Superpixel quality in microscopy images: the impact of noise & denoising

J. Roels[1], J. De Vylder[2], J. Aelterman[2], S. Lippens[3], Y. Saeys[4] and W. Philips[2]

[1] Ghent University / VIB, Department of Telecommunications and Information Processing / Inflammation Research Center, Ghent, Belgium
[2] Ghent University, Department of Telecommunications and Information Processing, Ghent, Belgium
[3] VIB / Ghent University, Inflammation Research Center / Department of Biomedical Molecular Biology, Ghent, Belgium
[4] VIB / Ghent University, Inflammation Research Center / Department of Internal Medicine, Ghent, Belgium

Abstract— **Microscopy is a valuable imaging tool in various biomedical research areas. Recent developments have made high resolution acquisition possible within a relatively short time. State-of-the-art imaging equipment such as serial block-face electron microscopes acquire gigabytes of data in a matter of hours. In order to make these amounts of data manageable, a more data-efficient representation is required. A popular approach for such data efficiency are superpixels which are designed to cluster homogeneous regions without crossing object boundaries. The use of superpixels as a pre-processing step has shown significant improvements in making computationally intensive computer vision analysis algorithms more tractable on large amounts of data. However, microscopy datasets in particular can be degraded by noise and most superpixel algorithms do not take this artifact into account. In this paper, we give a quantitative and qualitative comparison of superpixels generated on original and denoised images. We show that several advanced superpixel techniques are hampered by noise artifacts and require denoising and parameter tuning as a pre-processing step. The evaluation is performed on the Berkeley segmentation dataset as well as on fluorescence and scanning electron microscopy data.**

Keywords— **Fluorescence microscopy, electron microscopy, segmentation, superpixels, denoising**

I. INTRODUCTION

Parkinson's and Alzheimer's disease, various types of cancer and other inflammatory diseases are examples of high-impact disorders that are currently being investigated extensively using microscopy. In essence, these diseases are caused by cellular disorders, making visualization of ultrastructural cell content a crucial step. Fluorescence and electron microscopy (FM and EM, respectively) allow biomedical users to do this by generating nanometer and even Angström resolution images.

One of the consequences of high resolution microscopy is that the user is forced to capture hundreds of millions of pixels in order to visualize whole cells. An important step in biomedical image analysis is to segment this enormous amount of data, which requires advanced computer vision algorithms because of the complex ultrastructural content. However, users are often forced to low-complexity algorithms [1] (mostly leading to low-quality results) or even manual labour, because more advanced and robust algorithms (e.g. model based techniques [2]) are simply not tractable for this kind of datasets. For example, in [3] a team of 224 people manually annotated 950 neurons in a 1 million μm^3 EM dataset at nanometer resolution, leading to more than 20000 annotator hours. As a result, microscopy image analysis becomes a tedious and time-consuming process.

Additionally, FM and EM tend to introduce acquisition-dependent noise artifacts [4, 5]. In order to reduce this noise as much as possible, ordinary denoising algorithms assuming white, Gaussian noise will be suboptimal. Instead, a more explicit noise model as in [6] will give better denoising results.

To make advanced segmentation algorithms usable, microscopy datasets can be represented as a connected graph of superpixels instead of pixels [7–9]. Superpixels are defined as a coherent set of *similar* pixels, where the similarity between pixels usually is defined in terms of intensity or texture. They are ideally designed not to cross object boundaries and are therefore excellent to limit the large number of pixels to a much smaller number of superpixels, without losing important image information. Boundary preservation is therefore a crucial property for these techniques. However, noise artifacts which are very common in microscopy imaging, typically cause uncertainty around object boundaries. Despite this fact, it has never been validated whether noise has a significant impact on superpixel quality in microscopy images.

In this paper, we illustrate the quality impact of state-of-the-art superpixel algorithms when confronted with different types of noise that appear in microscopy. We show that recent superpixel algorithms lose boundary information on noisy data and require a denoising pre-processing step in order to improve the latter. The results were derived on the Berkeley segmentation dataset [10] (BSD) and annotated FM and scanning EM (SEM) datasets.

The structure of this paper is as follows: in Section II, we

will give a description of the superpixel and denoising algorithms that were used in the experiments. Next, we will describe and discuss the performed experiments in more detail in Section III. Section IV finalizes this paper with a conclusion.

