

SCIENTIFIC REPORTS



OPEN

Segmentation, Tracing, and Quantification of Microglial Cells from 3D Image Stacks

Mahmoud Abdolhoseini¹, Murielle G. Kluge^{2,3}, Frederick R. Walker^{2,3} & Sarah J. Johnson^{1,3}

Microglia play a central role in modulating synaptic structure and physiology, learning and memory processes. They exhibit morphological changes to perform these roles, therefore the morphological study of microglia can help to understand their functionality. Many promising methods are proposed to automatically segment the blood vessels or reconstruct the neuronal morphology. However, they often fail to accurately capture microglia organizations due to the striking structural differences. This requires a more sophisticated approach of reconstruction taking into account the varying nature of branch structures and soma sizes. To this end, we propose an automated method to reconstruct microglia, and quantify their features from 2D/3D image datasets. We first employ multilevel thresholding to segment soma volumes(3D)/areas(2D) and recognize foreground voxels/pixels. Seed points sampled from the foreground, are connected to form the skeleton of the branches via the tracing process. The reconstructed data is quantified and written in SWC standard file format. We have applied our method to 3D image datasets of microglia, then evaluated the results using ground truth data, and compared them to those achieved via the state-of-the-art methods. Our method outperforms the others both in accuracy and computational time.

Glial cells are a family of cells that perform highly specialized roles within the central nervous system of mammals. Microglia, astrocytes and oligodendrocytes are three major sub-types of this family. Microglia have recently been identified to play a central role in modulating synaptic structure, synaptic physiology as well as learning and memory processes¹⁻³. Microglia are not in permanent physical contact with other cells in their micro-environment. Therefore they need to engage in significant structural change to perform any of their known activities⁴. They exhibit distinct morphological transition states when undertaking specific biological functions. Therefore the morphological study of microglia is an important tool in understanding their functionality.

Manual reconstruction and quantification of the cells are very time consuming and prone to error, especially when dealing with 3D image stacks. Despite many promising automated methods in segmentation of blood vessels⁵, and reconstruction of neuronal morphology⁶⁻¹¹, our evaluation results reveal that these methods often fail to accurately capture microglia organizations due to the striking structural differences. Soma size and branch architecture vary considerably between neurons, microglia and astrocytes but most importantly the morphology among microglia cells is highly variable. Neurons resemble a relative stable and reliable tree structure with a primary process, the axon, and multiple dendrites representing the tree trunk and canopy, respectively. Microglial architecture is more diffuse and complex with multiple primary and secondary branches not confined to a generalizable structure. Additionally microglia morphology is highly diverse and changes by increasing and decreasing branches. This requires a more sophisticated approach of reconstruction taking into account the varying nature of branch structures and soma sizes.

Microglial cells consist of processes (branches) that form a tree structure and a central soma which envelops the cell nucleus. The easiest way to segment the outline of the cells is to distinguish the foreground from the background by calculating a single threshold based on the image intensity histogram¹². Using this approach, a local iterative thresholding has been proposed to reconstruct microglial morphology¹³. This method is fast and easy, however a single threshold is not enough to preserve the topology of cells. Depending on the threshold value, either many fine processes may be missed or many delicate gaps may be filled with noise or unwanted objects.

¹The University of Newcastle, School of Electrical Engineering and Computing, Callaghan, NSW, 2308, Australia.

²The University of Newcastle, School of Biomedical Sciences and Pharmacy, Callaghan, NSW, 2308, Australia. ³The Hunter Medical Research Institute, New Lambton, NSW, 2305, Australia. Correspondence and requests for materials should be addressed to M.A. (email: mahmoud.abdolhoseini@uon.edu.au)

A 2D segmentation of microglial cells via denoising and thresholding has been introduced¹⁴. A multifractal analysis has been performed on the segmented results to extract some spectrum features. Then a classifier is trained based on these features to classify microglial activation states. This approach is only applicable to 2D images, and the tree topology of the microglial cell is not considered. Therefore special features of the cells, such as branch points and branch length, cannot be truly quantified.

Discriminative dictionaries are another tool to recognize special patterns in images¹⁵. An over-complete dictionary via label consistent K-SVD¹⁶ has been learned to detect a sparse set of points (called seed points) that mostly lie on the image foreground¹⁷. A tree is grown for each cell by starting from a root node at the centre of its soma (previously segmented via a separate nuclear label) and connecting the seed points using geodesic distances and minimum spanning tree¹⁸. This approach requires accurate models (ground truth images) to learn a dictionary for each specific data set and the performance of the dictionary highly depends on the accuracy of the models. This approach outperforms open-curve active contour¹⁹ in which the foreground seed points are detected using a gradient vector flow field²⁰. However, it has been reported that 27.8% of the seed points output by the dictionary are in the background¹⁷, which indicates a high potential of false traces.

Minimal path techniques have been employed to trace tree structural morphologies^{21–23}. The tracing algorithm is started with a source point (manually chosen) on the tree structure and will stop at some end points which are either detected by a Harris detector²⁴ or chosen from image border pixels. Each path between the source point and an end point is traversed by fast marching method²⁵. This approach only segments an image containing a single cell from its border pixels and is very time consuming (computational time for a single structure in a synthetic image, size 220×300 is 94.94 seconds²²). It also requires a user to manually choose a source point.

In this paper we propose a novel approach to segment and trace the microglial cells, and quantify their features from 2D/3D image datasets. The segmentation of the soma, and background removal via multilevel thresholding is the first step of our algorithm. Then a tracing step is initiated with the centroid of the somas and connects the prioritized seed points extracted from the cell voxels to build a tree structure for each cell. The thickness of the traced skeleton is estimated as the next step. Finally, the output data is quantified to extract the cell features such as: number of primary branches and branch points, branch length, soma size, and cell size. The whole of the process is automated and results in an objective morphological analysis of microglia with high speed and accuracy. The reconstructed data is written in SWC file format, so that it can be visualized or quantified via many toolboxes designed to read this standard format. As the experimental results, we have applied our method to 3D image datasets of microglia and compared the results with those achieved via the state-of-the-art algorithms. The comparison shows our method outperforms the others in terms of accuracy (assessed via four different metrics), and computational time.

