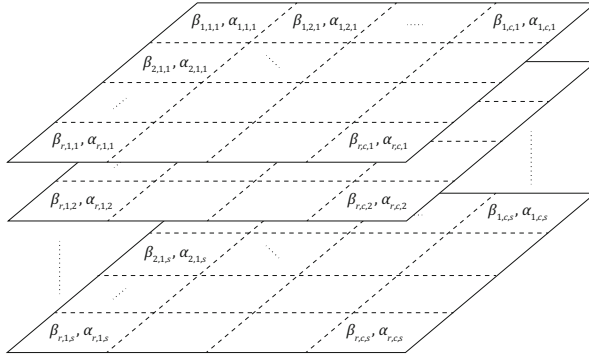
Fig. 3. EM image stack as a set of smaller blocks. The total number of blocks in the x-direction, y-direction, and z-direction are given by r, c, and s, respectively, resulting in a total of $r \cdot c \cdot s$ blocks. We have unique β and α parameters for each block.

2.1 Problem Formulation

As shown in Fig. 3, we partition each slice of the 3-D EM volume into small fixed-size two-dimensional blocks. Voxel intensity values are corrected by a linear transformation given by:

$$I'_{x,y,z} = \beta_{\lfloor x/w_{\mathrm{x}} \rfloor, \lfloor y/w_{\mathrm{y}} \rfloor, z} \cdot I_{x,y,z} + \alpha_{\lfloor x/w_{\mathrm{x}} \rfloor, \lfloor y/w_{\mathrm{y}} \rfloor, z} \tag{1}$$

where $I(x, y, z)$ refers to the scalar intensity value at position (x, y, z) of the volume, and $(w_{\mathrm{x}}, w_{\mathrm{y}})$ is the block size. The smaller the block size, the larger the number of parameters. Note that β and α correspond to correction of contrast and brightness, respectively. This block-based scheme effectively captures the local similarities within each slice and reduces the computational complexity.

We express our assumption of continuity in I' along z as a penalty on the squared norm of the gradient with respect to z. Likewise, the smoothness assumption on the the affine transforms is expressed as a penalty on the squared norms of the gradients of α and β with respect to the in-slice block positions. Thus, we formulate the optimization problem as follows:

$$\min_{\beta, \alpha} \sum_x \sum_y \sum_z (I'_{x,y,z+1} - I'_{x,y,z})^2$$
$$+ \gamma \sum_i \sum_j \sum_z \left[(\beta_{i+1,j,z} - \beta_{i,j,z})^2 + (\beta_{i,j+1,z} - \beta_{i,j,z})^2 \right. \tag{2}$$
$$\left. + (\alpha_{i+1,j,z} - \alpha_{i,j,z})^2 + (\alpha_{i,j+1,z} - \alpha_{i,j,z})^2 \right]$$
$$\text{s.t.} \quad \beta_{i,j,z} \geq 1, \quad \forall i, j, z.$$

The β parameters must be bounded to avoid the trivial null solution.

We reformulate the optimization problem given in Eq. (2) as the convex quadratic problem

$$\min_X \|D_z X\|_2^2 + \gamma(\|D_x X\|_2^2 + \|D_y X\|_2^2) \qquad \text{s.t.} \quad \beta \geq 1, \qquad (3)$$

where $X=[\beta, \alpha]$, $D_z = G_z A$, $D_x = G_x$, $D_y = G_y$, A is a matrix that maps the parameters X to a vector expressing I', and G_a is a matrix that maps a vectorized volume to a vector containing the finite-differences approximation of its gradient along axis a.

We use L-BFGS-B [37] to solve the optimization problem in Eq. (3). The optimization is stopped when the fractional decrease in objective between consecutive iterations falls below $\epsilon \cdot f$, where $\epsilon = 2^{-52}$ is the machine precision. To speed up the optimization process, we employ a coarse to fine estimation procedure detailed below.