II. BACKGROUND

Before giving an overview of the experiments, we will give a brief description of the superpixel and denoising techniques that were evaluated.

A. Superpixels

A superpixel segmentation can be described as a *controlled* oversegmentation of an image. This *control* is usually implemented as e.g. desired intensity or texture similarity within superpixels, boundary preservation, similar size or shape, etc.

Superpixel algorithms can be categorized in two types, based on the superpixel compactness. Compact superpixels are usually described as nearly circularly shaped segments of approximately the same size. Techniques producing this kind of superpixels such as [11–13] force homogeneity in the superpixel shapes and sizes. Non-compact superpixel algorithms such as [14, 15], allow shape and size variability. Compact techniques have the advantage of superpixel size flexibility and shape consistency, which is convenient for developing e.g. subsequent segmentation algorithms. However, in some cases, the compactness constraint is too strong, resulting in a loss of boundary preservation. Non-compact algorithms may solve this problem. The downside of these algorithms is that they are sensitive to *leaking*, when an edge is less apparent (due to e.g. noise or blur) and allows a segment to 'leak through' the object boundary. Therefore, the choice of compact or non-compact superpixel algorithms is application-dependent in practice. In our experiments, we will compare the most recently developed compact and non-compact superpixel algorithms: Simple Linear Iterative Clustering (SLIC, [13]), TurboPixels (TP, [12]) and graph-based superpixels (GB, [15]).

Since our main focus here does not lie on the techniques themselves, we will only give a brief algorithmic description. For more in-depth explanation, we refer the reader to the respective references. SLIC [13] is a k-means clustering based algorithm, producing a number of compact superpixels by limiting the search space proportional to the desired superpixel size and combining intensity and spatial distance measures. TP [12] produces a number of compact superpixels
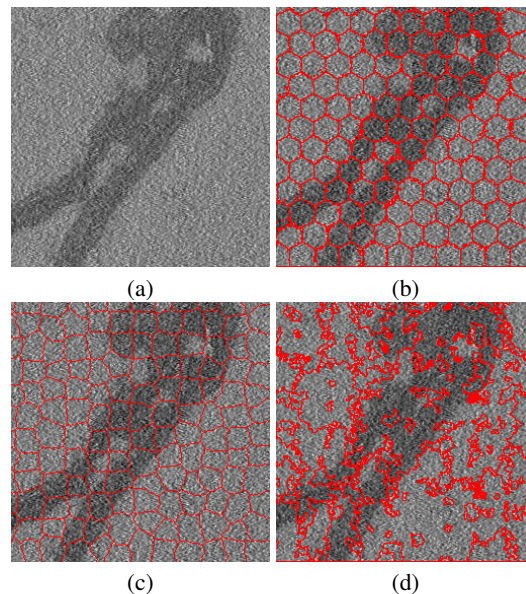

Fig. 1: (a) A cropped EM image, (b) SLIC, (c) TP and (d) GB superpixels. Both TP and SLIC produce compact superpixels as opposed to GB which produces superpixels of variable sizes and shapes.

by iteratively extending clusters according to the geometric-flow, which is attracted to edges, but smaller in magnitude in their vicinity. GB [15] forms non-compact superpixels by representing an image as a weighted graph and merging two subgraphs by evaluating if there is a boundary between the two subgraphs. Figure 1 shows the segmentation results of SLIC, TP and GB on a cropped EM image.

B. Denoising

Denoising is an image restoration technique where the latent image has to be restored from an acquired, noisy image. In general, one can distinguish between local [16], non-local [17, 18] and regularized probabilistic [19] denoising algorithms.

In our experiments, we will use our previously developed NLMS-SC algorithm [6], which is based on the non-local means algorithm [17] and specifically designed to handle SEM noise (i.e. signal-dependent and correlated). Moreover, it is also possible to handle signal-dependent (FM) or even white noise using a special case of NLMS-SC. The algorithm restores a noisy pixel by computing a weighted average of pixels with similar regions, where the weights are pre-whitened to cancel out correlation and signal dependency artifacts. Figure 2 illustrates denoising in EM with NLMS-SC.
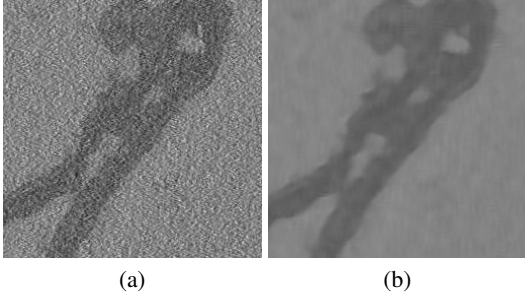
Fig. 2: (a) A cropped EM image and (b) the restored image with NLMS-SC.