Method

Soma segmentation and background removal. The first step is to locate the somas and extract the cell structures by removing the background. Let's assume a 3D microscopy image stack of microglia (all the steps are applicable to 2D images as well) is converted to a gray-scale image in which foreground intensity is lower than background (i.e. dark cells on a light background). A 3D binary array, B with the same size as the image, I , is defined as follows:

$$B_t(x, y, z) = \begin{cases} 1 & I(x, y, z) \leq t; \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where t is a threshold value from the image intensity range. Each set of connected '1's in B makes a 3D object. The number of 3D objects in B that are larger than a minimum (a suitable size for this minimum is discussed later) at a threshold t is denoted by n_t . We are interested in a t in which these objects form the soma volumes with an acceptable precision. This threshold value is called *soma threshold*, denoted by t_s . In the following, we propose a process of counting objects in different threshold values to find t_s .

Intuitively as t increases, n_t will increase until it reaches the maximum. After that, growing t will decrease n_t due to objects starting to merge together. Tracking the changes of n with a proper set of ascending thresholds can help us to find t_s . This proper set can be obtained from the image intensity histogram using multi-level Otsu's method¹². Using this method, 20 thresholds are chosen to assure smooth changes in n . This is more valuable than an arbitrary range of equally spaced thresholds which may include thresholds at levels of no value in demarcating image voxels.

To consider changes in n , a normalized differential criterion, c is defined as follows:

$$c_i = \frac{n_{t_{i+1}} - n_{t_i}}{n_{t_{i+1}} + n_{t_i}}; \quad n_t > 0, \quad (2)$$

in which $t_i \in \{t_1, \dots, t_{20}\}$ is the set of 20 thresholds obtained by Otsu's method, and sorted in ascending order. Assume for the first q thresholds, n_{t_i} is growing, which means c_i is non-negative (i.e. $c_i \geq 0$ where $i = \{1, \dots, q\}$). In this range, the t_s is empirically found to be where n has the minimum growing rate, which means c is minimum. If c reaches the minimum, c_{\min} , in more than one place (e.g. m places), i.e. $c_{\min} = \min c_i$ where $i = \{p_1, \dots, p_m\}$, t_s is specified by the place where $c_{i+1} - c_i$, $i = \{p_1, \dots, p_m\}$ becomes minimum (Fig. 1(h)). The advantage of this approach is that it does not require the soma to be separately stained, however if this information is available, it can be used instead to locate the somas.

Also a threshold in which n reaches the absolute maximum can be considered as the *background threshold* denoted by t_b . Because after t_b objects start merging together and forming one large solid which covers all the

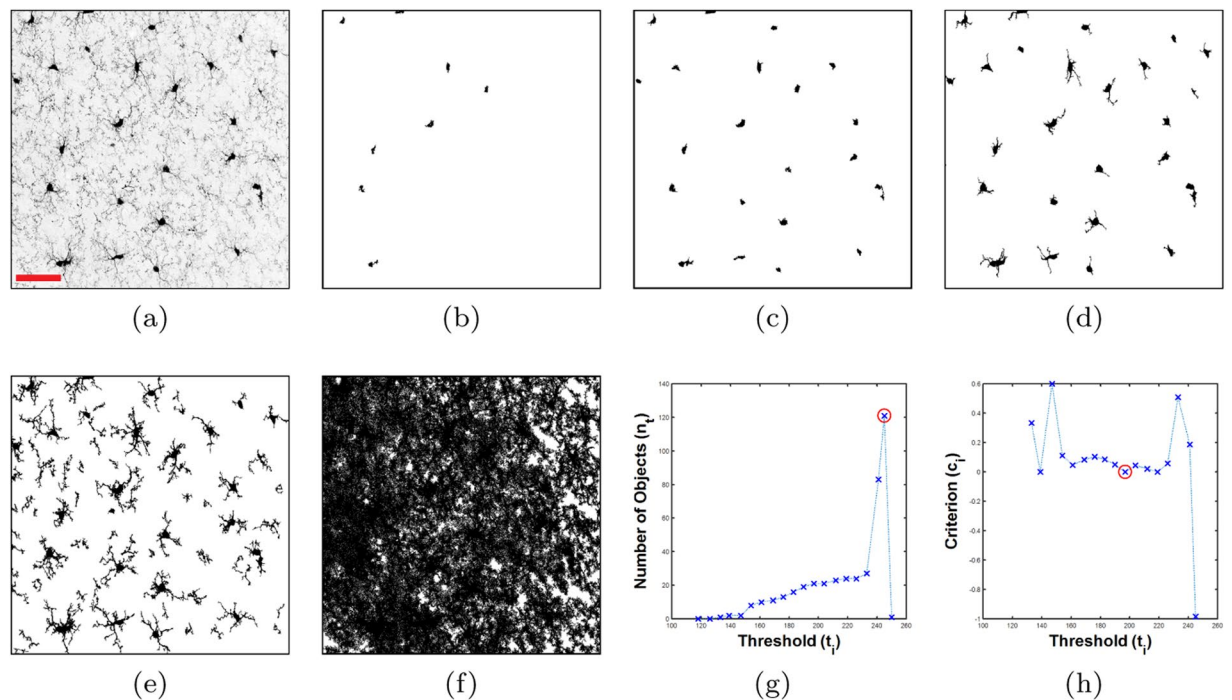


Figure 1. Soma segmentation and background removal using multiple thresholds. (a) Gray-scale MIP image, scale bar = 40 μm . (b) 2D MIP of B_{t_7} , $t_7 = 161$. (c) 2D MIP of the soma array, B_{t_s} , $t_s = t_{12} = 197$. (d) 2D MIP of $B_{t_{17}}$, $t_{17} = 233$. (e) 2D MIP of the cell voxel array, B_{t_b} , $t_b = t_{19} = 245$. (f) 2D MIP of $B_{t_{20}}$, $t_{20} = 250$. (g) Number of objects (n_i) vs. threshold (t_i), circled cross represents the background level. (h) Criterion (c_i) vs. threshold (t_i), circled cross represents the soma level.