2.2   Coarse-to-Fine Procedure

Rather than solving Eq. (3) with the final desired block size directly, we solve a sequence of optimization problems of the form shown in Eq. (3) with increasing image resolution in the x-y plane and/or decreasing block size. Each optimization after the first is initialized with the (appropriately upsampled) parameters that solved the previous optimization. In practice, we used (a single succession of) the following three steps:

1. The image resolution is reduced by a factor of 2 in x and y (using 2×2 averaging), as are the number of blocks.
2. The full image resolution is used, but the number of blocks remains reduced by a factor 2 in x and y.
3. The number of blocks is increased by a factor of 2 in x and y to its final size.

As the optimization is convex, the final solution obtained is unaffected by this procedure, but typically the running time is greatly reduced.

2.3   Removing the Blocking Effects

Ideally, the coarse-to-fine procedure is continued all the way down to a block size of $(1, 1)$. However, to reduce running time, it may be desirable to stop the procedure at a non-trivial block size. To avoid introducing artifacts from the blocking, after performing the final optimization, we upsample the α and β parameters to a block size of 1 using linear interpolation (rather than nearest neighbor interpolation).

2.4   Parallelization

For large image volumes, we can partition the volume into a grid along the x and y axes, which can be distributed across multiple processors or machines.

For simplicity, the grid cell boundaries should be aligned to block boundaries. Only the parameter gradients for blocks on the border of each grid cell must be communicated, at each iteration of the optimization, to the processors responsible for neighboring grid cells, which is in general a very small amount of data relative to the size of the image volume. As a simplifying approximation, we could even ignore the regularization term between neighboring grid cells and thereby require no communication between machines. In practice this may not significantly affect the result provided that the grid cells are large enough. In our experiments, we observed no loss of accuracy (in NIQE score) from using no communication between blocks, except for the final upsampling step.

3   Evaluation on Electron Microscopy Data

We tested EMISAC on a publicly available ATUM-based SEM volume of mouse cortex released by Kasthuri *et al.* [20] For a $1024 \times 1024 \times 100$ voxel portion of this dataset, a dense segmentation of neurites traced by a human expert is also publicly available as part of the SNEMI3D neurite segmentation challenge [5], which enabled us to evaluate the effect of EMISAC on segmentation accuracy. The dataset was acquired at a resolution of $3 \times 3 \times 30$ nm, but as the human-traced segmentation is provided only at the downsampled resolution of $6 \times 6 \times 30$ nm, we used the same downsampled resolution for all of our experiments. In addition, we tested EMISAC on another publicly available dataset, Cardona *et al.* 2010, collecting using a different imaging technique, ssTEM rather than ATUM-SEM, and of Drosophila larva ventral nerve cord rather than mouse cortex, with a resolution of $4 \times 4 \times 50$ nm [7,8].

We compared EMISAC against the original aligned but uncorrected image volume. We are aware of only one other method designed to address this same problem (of which we only became aware after independently developing our own algorithm), which we refer to as Kazhdan2013 [23]. As the authors of that method made publicly available their output [22] on the same mouse cortex volume on which we tested EMISAC, we were able to include the Kazhdan2013 algorithm in our evaluations without having to reimplement it. We also compared EMISAC against histogram equalization, contrast-limited adaptive histogram equalization (CLAHE) [38], and local normalization [30].

For EMISAC, we set the affine transform regularization parameter $\gamma = 0.4 \frac{w_\mathrm{x} w_\mathrm{y}}{d_\mathrm{x} d_\mathrm{y}}$ for all electron microscopy datasets, where d_x and d_y are the image downsampling factors in x and y respectively (relative to the $6 \times 6 \times 30$ and $4 \times 4 \times 30$ nm resolution data). The coefficient was selected to minimize NIQE score, which does not depend on any labeled data. Furthermore, results were fairly insensitive to several orders of magnitude change in γ.

3.1   Image Quality Evaluation

Although any corrected version of the data is only useful in so far as it aids an image analysis task of interest, such as neurite segmentation, it is convenient to

be able to directly quantify the image quality independent of any particular later analysis step. A visual assessment by humans would be inherently subjective (and also inconvenient), and it is impossible to obtain a "ground truth" version of the data without any imaging artifacts, against which the corrected version might be compared. Furthermore, we have no way of knowing the true distortion model. We therefore rely on the "completely blind" Natural Image Quality Evaluator (NIQE) [28], which requires neither a model of expected distortions nor human assessments of distorted images as training data, but merely a set of high quality images from which to estimate a model of natural image statistics. On natural image benchmarks, this method is comparable to the best methods that *do* rely on human assessments as training data.