Fig. 3: SLIC superpixels (left) on noisefree data ($m = 60$), (middle) on noisy data with compactness adjustment ($m = 140$), (right) on noisy data without compactness adjustment ($m = 60$).

III. EXPERIMENTS AND DISCUSSION

In our experimental setup, we wish to evaluate superpixel quality with noisy and denoised input data. This requires a quantitative evaluation metric that expresses the desired superpixel properties.

A. Evaluation metrics

Quantitative superpixel evaluation metrics should express how well the superpixel boundaries preserve object boundary information. For example, in a superpixel-based segmentation application, the detected segments will be less accurate whenever superpixels are crossing object boundaries. We will express boundary preservation in terms of a Hausdorff distance variant [20]. Given a groundtruth segmentation $G = \{g_1, g_2, \ldots, g_n\}$ and a superpixel segmentation $S = \{s_1, s_2, \ldots, s_m\}$, where g_i and s_j are boundary pixels corresponding to segmentation or superpixel partitions (i.e. the red borders in Figure 1). The proposed distance $d(g, S)$ between a groundtruth boundary pixel $g \in G$ and a superpixel partition S is defined as the Euclidean distance of g to its closest superpixel boundary pixel:

$$d(g, S) = \min_{s \in S} d_E(g, s) \quad (1)$$

where d_E is defined as the spatial Euclidean distance. A superpixel segmentation S will be validated on a complete groundtruth segmentation G based on the histogram of $d(g, S)$ for all $g \in G$. High-quality superpixel segmentations, preserving boundary information will have more superpixel edges in the vicinity of groundtruth boundaries and less groundtruth boundary pixels that are not in the vicinity of superpixel edges. Therefore, high-quality superpixel segmentations will correspond to histograms that are more skewed towards zero compared to low-quality superpixels.
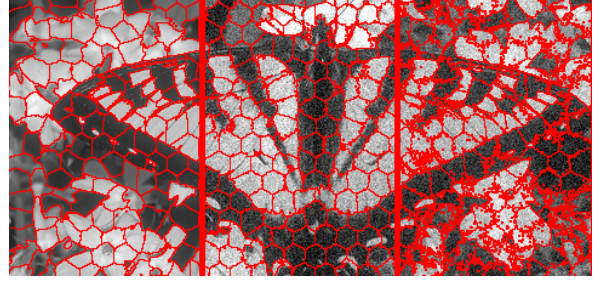
B. Experiments

B1 Experimental setup

We will evaluate superpixel quality on noisy and denoised data both visually and quantitatively. Therefore, we obtained manually annotated Fluorescence Microscopy (FM) and Scanning Electron Microscopy (SEM) data. The FM and SEM datasets consist of 13 1388×1040 and 112 1000×1100 images, respectively. Recent research has shown that these types of microscopy are typically degraded by signal-dependent and signal-dependent, correlated noise, respectively [4,5]. As a third dataset, we used the Berkeley Segmentation Dataset (BSD) consisting of 500 manually segmented images which we artificially corrupted with white, Gaussian noise.