image array (Fig. 1(f)). Therefore voxel intensities larger than t_b have no value in demarcation of the foreground and will not participate in the cell tracing process. One way to find t_b is to calculate n for all the intensities and then find the absolute maximum. This is a slow process specially when the size of the image array is large. Another way of finding t_b is to use the values of n that have been already calculated for the threshold set, $\{t_1, \dots, t_{20}\}$. Assume t_{\max} is a member of this set for which $n_{t_{\max}}$ becomes maximum, so t_{\max} will be either equal or very close to t_b . Therefore, instead of searching all the intensities to find t_b , we only need to search a small neighbourhood around t_{\max} . This makes it much faster to locate t_b for which n_{t_b} is the absolute maximum (Fig. 1(g)).

To accelerate the process of finding t_s and t_b , a 2D version of the image I is defined as follows:

$$M(x, y) = \underset{z}{\text{mjin}} I(x, y, z), \quad (3)$$

where mjin is the minimum along z direction. M is defined as a maximum intensity projection (MIP) of a 3D image stack when the foreground intensity is lighter (has larger value) than the background. Here, the reverse is true and the foreground is darker, therefore in (3) the minimum is applied instead of the maximum. For the purpose of finding t_s and t_b only a small thickness portion of I is used to create M . Specially it is important to consider a portion with the width of one cell approximately to avoid overlapping of too many cells (we use a thickness of 30 μm).

Definition (3) allows to find t_s and t_b through the 2D array M instead of the 3D image array I . This simplifies the process of counting objects from 3D to 2D and accelerates it a lot. To avoid counting noise or small isolated branches, we empirically found out only objects greater than 16 μm^2 needed to be counted in this process. Note, although t_s and t_b are found using the 2D MIP, M , the full 3D image is thresholded to obtain B as shown in (1).

Voxels with intensities between $\{t_s, \dots, t_b\}$ provide us with all the valuable information needed to trace paths through the centrelines of the processes. This is the advantage of multiple thresholds which provide several binary arrays, i.e. $B_{t_s}, B_{t_{s+1}}, B_{t_b}$, that participate in reconstruction: B_{t_s} shows all the soma volumes (Fig. 1(c)), B_{t_b} illustrates all the cell voxels (Fig. 1(e)), and B_{t_i} , $i \in \{s+1, \dots, b-1\}$ include all the branch details that are necessary for tracing.

Creating prioritized seed points. Only a small portion of the cell voxels (called seed points) are needed for the tracing stage. To extract the seed points, we define another binary array, L for the chosen threshold levels as follows:

$$L_i = B_{t_i} - B_{t_{i-1}}, \quad i \in \{s+1, \dots, b\}. \quad (4)$$

In each L_i , '1's specify places where voxel intensities of the image are between t_{i-1} and t_i . These voxels are sampled to create the seed points.

In general, the closer a voxel is to the centreline of the processes, the darker intensity it has. Therefore, the target is to sample the voxels with low intensities in each level to be assured of smooth tracing on the centreline of the processes. Hence, the actual intensities of the cell voxels are needed during the sampling. To this end, \hat{L}_i is created by element-wise multiplication of L_i by I , which makes the elements of \hat{L}_i be equal to the image elements where $L_i(x, y, z) = 1$, and '0' otherwise.

The seed points are created by first dividing \hat{L}_i into small volumes. Then the minimum intensity (i.e. darkest) voxel of each volume that consists of nonzero elements, is selected as a seed point. The size of the small volume specifies the sampling rate, the smaller the volume the higher the sampling rate. The sampling rate must be high enough to preserve the delicate curvy shape of the branches (in our experience, one seed point per $5\mu\text{m}^3$ cube is sufficient).

In addition to the seed points sampled from the \hat{L} arrays, a set of seed points sampled from soma surfaces is needed to assure us of tracing all the primary branches (branches that are directly connected to the soma). The surface of each soma is extracted from B_i and sampled (using the same procedure mentioned above) to form the set of seed points that initializes the tracing algorithm.

Tracing algorithm. The seed points are sampled from different levels of intensity, starting from the soma surface with the lowest intensity to the last significant level, L_b , with the highest intensity. For each cell, the tracing process is started with the centroid of the soma, 'root point', and ended with the last level, L_b . This helps to direct the path through the centreline of the processes and smoothly reconstruct the skeleton of the branches. Each cell is traced individually inside a defined boundary to build a tree-like structure. The default boundary is set to be large enough (a cube with $60\mu\text{m}$ each side) to cover all the microglial cell structure.

A connected set is defined as a set that consists of the seed points already joined to the cell tree structure during the tracing process. A disconnected set is defined as a set of seed points that are not part of the tree yet but are sufficiently close to an element in the connected set. The connected set is initialized to be the root point and all the seed points from the soma surface connected to it. The disconnected set is initialized with seed points that exist within a small volume around each seed point of the soma surface. The small volume size specifies a desired distance within which the seed points are reachable. Depending on the sampling rate, this size must be large enough (a cube of $40\mu\text{m}^3$ in our experiment) to assure all the seed points lying on the cell structures are reachable.