We trained a NIQE model on a random set of x-y sections from the dataset that were not part of any of the volumes on which we tested our approach. Thus, the NIQE score of a y-z or x-z cross section represents the statistical similarity of the orthogonal cross sections to the original image sections, a very reasonable metric given that the ultimate goal is to be able to analyze the 3-D volume without regard to a preferred orientation.

3.2 Segmentation Accuracy Evaluation

As one of the primary goals of our work on the artifact correction is the improvement of automated segmentation results for neural circuit reconstruction, we directly evaluated the impact of EMISAC on 3-D segmentation accuracy. The training of our machine-learning-based segmentation algorithm, as well as the evaluation of segmentation accuracy, were based on the SNEMI3D human expert-traced segmentation, which we treat as "ground truth."

Although the focus of this work is not on segmentation algorithms, for completeness we describe the three-step segmentation procedure that we used:

1. We use an unsupervised procedure to transform each position in the image volume into a 1-of-k binary feature vector, with $k = 624 \cdot 16$. We cluster $16 \times 16 \times 4$ patches of the image volume using k-means based on L_1 distance; the binary feature vector for each position is obtained by vector quantizing the image patch centered at that location. We take advantage of the assumed rotational covariance of the data (namely transposition and reflection in the x-y plane, and reflection along the z axis), which reduces the effective number of parameters by a factor of 16.

2. To predict the presence of a cell boundary between two adjacent voxels along the x, y, or z axes, we train a logistic regression classifier for each of the 3 axes. The feature vector for classification is obtained by concatenating all of the 1-of-k binary feature vectors within a $16 \times 16 \times 16$ window around the boundary (producing a very high-dimensional feature vector). We extracted boundary information for training examples directly from the human-provided ground-truth segmentation, and ensured that equal total weight was given to positive and negative training examples. We optimized the classification model using L-BFGS, using a quadratic approximation to dropout [35] (with $p = 0.5$) for

regularization. As for the unsupervised feature learning, we take advantage of the assumed rotational covariance to reduce the number of parameters to be learned by a factor of 16.

3. To produce a segmentation, we employ Gala [29], a state-of-the-art electron microscopy image segmentation algorithm, based on agglomeration of supervoxels, for which we use the cell boundary predictions as input.

We use half of the ground truth segmentation to train the boundary classifier, half of the remaining portion to train Gala, and the remainder for evaluation. The same training/testing procedures were used for the original data, the EMISAC output, and the Kazhdan2013 output. The results are averaged over 4 splits.

We evaluate the segmentation accuracy relative to the human-labeled ground-truth using Variation of Information (VI) [27], which has been used by recent prior work for evaluating EM segmentations. [29,3] The variation of information between two segmentations, S and U, is defined as the sum of two conditional entropy terms, $H(U|S)$ and $H(S|U)$, where each segmentation assigns to each voxel position a segment identifier, and the entropy terms are with respect to the distribution over joint segment assignments $(S(x), U(x))$ obtained by sampling positions x within the volume uniformly at random. Thus, $H(S|U)$ can be seen as a measure of false splits in S with respect to U, and $H(U|S)$ can be seen as a measure of false merges in S with respect to U. A lower score corresponds to greater accuracy.

4   Electron Microscopy Results

To guide later experiments, we initially evaluated the effect of varying the block size on running time and image quality (measured by NIQE score); the results are shown in Fig. 4. The running times reported for this and later experiments are based on our Python implementation running on an 8-core Intel Xeon X5570 2.93 GHz system, which consumed about 15 GB memory for each $1024 \times 1024 \times 100$ volume. Based on the observed trade-off between image quality and running time, we used a final block size of $(16, 16)$ for later experiments. We found that the NIQE score typically converged before reaching the threshold of $f = 10^{10}$.