B2 Parameter settings

We should note that, as opposed to TP and GB, SLIC is provided with a compactness parameter m. SLIC uses an adjusted distance metric that is a weighted sum of spatial and intensity distance. The parameter m assigns a relative importance to these two. As a consequence, high values for this parameter will enforce the superpixel boundaries to form a Voronoi diagram, i.e. they become more compact. Noise, however, typically causes artifacts on superpixel boundaries, making them highly irregular (Figure 3). In order to preserve compactness, we have to increase the compactness parameter on noisy input data. The only question remaining is how much m should be increased to get similar compactness. We express the compactness of a superpixel $S' \subset S$ of a superpixel segmentation S quantitatively in terms of stockiness $\left(\frac{4 \times \pi \times \text{Area}(S')}{\text{Perimeter}(S')^2} \right)$ where $\text{Area}(S')$ and $\text{Perimeter}(S')$ denote the area and perimeter induced by the set of superpixel boundary pixels S'. Superpixels having stockiness equal to 1 correspond to perfectly circular shaped and therefore compact superpix-
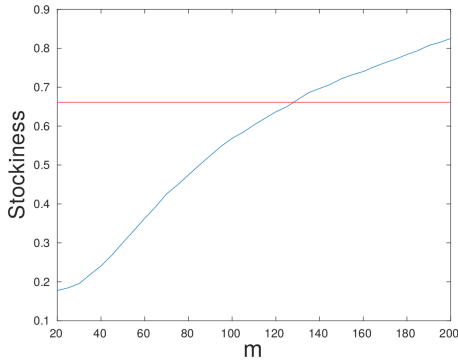
Fig. 4: Mean SLIC superpixel stockiness for variable m on noisy data (blue) and the mean stockiness of superpixels computed on noisefree data where $m = 60$ (red).

els, whereas low stockiness indicates non-compactness.

Figure 4 illustrates the impact of the parameter m on superpixel stockiness in a noisy image. The noisy image used in Figure 3 was employed for this purpose. The red line corresponds with the mean stockiness of superpixel segments on the original image (i.e. the left part of Figure 3), where the compactness parameter was set to 60. The blue line indicates the mean stockiness of superpixels when computed on noisy data for various values of m. The intersection of the lines indicates how the compactness parameter should be set, in order to get similar stockiness (or — implicitely — compactness) on noisy and noisefree images. This way, based on a limited training set of images from all three datasets, we determined the value for m whenever we are computing SLIC superpixels on noisy data.

B3 Superpixel quality with noisy and denoised data

We wish to evaluate superpixel quality with noisy and denoised input data. In practice, most superpixel algorithms ignore the presence of noise in images. As a consequence, it is expected that several techniques might not be noise-robust and require a denoising step in advance. For this, we denoised FM and SEM images using the NLMS-SC algorithm and compared SLIC, TP and GB superpixel quality between noisy and denoised data.

Rows one through four in Figure 5 illustrate superpixels on noisy SEM, denoised SEM, noisy FM and denoised FM data, respectively. Rows five and six illustrate the impact of white Gaussian noise ($\sigma = 20$) on an image taken from the BSD dataset[1]. Notice we are considering three types

---

[1]For visualizations purposes, only crops of the results are shown. A more extensive visualization can be found at http://telin.ugent.be/~jbroels/medicon_2016/
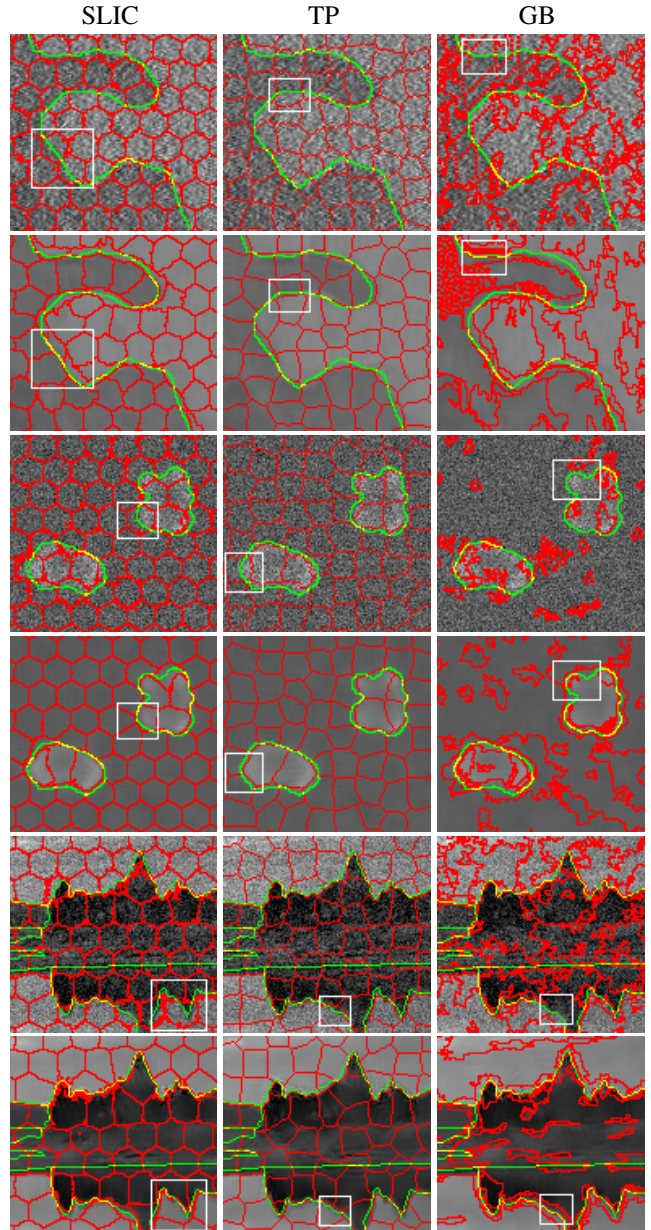