After initialization, a loop is designed to withdraw a seed point from the disconnected set and place it inside the connected set in each iteration. The chosen seed point is the one with a minimum geodesic distance to a member of the connected set. The geodesic distances are calculated between the members of the sets as follows. Assume 'k' is a member of the connected set for which we want to calculate the geodesic distances. B_b is cropped with the specified volume around 'k' to obtain an object which includes 'k'. Then Euclidean distances are calculated from 'k' to members of the disconnected set which belongs to this object. Those which do not belong to this object will be considered as the seed points with infinite distances and discarded.

After joining a new point to the connected set, the disconnected set will be updated by the new point's surrounding seed points within the specified volume. Then the same process will be repeated in the next iteration to specify a new connection. The tracing process will be continued until all the reachable seed points that belong to the first level, L_{s+1} are connected. Then, it steps up to the next level, L_{s+2} and connects all the reachable seed points in this level to the tree with the same procedure. After connecting all the reachable seed points in each level, it steps up to the next higher level until all the reachable seed points from all the levels are connected. The entire process is automated with zero user interaction required.

During the process of tracing centrelines, many small branches are added to the cell tree structures. A pruning stage is implemented after the tracing to remove branches shorter than a specified minimum. We set this minimum to $5\mu\text{m}$ as microglial branches smaller than this length are difficult to resolve.

In 2D reconstruction, there might be some branches that start from a cell and go to another with the same intensity along the way due to the maximum intensity projection. In these cases, there must be another decision metric other than intensity to cut the branch at some point along its path. We found out the watershed algorithm^{26–28} is an appropriate tool to make such a decision. Note, this problem no longer exists in 3D reconstruction of the cells from 3D image stacks, since there is no projection and 3D branches have different intensity along the way. Therefore, the application of watershed is no longer recommended.

Finding the thickness of the branches. Although the pruned short branches are not part of the skeleton, they in fact represent the thickness (or the radius of the cross section) of the main branches that form the skeleton. Therefore their locations and lengths are employed to estimate the branch radii.

After the pruning process, all the points (except the root point) which form the tree structure of a cell can be divided into two categories. (1) Points to which small branches were attached before pruning, where the radii are defined as equal to the length of these branches. (2) Points which did not hold any extra small branches, where the radii must be estimated using the radii at the neighbouring points.

For each traced cell skeleton, the radius estimation starts from the root point and spans the tree from the root to the tip of the branches. The root point's children (points on the soma surface that are directly connected to the root point) are the starting points of the branches. If the radius is not specified at a starting point, the radius at its child location can be a good estimation, since branch thickness is generally a smoothly changing value along the length of a branch. Except the starting points, if the radius is not specified at any other point on the branch centreline, it is estimated by the average of the radii at its parent and child locations. After estimating the radii at all the points, they are stored in a radii vector.

Quantification. Some features that help neuroscientist to gain good insight into the cell functionality are extracted and quantified using the output of the proposed method. These quantified features are: number of primary branches, number of branch points, branch length, soma size, and cell size.

Primary branches are the main branches that are directly connected to the soma. Secondary and tertiary branches divide from primary as it grows. The number of primary branches equals the number of the root point's children which is an output in our method. A branch point is a place where the processes bifurcate. Therefore the number of bifurcations equals the number of points that have more than one child. Branch length is equal to the summation of all the segment lengths that form the cell tree structure.

Soma size is directly obtained from the B_{i_s} array that consists of the soma volumes (3D) or area (2D). Cell size is quantified by calculating the cell volume in 3D or cell area in 2D. The radii vector calculated in previous section, represents the radius of a branch cross section in 3D (or the length of a perpendicular line to the centreline in 2D) at each point that lies on the centreline of the branches. Therefore the volume, v , of a solid segment between each pair of points is calculated as follows:

$$v = \frac{\pi}{3} \times h \times (r_1^2 + r_1r_2 + r_2^2), \quad (5)$$

if it is 3D or the segment area, a :

$$a = h \times (r_1 + r_2), \quad (6)$$

if it is 2D, where r_1 and r_2 are the radii at the beginning and end points of the segment respectively, which are obtained from the radii vector, and h is the length of the segment. The cell size is equal to the summation of all the segment volumes/areas (starting from root point's children to the end points of the branches) calculated via (5) or (6), plus soma size.

Output in SWC file format. The traced path output has a tree structure for each cell and is written in SWC file format. SWC is a standard ASCII format including all the information necessary for quantification and visualization of the cells. Each point on the traced path is assigned a row with seven fields in the SWC reconstruction file. These fields are: index number, tag number (which specifies different parts of a cell, e.g. '1' is assigned to soma), three spatial coordinates, radius, and parent index²⁹. There are some tools publicly available to read, visualize, and quantify this format, such as: Neuromantic³⁰, SharkViewer³¹, Vaa3D³², and L-Measure³³.

The SWC format is also employed to assess the quality of our reconstruction. One way to evaluate the performance of a method and compare it to other methods is using DIADEM metric³⁴ to benchmark automated reconstructions against manual (ground truth). DIADEM is a multi-step scoring process designed for tree-like structures, which takes into account the connections between points, bifurcation coordinates, excess and missing branches. Depending on the dataset, DIADEM requires some parameters to be set before running the metric. For all of our experiments, we set these parameters as follows: `xy-path-thresh = 1`, `z-path-thresh = 1`, `terminal-path-threshold = 20`, `terminal-threshold = 15`, `xy-threshold = 10`, `z-threshold = 8`, and left the remainder as default.

Another way of performance evaluation is to measure spatial Euclidean distances between the nodes of two reconstructions (automated and ground truth). These measures are spatial distance (SD), substantial spatial distance (SSD), and the percentage of SSD nodes (SSD%)³². We calculate these measures via NDIST³⁵ with the default value of 2 voxels for the substantial distance. Both metrics receive SWC files to perform their evaluation.