4.1   Comparison of NIQE Scores

Using these parameters, we evaluated the improvement in NIQE score relative to the original data of EMISAC, Kazhdan2013, histogram equalization, contrast-limited adaptive histogram equalization [38], and local normalization. Figure 5 shows the distribution of NIQE scores for the SNEMI3D volume. Table 1 shows the improvement in NIQE score on both datasets. Under Welch's t-test, EMISAC attains a large and highly statistically significant improvement in both x-z ($p < 0.0001$) and y-z ($p < 0.0001$) NIQE score relative to Kazhdan2013 on the ATUM-SEM volume, without any considerable loss in x-y NIQE score. The preservation of detail is confirmed by the very high structural similarity (SS) [36] between the original and corrected x-y cross-sections. See Fig. 6 for a visual comparison.
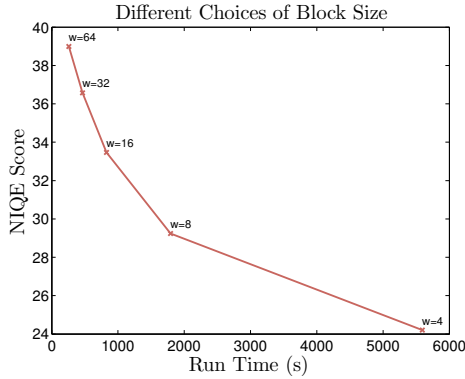
Fig. 4. Plot of average NIQE score versus EMISAC running time on the $1024 \times 1024 \times 100$ voxel SNEMI3D portion of the mouse cortex volume. The NIQE scores for all x-y, x-z, and y-z cross sections within the volume are averaged. w stands for the block size of (w, w). A lower NIQE score corresponds to higher image quality. The optimization was run in all cases with a stopping threshold of $f = 10^{10}$. These NIQE scores are consistent with the higher rate of visually apparent artifacts present when larger block sizes are used, which provides some confirmation of the validity of the NIQE score.
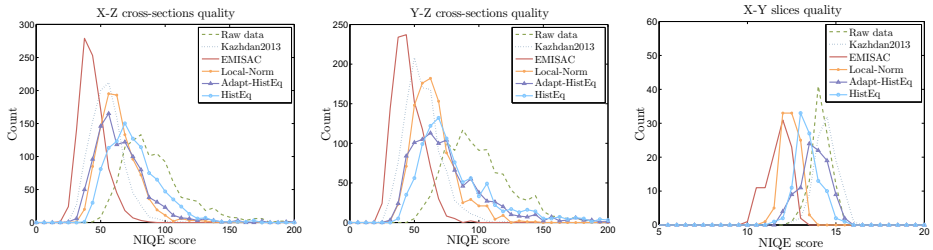

Fig. 5. Histogram of NIQE scores for the SNEMI3D volume. EMISAC uses a block size of $(16, 16)$ and stopping criteria of $f = 10^{10}$. Lower scores are better.

4.2 Comparison of Segmentation Accuracy

For the original data and each correction method, we evaluate the segmentation accuracy on each of the 4 boundary training/agglomeration training/test splits of the SNEMI3D volume. The Gala segmentation algorithm has a threshold parameter that trades off between false merges and false splits, as shown in Fig. 7; the optimal trade-off depends on the particular application, but in order to summarize results, we simply compute the minimum VI score (which gives equal weight to false splits and merges) over all thresholds.[4] For each correction method on each split, we compute the percent decrease in *minimum* VI score relative