Fig. 5: SLIC, TP and GB superpixels (red) in noisy and denoised images. Rows one and two illustrate superpixel quality on noisy and denoised SEM data, respectively. Rows three and four illustrate superpixel quality on noisy and denoised FM data, respectively. Rows five and six illustrate superpixel quality on noisy and denoised BSD images, respectively. Groundtruth segmentation borders are annotated in green, overlapping boundaries in yellow.

of noise: correlated, signal-dependent in case of EM, signal-dependent in case of FM and white Gaussian noise in case of BSD. Superpixels should preserve boundary information as
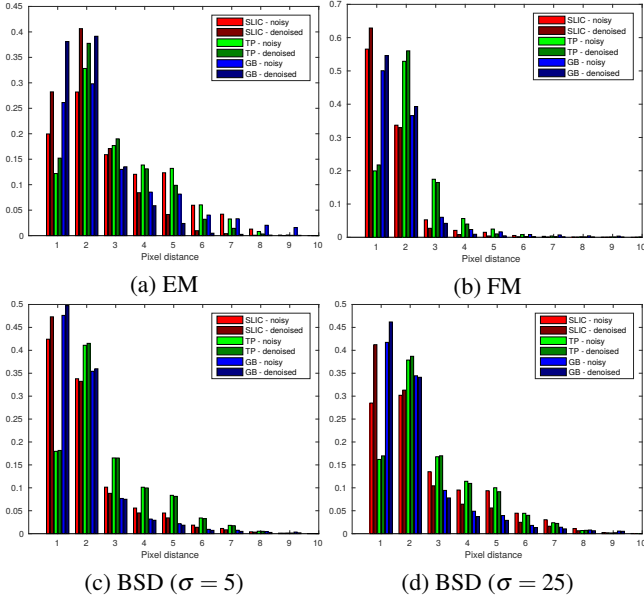
Fig. 6: Hausdorff distance distributions for superpixel segmentations computed on noisy and denoised versions of the (a) EM, (b) FM and (c-d) BSD dataset. For the BSD dataset, we artificially added white, Gaussian noise with standard deviation (c) $\sigma = 5$ and (d) $\sigma = 25$.

much as possible. In order to preserve stockiness, superpixels are forced to a more compact shape, leading to inaccurate boundary preservation. After a denoising step, a lower compactness parameter value will result in more shape variability and more accurate boundary preservation. This is clearly noticeable throughout all the SLIC results. The compactness constraint in TP is stronger, leading to more robustness to noise. Nevertheless, boundary information is better preserved in denoised data. The GB algorithm provides non-compact superpixels. Especially in the noisy data this results in highly irregular superpixel shapes, sizes and leaking. However, in the denoised data this seems to be less apparent. Denoising improves superpixel quality significantly, especially because there are much less superpixels leaking through groundtruth borders.