Image acquisition. All experiments were carried out in 6–8 weeks old male Cx3CR1^{GFP/+} mice, expressing fluorescently labelled microglia. Animals were obtained from the Garvan Institute of Medical Research and Australian BioResources, Sydney. All experiments were approved by the University of Newcastle Animal Care and Ethics Committee, and conducted in accordance with the New South Wales Animals Research Act and the Australian Code of Practice for the use of animals for scientific purposes. High resolution confocal images of fixed brain sections were taken on a Leica TCS SP8 confocal microscope with a Leica HC PLC APO 40x/1.30 OIL objective. Image stacks of the entire 30 μ m slices were taken at 0.2, 0.5 and 1.0 μ m z-step sizes.

Results

Two experiments are performed in this section: 1- reconstruction of 21 microglial cells and benchmarking the results against the ground truth images and comparing them to that of the state-of-the-art methods, 2- testing the ability of our method in reconstruction and quantification of large 3D image stack.

Performance evaluation. We have applied our method to 21 microglial images including 3D image patches of different sizes from $201 \times 201 \times 26$ to $401 \times 401 \times 108$ voxels. 18 of these images were imaged in our own lab, and three of them are available here¹⁷. We have benchmarked the quality of our reconstructions against ground truth images (manually reconstructed images by experts via Neuromantic³⁰) using 4 different metrics: SD, SSD, SSD%, and DIADEM to evaluate the efficiency of our method. The results have been compared to that of four state-of-the-art methods: Multiple Neuron Tracer (MNT)¹⁷, App2¹⁰, Simple Tracing (ST)¹¹, and Ensemble Neuron Tracer (ENT)⁹ (last three methods are implemented as plug-ins inside Vaa3D framework, version 3.200³²). The parameters in each application were tuned through an exhaustive search to get the best results as follows. Ensemble Neuron Tracer (ENT) and Simple Tracing (ST) do not require any parameter. As for APP2 the parameters were set as follows, `length_thresh = 5`, `cnn_type = 2`, and `background_threshold` was set to be automatically chosen since any changes result in invalid outputs. As for MNT the parameters were set as follows, `cost_threshold = 1000`, `intensity_threshold = 0.005`, and `contrast_threshold = 0.0003`. As

		Ours	MNT	App2	ST	ENT
Mean	SD	6.34	9.32	12.53	14.16	18.60
	SSD	8.43	11.74	15.54	16.14	19.36
	SSD%	0.67	0.69	0.76	0.80	0.91
	DIADDEM	0.57	0.44	0.39	0.15	NA
Std. error	SD	0.39	0.93	0.78	1.69	2.52
	SSD	0.37	1.02	0.91	1.61	2.22
	SSD%	0.02	0.02	0.01	0.03	0.05
	DIADDEM	0.03	0.04	0.04	0.02	NA

Table 1. Average and standard errors of evaluation results. MNT: Multiple Neuron Tracer¹⁷, App2¹⁰, ST: Simple Tracing¹¹, ENT: Ensemble Neuron Tracer⁹.

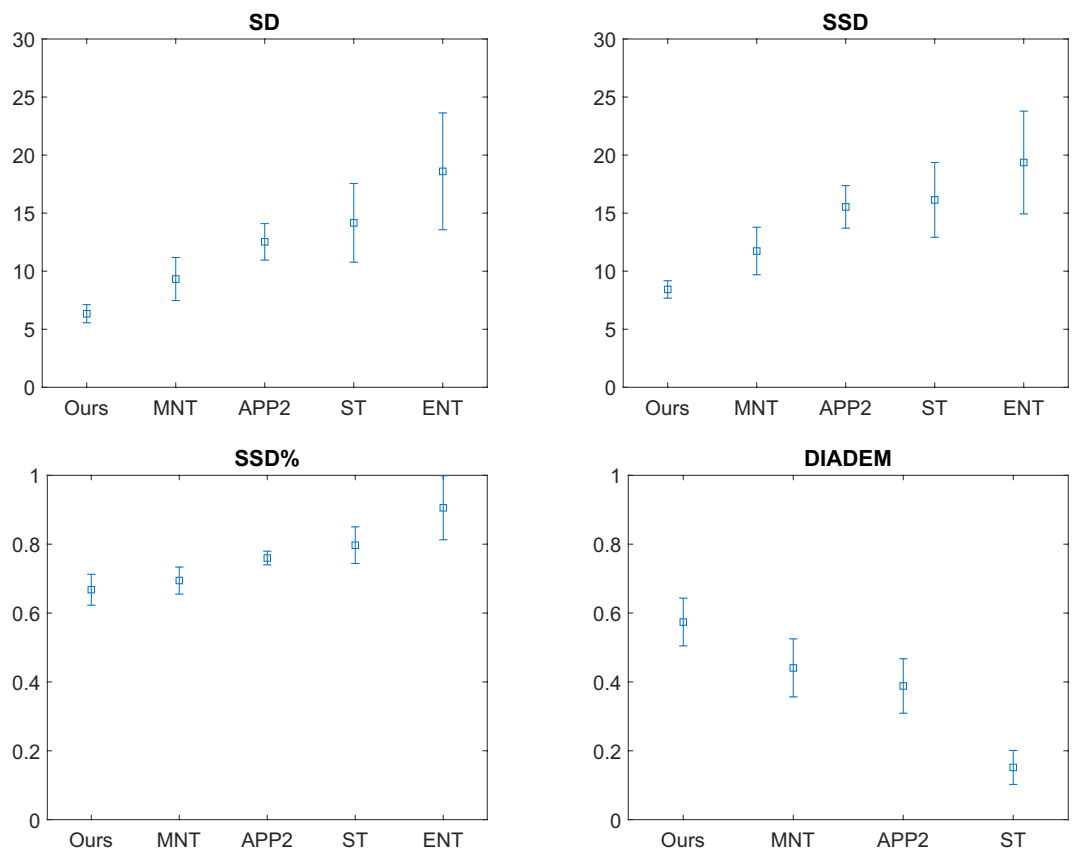


Figure 2. Mean and its 95% confidence interval plots of the evaluation results summarized in Table 1, achieved via our method, Multiple Neuron Tracer (MNT), App2, Simple Tracing (ST), and Ensemble Neuron Tracer (ENT).