---

[4] In actual use, we would have to pick the threshold based on cross-validation, as there would be no way to determine the true VI score for each threshold. However, this added complexity is irrelevant to our evaluation of artifact correction methods.
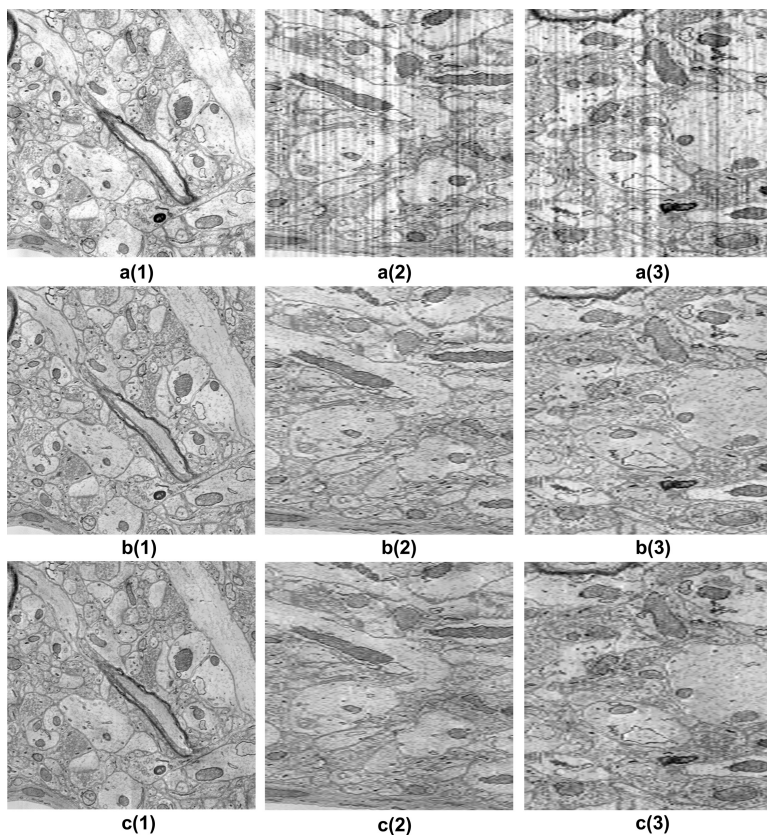
Fig. 6. Visual comparison of (a) the original data, and the corrected versions using (b) Kazhdan2013 and (c) our method EMISAC. From left to right we have the (1) x-y cross section, (2) x-z cross section, and (3) y-z cross-section. The two correction algorithms produce visually very similar results.

to the original data. To compare methods, we compute the mean and standard deviations of these decrease percentages. The results are shown in Table 2.

The nearly exact match in x-y NIQE scores between the original data and Kazhdan2013, as shown in Fig. 5 and Table 1, can be explained by the fact that Kazhdan2013 essentially copies the high-frequency content of the original x-y slices in its final step, and NIQE scores depend only on local (high-frequency) information.

While the NIQE scores were relatively insensitive to the stopping threshold f, we observed that the segmentation accuracy was highly sensitive, and therefore computed results for $f \in \{10^{10}, 10^9, 10^7, 10^6\}$, corresponding to increasing segmentation accuracy. Both Kazhdan2013 and EMISAC (for $f \leq 10^9$) achieve a similarly large improvement in accuracy over the original data. The difference between the two methods is not statistically significant ($p = 0.87$).

Table 1. NIQE score reduction (improvement) percentage, averaged over all x-y, x-z, and y-z cross-sections in the two EM datasets. The average structural similarity index (SS) [36] (as a percentage) between the original and corrected x-y cross-sections is also shown. Higher percentages are better. *First column:* Six $1024 \times 1024 \times 100$ voxel volumes of the mouse cortex ATUM-SEM volume [20] (five randomly sampled volumes plus the SNEMI3D volume). *Second column:* $512 \times 512 \times 30$ *Drosophila* larva ventral nerve cord ssTEM volume [8].

	Mouse cortex volume				Drosophila larva ventral nerve cord volume			
	x-y	x-z	y-z	SS	x-y	x-z	y-z	SS
EMISAC	6.59 ±6.86	**41.21** ±19.25	**39.56** ±19.82	96.4 ±2.20	-1.15 ±2.83	**11.04** ±25.05	**11.82** ±23.27	97.4 ±1.06
Kazhdan2013	-1.85 ±0.82	35.24 ±20.28	35.06 ±20.02	98.3 ±1.69	-	-	-	-
Histogram Equalization	-4.53 ±6.02	-35.80 ±86.31	-42.31 ±95.04	69.5 ±6.47	-21.16 ±17.79	-68.54 ±88.76	-67.92 ±84.56	81.0 ±3.64
CLAHE	-7.59 ±6.19	-39.77 ±95.09	-40.46 ±96.56	70.3 ±3.73	-14.02 ±9.43	-127 ±150	-133 ±154	80.1 ±3.76
Local Normalization	2.51 ±4.91	13.04 ±27.14	13.79 ±27.20	93.3 ±3.52	3.05 ±1.93	-6.38 ±26.72	-8.51 ±28.85	97.4 ±1.58

Figure 4 suggests that better results may be possible by using a block size smaller than $w = (16, 16)$, which was chosen for convenience in running experiments given the speed of our implementation.