Figure 6 shows a quantitative evaluation of our experiments. For each dataset and superpixel technique, the proposed distance (Equation 1) between each groundtruth boundary pixel and superpixel segmentation was computed for each image in the dataset. For each image in the dataset, these distances are merged and shown in a histogram. Figure 6 visualizes this histogram for each dataset (both noisy and denoised) and superpixel method. Because the BSD dataset images originally are not inflicted by any significant amount of noise, we artificially added white, Gaussian noise

of both low and high variance. A general remark is that the histograms of the denoised data are more skewed towards zero compared to the histograms of the noisy data. In other words, denoising improves superpixel segmentation in terms of boundary preservation. For SLIC and GB there are significantly more groundtruth boundary pixels within a small distance of a superpixel boundary pixel and significantly less that have no superpixel boundary pixel in their vicinity. This is much less the case with TP. Figure 5 confirms this, as the difference between superpixels on noisy and denoised data is less significant, compared to SLIC and GB. We also notice that denoising becomes more important for higher amounts of noise, since the histograms corresponding to the denoised data are significantly more skewed towards zero in Figure 6d compared to the corresponding histograms in Figure 6c.

The histograms in Figure 6 illustrate the general importance of denoising when generating superpixels on noisy data. However, it is still possible that some superpixels have been shifted towards groundtruth boundary pixels, and other have neglected boundary preservation (due to e.g. edges being blurred by the denoising procedure). To take this into account, we compute the relative improvement r_g of the corresponding distances with and without a denoising step:

$$r_g = \frac{d(g, S_{noise})}{d(g, S_{denoised})} - 1, \tag{2}$$

where $g \in G$ is a groundtruth boundary pixel and S_{noise} and $S_{denoised}$ correspond to superpixel segmentations on noisy and denoised data, respectively. Note r_g can also be negative: in this case denoising increases the distance and has a negative impact on boundary preservation. Figure 7 shows the relative distance improvement histograms for the same datasets as in Figure 6. We should note that for these histograms only those $g \in G$ were taken into account such that $d(g, S_{noise}) > 2$, since we experienced that for these pixels typically $d(g, S_{denoised}) \leq 3$ and relative improvements would not be very informative (whether $d(g, S_{denoised})$ is 0, 1 or 2, in practice this is within manual precision). Secondly, the histograms are saturated at 4, meaning these groundtruth boundary pixels correspond to higher values than 4. Most of these values are ∞ because the corresponding groundtruth boundary pixels become perfectly detected ($d(g, S_{denoised}) = 0$ in Equation 2). In every dataset, we notice a negligible amount of negative relative improvements, compared to the positive ones. In other words, denoising improves boundary preservation significantly and only a negligible amount of groundtruth boundary pixels suffer from denoising and become less accurately detected by the superpixel algorithm.

(a) EM

(b) FM

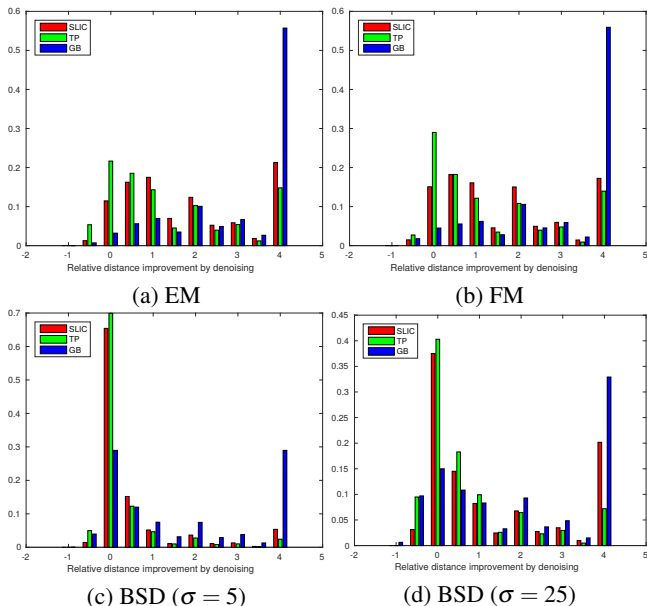(c) BSD ($\sigma = 5$)

(d) BSD ($\sigma = 25$)

Fig. 7: Relative distance improvement histograms for the same datasets as in Figure 6. The amount of groundtruth boundary pixels having a negative relative improvement is negligible compared to the positive ones.

IV. CONCLUSION

Superpixel algorithms are an essential step in making large amounts of data more tractable for computationally intensive computer vision algorithms and become indispensable in high-throughput biomedical imaging processes. However, noise causes quality issues for state-of-the-art superpixel techniques such as SLIC, TP and GB. We validated that denoising in advance and parameter adjustment improves superpixel boundary preservation (visually as well as quantitatively). We established significant quality improvements for SLIC and GB, whereas for TP this improvement is more subtle due to its stronger compactness constraint. Both SLIC and GB produce the most boundary preserving superpixels. Whether to use SLIC or GB is largely depending on the application and the necessity for compact or non-compact superpixels.