for our method, the parameters are chosen based on the expected size of microglial cells as explained in the method section and would not normally need to be adjusted by the user. Note, the fluorescent images are RGB color images with the green channel containing the signal. Thus the green channel was specified prior to running the plugins in Vaa3D for each plugin that required a channel specification. Also note, soma centroids must be provided before running MNT which is a disadvantage of this method, and SWC files produced by ENT are slightly different from the standard definition and cannot be assessed by DIADDEM metric.

Table 1 shows the means and standard errors of the results. They are also plotted with 95% confidence interval of the mean in Fig. 2. The comparison shows our method achieves the lower spatial distances (SD, SSD, and SSD%) to the ground truths, and the higher DIADDEM scores. This means that the images reconstructed via our method are more similar to the ground truths than the results produced via the state-of-the-art, therefore our method outperforms them in terms of quality of reconstruction assessed via these metrics.

Four SWC files of the reconstructed cells are randomly selected to be visualized via SharkViewer³¹ (Fig. 3). It can be seen that our 3D reconstructions are the closest to the ground truths. Since soma volumes are automatically segmented in our method, they are also visualized in Fig. 3. Note, somas are not considered in the performance evaluation via the employed metrics.

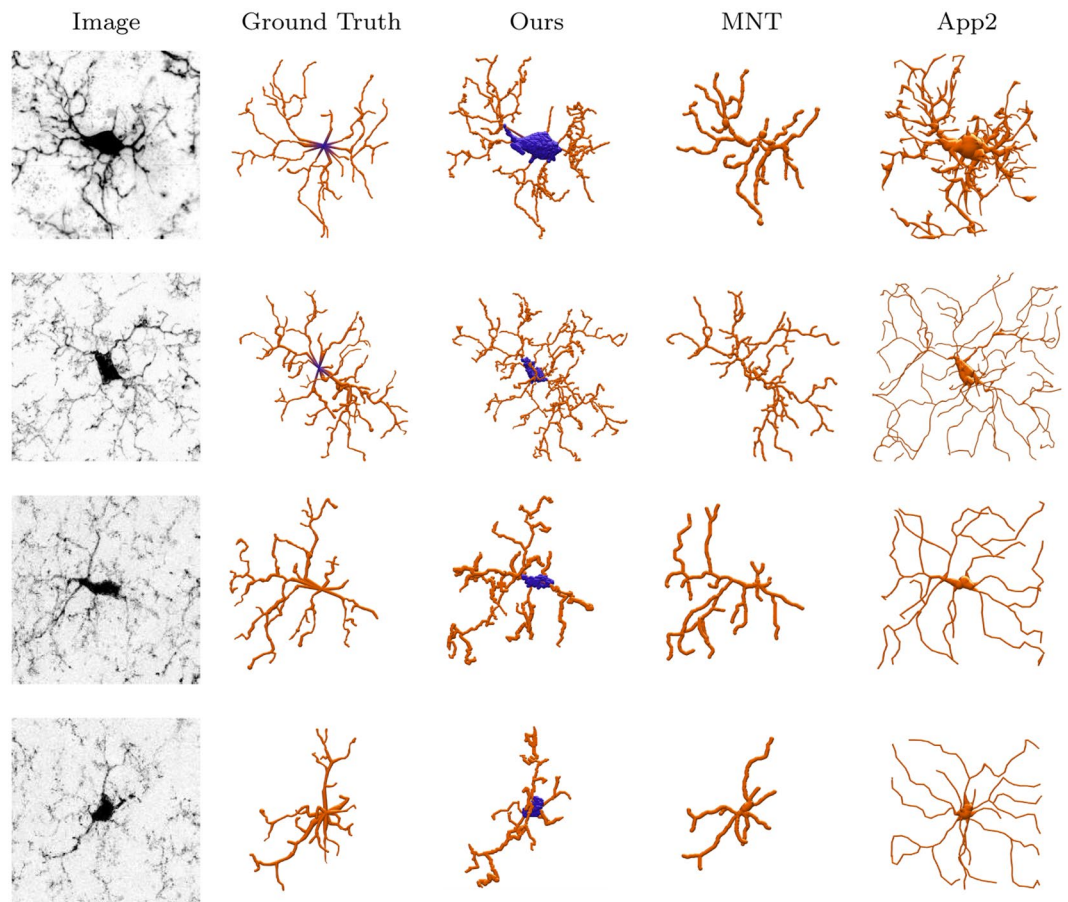


Figure 3. 3D visualization of 4 microglial samples reconstructed via our method, MNT¹⁷, and App2¹⁰.

	Ours	MNT	App2	ST	ENT
Time (seconds)	1.19	8.46	4.56	1.23	34.07

Table 2. Computational time of different methods. System specifications: Windows 10 running on Dell, Intel core i7 with 16 GB memory. Image size: $201 \times 201 \times 26$.

It is hard to compare the computational time of these algorithms, since they are coded in different programming languages (ours in MATLAB, the other methods in C++) and optimized for different platforms. However, Table 2 summarize the computational time of these methods to give us a clue about their computational order. These times have been recorded while these methods are applied to the same image with a size of $201 \times 201 \times 26$, running on a DELL core i7 laptop with 16 GB RAM. It shows that our method reconstructs the image faster than others.