We report running times of our Python implementation for comparison purposes, but by no means expect them to be comparable to those of a highly-optimized CPU or GPU implementation. Furthermore, L-BFGS-B is by no means the most effective algorithm for optimizing Eq. (3). The focus of our work was in evaluating correction models, rather than implementation speed.

Convolutional neural networks have shown good performance for neurite boundary detection [32,17,10], and may well perform better than the boundary classification method we used for our segmentation evaluation. Our choice was motivated by the fact that the state-of-the-art 3-D convolutional neural network approach for this problem is currently far from a settled matter, and we believe our method to be similar in performance; furthermore, it would have been highly impractical to spend the several weeks to months[5] of GPU time required to the train the network for each variant and data split that we tested.

---

[5] State-of-the-art 2-D networks often require several weeks of GPU training [10,25]; a comparable 3-D network can be expected to take at least as long, and possibly several times longer due to the larger number of parameters.
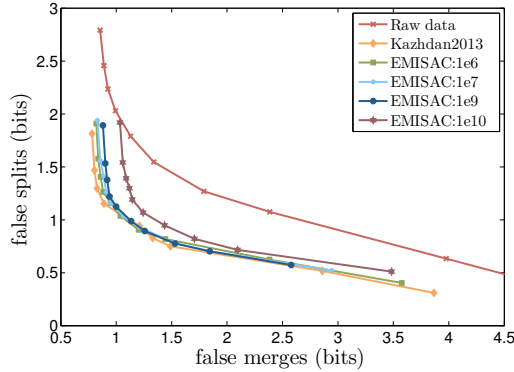
Fig. 7. Plot of the $H(S|U)$ (false split) vs. $H(U|S)$ (false merge) trade-off for segmentations based on the original image volume, Kazhdan2013, and EMISAC with several values of the stopping criteria f. Lower scores are better. Results are shown just for a single split, but results on other splits are similar. Note that the variation of information is simply $H(S|U) + H(U|S)$.

Table 2. Effect of artifact correction on segmentation accuracy ($n = 4$)

	Kazhdan2013	EMISAC			
		1e10	1e9	1e7	1e6
VI improvement(%)	29.19	19.83	26.43	27.23	27.97
	±13.12	±11.24	±10.03	±2.79	±1.88
Run Time (s)	-	828	1847	3963	4874

5   Lighting Correction of Time-Lapse Photography

Raw time-lapse photography sequences typically exhibit substantial flickering due to variations in lighting and exposure between frames [9]. Due to scene geometry, these variations are often local, such that a global brightness and contrast adjustment per frame is insufficient. We can directly apply EMISAC to the problem of correcting such lighting issues by treating time as the z-axis (taking the place of the section axis for the EM data); a separate set of α and β parameters are used for the red, green, and blue channels.

Quantitative evaluation of these time-lapse sequences cannot be done in the same way as for electron microscopy stacks, since the data distribution is obviously not invariant to transpositions between time and the x or y axis, as would be implied by comparing x-z and y-z cross-sections to the original x-y frames. In fact, we are unaware of any established method for quantitatively measuring lighting discontinuity in time-lapse sequences; while EMISAC's own objective function does measure this in some sense, it cannot reasonably be used for comparison to other methods nor can it be aggregated across datasets. Therefore, we are limited to qualitative assessment.

For all time-lapse sequences, we used a final block size of $(w_x = 4, w_y = 4)$ and manually set γ in the range of $[\frac{39 \times 10^2 w_x w_y}{d_x d_y}, \frac{39 \times 10^4 w_x w_y}{d_x d_y}]$; γ trades off preservation of detail and temporal smoothness, and is fundamentally a matter of user preference.

For evaluation, we used 8 publicly available time-lapse sequences that exhibited lighting discontinuities between frames. For several of these sequences, a demonstration result obtained by manual editing using the commercial LRTime-Lapse software was also available. The corresponding video files can be found at http://rll.berkeley.edu/2014_ECCV_EMISAC. Qualitatively, EMISAC essentially eliminates all flickering without reducing the apparent quality of individual frames. It appears to give a very similar quality result, in terms of correcting lighting discontinuities, to that obtained by manual editing with the specialized LRTimeLapse software.