REFERENCES

1. Bene Miroslav, Zitova Barbar. Performance evaluation of image segmentation algorithms on microscopic image data *Journal of Microscopy.* 2015;257:65–85.
2. Osher Stanley, Paragios Nikos. *Geometric Level Set Methods in Imaging, Vision, and Graphics.* Springer 2003.
3. Helmstaedter Moritz, Briggman Kevin L., Turaga Srinivas C., Jain Viren, Seung H. Sebastian, Denk Winfried. Connectomic reconstruction of the inner plexiform layer in the mouse retina *Nature.* 2013.
4. Jezierska Anna, Talbot Hugues. Poisson-Gaussian Noise Parameter Estimation in Fluorescence Microscopy Imaging *IEEE International Symposium on Biomedical Imaging.* 2012:1663–1666.
5. Roels Joris, Aelterman Jan, De Vylder Jonas, et al. Noise Analysis in 3D Electron Microscopy in *Proc. 5th Dutch Bio-Medical Engineering Conference* 2015.
6. Roels Joris, Aelterman Jan, De Vylder Jonas, et al. Noise Analysis and Removal in 3D Electron Microscopy *Lecture Notes in Computer Science (Advances in Visual Computing).* 2014:31–40.
7. Lucchi Aurélien, Smith Kevin, Achanta Radhakrishna, Lepetit Vincent, Fua Pascal. A Fully Automated Approach to Segmentation of Irregularly Shaped Cellular Structures in EM Images. *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2010;13:463–471.
8. Jain Viren, Turaga Srinivas C.. Learning to Agglomerate Superpixel Hierarchies *Advances in Neural Information Processing Systems.* 2011:1–9.
9. Navlakha Saket, Ahammad Parvez, Myers Eugene W. Unsupervised Segmentation of Noisy Electron Microscopy Images using Salient Watersheds and Region Merging. *BMC Bioinformatics.* 2013;14:294.
10. Arbeláez Pablo, Maire Michael, Fowlkes Charless, Malik Jitendra. Contour detection and hierarchical image segmentation *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2011;33:898–916.
11. Moore Alastair P., Prince Simon J. D., Warrell Jonathan, Mohammed Umar, Jones Graham. Superpixel Lattices *Proc. IEEE Conference on Computer Vision and Pattern Recognition.* 2008:1–8.
12. Levinshtein Alex, Stere Adrian, Kutulakos Kiriakos N., Fleet David J., Dickinson Sven J., Siddiqi Kaleem. TurboPixels: Fast Superpixels using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2009;31:2290–7.
13. Achanta Radhakrishna, Shaji Appu, Smith Kevin, Lucchi Aurélien. SLIC Superpixels Compared to State-of-the-art Superpixel Methods *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2012;34:2274–2281.
14. Comaniciu Dorin, Meer Peter. Mean Shift: A Robust Approach Toward Feature Space Analysis *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2002;24:603–619.
15. Felzenszwalb Pedro F., Huttenlocher Daniel P.. Efficient Graph-Based Image Segmentation *International Journal of Computer Vision.* 2004;59:167–181.
16. Goossens Bart, Pižurica Aleksandra, Philips Wilfried. Image Denoising using Mixtures of Projected Gaussian Scale Mixtures *IEEE Transactions on Image Processing.* 2009;18:1689–1702.
17. Buades Antoni, Coll Bartomeu, Morel Jean-Michel. A Non-local Algorithm for Image Denoising in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*;2:60–65 vol. 2 2005.
18. Dabov Kostadin, Foi Alessandro. Image Denoising with Block-matching and 3D Filtering *Electronic Imaging.* 2006;6064:1–12.
19. Aelterman Jan, Goossens Bart. Combined Non-local and Multi-Resolution Sparsity Prior in Image Restoration *IEEE Transactions on Image Processing.* 2012:3049–3052.
20. Huttenlocher Daniel P., Klanderman Gregory A., Rucklidge William J.. Comparing images using the Hausdorff distance *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1993;15:850–863.