Reconstruction and quantification of a large size image. We have applied our method to a large 3D fluorescent image stack (dimensions: $3070 \times 2047 \times 21$ voxel equals $1108 \mu\text{m} \times 738 \mu\text{m} \times 30 \mu\text{m}$ physical size) of microglia including 532 cells. The 3D reconstruction is visualized in Fig. 4 via SharkViewer³¹, and the average quantitative results of the cell features are presented in Table 3. The computational time required to automatically reconstruct and quantify the whole of the 3D stack was 415 seconds. These results show that our method is successful in reconstruction of large 3D microglial images, and quantification of their features with high speed and accuracy.

Discussion

We have presented a completely automated fast algorithm for reconstruction of microglial cells from 3D and 2D microscopy images which allows accurate quantification and visualization of the traced cells. Automatic soma segmentation and background removal are performed using multilevel thresholding as the first step of our algorithm. Different intensity levels output by the first step are employed to create the prioritized seed points. The tracing step is started with the centroid of the soma found in the segmentation step and ended with the background level found in the background removal step.

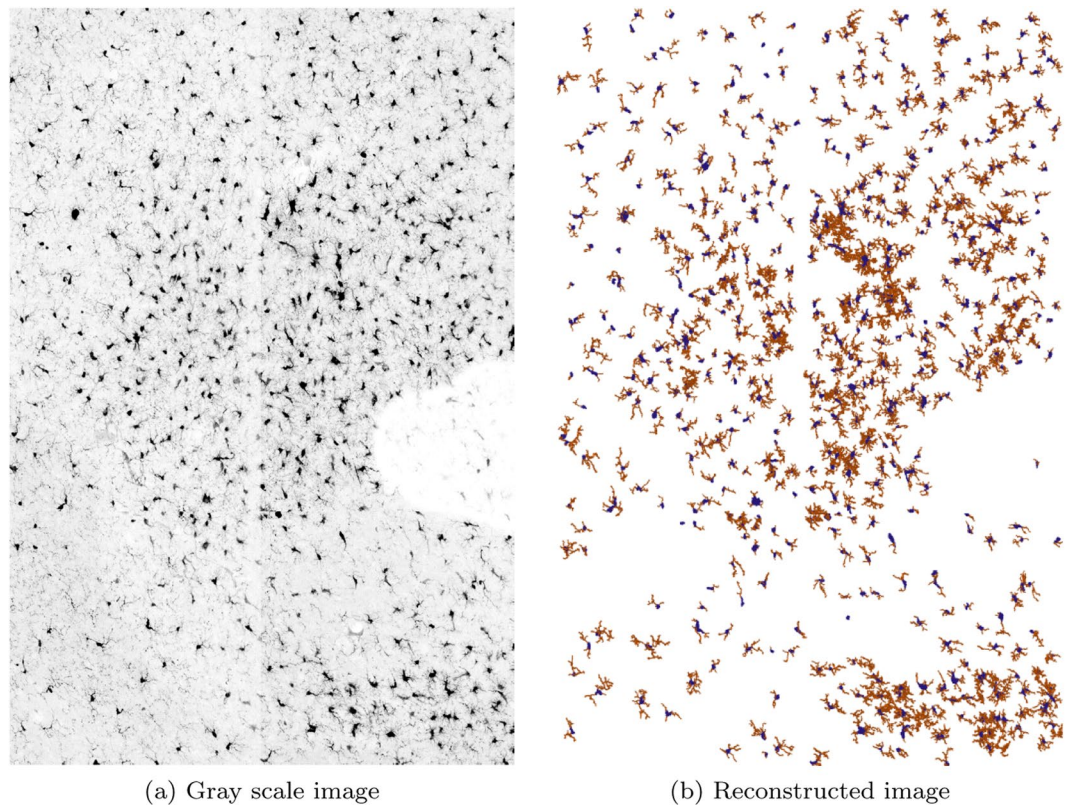


Figure 4. 3D reconstruction of microglial cells from a fluorescent image stack. Dimensions: $3070 \times 2047 \times 21$ voxel equals $1108 \mu\text{m} \times 738 \mu\text{m} \times 30 \mu\text{m}$ physical size. **(a)** Gray scale version of the maximum intensity projection of the original image, **(b)** 3D reconstruction.

	NPB	NBP	Branch length	Soma size	Cell size	Time (seconds)
Average per cell	5.12	11.31	230	124	520	0.78

Table 3. Quantification results. NPB: number of primary branches, NBP: number of branch points. The branch length is in μm , and the soma and cell sizes are in μm^3 .

Prioritizing the seed points based on their intensities helps to smoothly direct the paths through the centreline of the processes. The definition of the geodesic distance via the cell voxel array, B_{t_b} , helps to avoid incorrect connection between the seed points during the tracing process. This process is designed to connect one disconnected seed point to the path in each iteration. This prevents a loop in the path and ensures a tree-like structure output for each cell which fulfils the topology of the microglial cell. The thickness of the traced centerlines are estimated through the length of small branches attached to them and used to derive area/volume measures. The reconstructed cells are written in SWC format and can be quantified and visualized to provide the morphological features of the cells which are crucial to gain insight into the cell functionality.

We evaluated the performance of our method and compared it with four state-of-the-art methods by applying them to 3D image stacks of microglia. The results were benchmarked against the ground truth images using four different metrics. The comparison has shown our method achieves the lower spacial distances to the ground truths and the higher DIADEM scores which means our method outperforms the others in terms of the reconstruction accuracy. We have also demonstrated the satisfying performance of our method in reconstruction of large 3D microglial images, and quantification of their features.

So far, the branch growing algorithms use intensity and local proximity information to decide how to trace the cells. On the other hand in the manual reconstruction, an expert uses information about the angle of the branch at the connection and the slight thickening at the connection point to determine where the branches should be broken, or might decide to ignore a gap and connect a broken branch due to biological features. As a future work, we suggest to consider these biological features of the microglia to create a smart tracing algorithm which incorporates this expert knowledge directly.