6   Discussion

Imaging artifacts, most notably discontinuities along the section (z) axis, have so far limited the use of image volumes acquired by ATUM-based SEM, one of the most promising high-throughput volume electron microscopy techniques, to essentially 2.5-D analysis [34,10]. Our limited assumptions about the data and distortion process lead naturally to a simple but highly effective optimization-based procedure: our method EMISAC appears to eliminate all visible discontinuities, without any loss of intra-section detail. On the key task of neurite segmentation, EMISAC substantially improves accuracy relative to the original data by about 28%, matching the improvement achieved by the recent independently-developed alternative method Kazhdan2013 [23]. In terms of NIQE score [28], our method significantly outperforms Kazhdan2013. Furthermore, the significant qualitative improvement in the video results demonstrates the applicability of EMISAC to time-lapse photography.

One explanation for the superior NIQE scores attained by EMISAC compared to Kazhdan2013 may be that EMISAC supports both local brightness as well as local contrast correction. Kazhdan2013 solves two sequential optimization problems, both of which apply an additive correction term to the original data, and penalize deviations in intra-slice gradients of the correction term. This implicitly allows smooth changes in brightness, as the gradient of the correction term is only affected by the gradient of the brightness factor. Even constant changes in contrast, however, are penalized heavily, as they have a *multiplicative* effect on the gradient of the correction term. While the lower NIQE scores of EMISAC relative to Kazhdan2013 did not correspond to better segmentation accuracy in our experiments, the segmentation performance may have been limited by the quality of the feature representation, and a future segmentation algorithm may indeed show improvement.

References

1. Alexander, A.L., Tsuruda, J.S., Parker, D.L.: Elimination of eddy current artifacts in diffusion-weighted echo-planar images: the use of bipolar gradients. Magnetic Resonance in Medicine 38(6), 1016–1021 (1997)
2. Andersson, J.L., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20(2), 870–888 (2003)
3. Andres, B., Kroeger, T., Briggman, K.L., Denk, W., Korogod, N., Knott, G., Koethe, U., Hamprecht, F.A.: Globally optimal closed-surface segmentation for connectomics. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 778–791. Springer, Heidelberg (2012)
4. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
5. Berger, D.R., Schalek, R., Kasthuri, N., Tapia, J.C., Hayworth, K., Seung, H.S., Lichtman, J.W.: SNEMI3D challenge, `http://brainiac2.mit.edu/SNEMI3D/home` (training volume)
6. Briggman, K.L., Bock, D.D.: Volume electron microscopy for neuronal circuit reconstruction. Current Opinion in Neurobiology 22(1), 154–161 (2012)
7. Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., Hartenstein, V.: An integrated micro-and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. PLoS Biology 8(10), e1000502 (2010)
8. Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., Hartenstein, V.: Segmented serial section Transmission Electron Microscopy (ssTEM) data set of the Drosophila first instar larva ventral nerve cord (VNC) (2010), `http://www.ini.uzh.ch/~acardona/data.html`
9. Chylinski, R.: Time-Lapse Photography: A Complete Guide to Shooting, Processing and Rendering Time-Lapse Movies. Cedar Wings Creative (2012), `http://books.google.com/books?id=7fDaLPhJB5IC`
10. Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS, pp. 2852–2860 (2012)
11. Craddock, R.C., Jbabdi, S., Yan, C.G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P.: Imaging human connectomes at the macroscale. Nature Methods 10(6), 524–539 (2013)
12. Denk, W., Horstmann, H.: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biology 2(11), e329 (2004)
13. Haselgrove, J.C., Moore, J.R.: Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. Magnetic Resonance in Medicine 36(6), 960–964 (1996)
14. Hayworth, K., Kasthuri, N., Schalek, R., Lichtman, J.: Automating the collection of ultrathin serial sections for large volume tem reconstructions. Microscopy and Microanalysis 12(S02), 86–87 (2006)
15. Helmstaedter, M.: Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. Nature Methods 10(6), 501–507 (2013)