References

- Kettenmann, H., Kirchhoff, F. & Verkhratsky, A. Microglia: New roles for the synaptic stripper. *Neuron* **77**, 10–18 (2013).
- Tremblay, M.-È., Lowery, R. L. & Majewska, A. K. Microglial interactions with synapses are modulated by visual experience. *PLoS Biol* **8**, 1–16 (2010).
- Wake, H., Moorhouse, A. J., Jinno, S., Kohsaka, S. & Nabekura, J. Resting microglia directly monitor the functional state of synapses *in vivo* and determine the fate of ischemic terminals. *The Journal of Neuroscience* **29**, 3974–3980 (2009).
- Walker, F. R. *et al.* Dynamic structural remodelling of microglia in health and disease: A review of the models, the signals and the mechanisms. *Brain, Behavior, and Immunity* **37**, 1–14 (2014).
- Lesage, D., Angelini, E. D., Bloch, I. & Funka-Lea, G. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis* **13**, 819–845 (2009).
- Acciai, L., Soda, P. & Iannello, G. Automated neuron tracing methods: An updated account. *Neuroinformatics* **14**, 353–367 (2016).
- Donohue, D. E. & Ascoli, G. A. Automated reconstruction of neuronal morphology: An overview. *Brain Research Reviews* **67**, 94–102 (2011).
- Peng, H., Meijering, E. & Ascoli, G. A. From diadem to bigneuron. *Neuroinformatics* **13**, 259–260 (2015).
- Wang, C.-W., Lee, Y.-C., Pradana, H., Zhou, Z. & Peng, H. Ensemble neuron tracer for 3D neuron reconstruction. *Neuroinformatics* **15**, 185–198 (2017).
- Xiao, H. & Peng, H. APP2: automatic tracing of 3D neuron morphology based on hierarchical pruning of a gray-weighted image distance-tree. *Bioinformatics* **29**, 1448–1454 (2013).
- Yang, J., Gonzalez-Bellido, P. T. & Peng, H. A distance-field based automatic neuron tracing method. *BMC Bioinformatics* **14**, 93 (2013).
- Otsu, N. A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on* **9**, 62–66 (1979).
- Kozłowski, C. & Weimer, R. M. An automated method to quantify microglia morphology and application to monitor activation state longitudinally *in vivo*. *PLoS ONE* **7**, e31814 (2012).
- Ding, Y. *et al.* Novel methods for microglia segmentation, feature extraction and classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **PP**, 1–1 (2016).
- Rubinstein, R., Bruckstein, A. M. & Elad, M. Dictionaries for sparse representation modeling. *Proceedings of the IEEE* **98**, 1045–1057 (2010).
- Jiang, Z., Lin, Z. & Davis, L. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In *CVPR*, 1697–1704 (2011).
- Megjhani, M. *et al.* Population-scale three-dimensional reconstruction and quantitative profiling of microglia arbors. *Bioinformatics* **31**, 2190–2198 (2015).
- Prim, R. C. Shortest connection networks and some generalizations. *Bell System Technical Journal* **36**, 1389–1401 (1957).
- Wang, Y., Narayanaswamy, A. & Roysam, B. Novel 4-D open-curve active contour and curve completion approach for automated tree structure extraction. In *CVPR*, 1105–1112 (2011).
- Xu, C. & Prince, J. L. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* **7**, 359–369 (1998).
- Cohen, L. D. & Deschamps, T. Segmentation of 3D tubular objects with adaptive front propagation and minimal tree extraction for 3D medical imaging. *Computer Methods in Biomechanics and Biomedical Engineering* **10**, 289–305 (2007).
- Rouchdy, Y. & Cohen, L. D. Geodesic voting for the automatic extraction of tree structures. Methods and applications. *Computer Vision and Image Understanding* **117**, 1453–1467 (2013).
- Rouchdy, Y., Cohen, L. D., Pascual, O. & Bessis, A. Minimal path techniques for automatic extraction of microglia extensions. *Int. J. Comput. Vis. Biomech.* **4**, 35–42 (2011).
- Harris, C. & Stephens, M. A combined corner and edge detector. In *Proc. of Fourth Alvey Vision Conference*, 147–151 (1988).
- Sethian, J. *Level Set Methods and Fast Marching Methods. Cambridge Monographs on Applied and Computational Mathematics.* (Cambridge University Press, 1999).
- Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 583–598 (1991).
- Vincent, L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions on Image Processing* **2**, 176–201 (1993).
- Meyer, F. Topographic distance and watershed lines. *Signal Processing* **38**, 113–125 (1994).
- Cannon, R., Turner, D., Pyapali, G. & Wheal, H. An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods* **84**, 49–54 (1998).
- Myatt, D., Hadlington, T., Ascoli, G. & Nasuto, S. Neuromantic—from semi-manual to semi-automatic reconstruction of neuron morphology. *Frontiers in Neuroinformatics* **6**, 4 (2012).
- Weaver, C. & Bruns, C. Sharkviewer, <http://www.janelia.org/sharkviewer> (2014).
- Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nature biotechnology* **28**, 348–53 (2010).
- Scorcioni, R., Polavaram, S. & Ascoli, G. A. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat. Protocols* **3**, 866–876 (2008).
- Gillette, T. A., Brown, K. M. & Ascoli, G. A. The DIADEM metric: Comparing multiple reconstructions of the same neuron. *Neuroinformatics* **9**, 233 (2011).
- Radojević, M. NDIS, <http://www.braincadet.com/compare-neuron-trees/> (2017).

Author Contributions

M.A. developed the algorithms and methods, wrote the programming code, conducted the experiments, and wrote the main manuscript text, M.G.K. took the microscopy images and provided the ground truth data, F.R.W. supervised the work, S.J.J. conceived the research idea, edited and supervised the work. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019