16. Helmstaedter, M., Briggman, K.L., Denk, W.: High-accuracy neurite reconstruction for high-throughput neuroanatomy. Nature Neuroscience 14(8), 1081–1088 (2011)

17. Jain, V., Bollmann, B., Richardson, M., Berger, D.R., Helmstaedter, M.N., Briggman, K.L., Denk, W., Bowden, J.B., Mendenhall, J.M., Abraham, W.C., et al.: Boundary learning by optimization with topological constraints. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2488–2495. IEEE (2010)

18. Jain, V., Murray, J.F., Roth, F., Turaga, S., Zhigulin, V., Briggman, K.L., Helmstaedter, M.N., Denk, W., Seung, H.S.: Supervised learning of image restoration with convolutional networks. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007)

19. Jain, V., Turaga, S.C., Briggman, K.L., Helmstaedter, M.N., Denk, W., Seung, H.S.: Learning to agglomerate superpixel hierarchies. Advances in Neural Information Processing Systems 2(5) (2011)

20. Kasthuri, N., Lichtman, J.W.: Mouse S1 cortex Automatic Tape-Collecting Ultra Microtome (ATUM)-based Scanning Electron Microscopy (SEM) volume (2011), http://www.openconnectomeproject.org

21. Kaynig, V., Fischer, B., Müller, E., Buhmann, J.M.: Fully automatic stitching and distortion correction of transmission electron microscope images. Journal of Structural Biology 171(2), 163–173 (2010)

22. Kazhdan, M., Burns, R., Kasthuri, B., Lichtman, J., Vogelstein, J., Vogelstein, J.: Color corrected mouse S1 cortex Automatic Tape-Collecting Ultra Microtome (ATUM)-based Scanning Electron Microscopy (SEM) volume (2013), http://www.openconnectomeproject.org

23. Kazhdan, M., Burns, R., Kasthuri, B., Lichtman, J., Vogelstein, J., Vogelstein, J.: Gradient-domain processing for large em image stacks. arXiv preprint arXiv:1310.0041 (2013)

24. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. The Journal of Neuroscience 28(12), 2959–2964 (2008)

25. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1106–1114 (2012)

26. Kuwajima, M., Mendenhall, J.M., Harris, K.M.: Large-volume reconstruction of brain tissue from high-resolution serial section images acquired by sem-based scanning transmission electron microscopy. In: Nanoimaging, pp. 253–273. Springer (2013)

27. Meilă, M.: Comparing clusterings–an information based distance. Journal of Multivariate Analysis 98(5), 873–895 (2007)

28. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Processing Letters 20(3), 209–212 (2013)

29. Nunez-Iglesias, J., Kennedy, R., Parag, T., Shi, J., Chklovskii, D.B.: Machine learning of hierarchical clustering to segment 2d and 3d images. PloS One 8(8), e71715 (2013)

30. Sage, D.: Local normalization filter to reduce the effect of non-uniform illumination (March 2011), http://bigwww.epfl.ch/sage/soft/localnormalization/

31. Schalek, R., Wilson, A., Lichtman, J., Josh, M., Kasthuri, N., Berger, D., Seung, S., Anger, P., Hayworth, K., Aderhold, D.: Atum-based sem for high-speed large-volume biological reconstructions. Microscopy and Microanalysis 18(S2), 572–573 (2012)

32. Turaga, S., Briggman, K., Helmstaedter, M., Denk, W., Seung, S.: Maximin affinity learning of image segmentation. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (eds.) Advances in Neural Information Processing Systems, vol. 22, pp. 1865–1873. MIT Press, Cambridge (2009)
33. Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H.S.: Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Comput. 22(2), 511–538 (2010)
34. Vazquez-Reina, A., Gelbart, M., Huang, D., Lichtman, J., Miller, E., Pfister, H.: Segmentation fusion for connectomics. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 177–184. IEEE (2011)
35. Wager, S., Wang, S., Liang, P.: Dropout training as adaptive regularization. In: Advances in Neural Information Processing Systems, pp. 351–359 (2013)
36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
37. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. ACM Transactions on Mathematical Software (TOMS) 23(4), 550–560 (1997)
38. Zuiderveld, K.: Contrast limited adaptive histograph equalization. Graphic Gems, pp. 474–485 (1994)