WAVELET-BASED SEGMENTATION AND CONVEX HULL APPROACHES FOR QUANTITATIVE ANALYSIS OF BIOLOGICAL IMAGING DATA

by

Tzu-Ching Wu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agricultural and Biological Engineering

West Lafayette, Indiana

December 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. David M. Umulis, Chair

Department of Agricultural and Biological Engineering

Dr. Jenna Rickus

Department of Agricultural and Biological Engineering

Dr. Mamara Kinzer-Ursem

Department of Biomedical Engineering

Dr. Meng Deng

Department of Agricultural and Biological Engineering

Dr. Sherry Harbin

Department of Biomedical Engineering

Approved by:

Dr. Nathan Mosier

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CWT	Continuous Wavelet Transform
DS	Derivatives Sum
GUI	Graphical User Interface
IO	Input and Output
CT	Computed tomography
TP	True Positive
FP	False Positive
FN	False Negative
SNR	Signal-to-Noise Ratio

ABSTRACT

Imaging-based analysis of developmental processes are crucial to understand the mechanisms controlling plant and animal development. In vertebrate embryos such as the zebrafish embryo, nuclei segmentation plays an important role to detect and quantify nuclei over space and time. However, limitations of the image quality and segmentation methods may affect the segmentation performance. In plant including studies on Arabidopsis epidermis growth, cellular shape change dictates organ size control and growth behavior, and quantitative image analysis of dynamics cell patterning is needed to link the cause and effect between cells and organs. Here we provide a series of new quantitative biological imaging methods a series of new quantitative biological imaging methods and tools including wavelet-based segmentation method in zebrafish embryo development studies and convex hull approach for quantitative shape analyses of lobed plant cells.

Identification of individual cells in tissues, organs, and in various developing systems is a well-studied problem because it is an essential part of objectively analyzing quantitative images in numerous biological contexts. In this paper we present a size dependent wavelet-based segmentation method that provides robust segmentation without any preprocessing, filtering or fine-tuning steps, and is robust to the signal-to-noise ratio (SNR). The program separates overlapping nuclei, identifies cell cycle states and minimizes intensity attenuation in object identification. The wavelet-based methods presented herein achieves robust segmentation results with respect to True Positive rate, Precision, and segmentation accuracy compared with other commonly used methods. We applied the segmentation program to Zebrafish embryonic development IN TOTO quantification and developed an automatic interactive imaging analysis platform named WaveletSEG, that integrates nuclei segmentation, image registration, and nuclei shape analysis. A set of additional functions we developed include a 3D ground truth annotation tool, a synthetic image generator, a segmented training datasets export tool, and data visualization interfaces are also incorporated in WaveletSEG for additional data analysis and data validation.

In addition to our work in Zebrafish, we developed image analysis tools for quantitative studies of cell-to-organ in plants. Given the importance of the epidermis and this particular cell type for leaf expansion, there is a strong need to understand how pavement cells morph from a simple

polyhedral shape into highly lobed and interdigitated cells. Currently, it is still unclear how and when patterns of lobing are initiated in pavement cells, and one major technological bottleneck to address the problem is the lack of a robust and objective methodology to identify and track lobing events during the transition from simple cell geometry to lobed cells. We develop a convex-hull-based algorithm termed LobeFinder to identify lobes, quantify geometric properties, and create a useful graphical output for further analysis. The algorithm is validated against manually curated cell images of pavement cells of widely varying sizes and shapes. The ability to objectively count and detect new lobe initiation events provides an improved quantitative framework to analyze mutant phenotypes, detect symmetry-breaking events in time-lapse image data, and quantify the time-dependent correlation between cell shape change and intracellular factors that may play a role in the morphogenesis process.

CHAPTER 1. WAVELET-BASED SEGMENTATION METHOD

1.1 Introduction

In developmental biology, quantification of cell morphology, tissue patterning, and gene expression within a developing embryo at different stages provides detailed information that drives our understanding of differentiation and the function of signaling networks driving cell fate specification. Several nuclei segmentation methods have been developed to track nuclei position in fluorescent images and extract the spatio-temporal expression of each nuclei [1-4]. Most of the nuclei segmentation algorithms need image pre-processing steps before segmentation, and the settings in pre-processing are highly dependent on the microscopy imaging environment [5]. Global and adaptive thresholding segmentation methods are the simplest and the most common nuclei segmentation methods that use single or multiple threshold values on image histogram to distinguish between an object and background [6-7].

However, most real images lack significant thresholding points or low contrast foreground and background8. The Watershed method is another common segmentation method also capable of dividing overlapping nuclei [9-10]. Nearby pixels with similar features are classified topologically and defined as a catchment basin that are separated by a watershed ridge line from the adjacent catchment basin. Using the watershed method can divide overlapping objects, but may cause an over-cutting problem. Other methods such as active contour models (ACMs) [11-12], level-set [13] and graph cuts [14] are also applied in the field of nuclei segmentation.

Various level-set and model-based methods have been developed for accurate embryo nuclei segmentation [15-17], although they are computationally expensive or require pre-knowledge [18]. To segment and track chromosome dynamics in the early Drosophila embryo, a fast level-set method was proposed [19]. A 3D level-set method was also applied to detect nuclei and mitotic chromosomes in Drosophila embryos [20]. A region-based active contour method was developed in 3D cell counting and segmentation of vertebrae in early embryogenesis [21]. Nuclei segmentation in large scale imaging such as whole mount embryo imaging, encounters a variety of challenges. Intensity attenuation effects due to absorption and scattering of fluorescent light in

deep layer specimens can lead to intensity decay and underestimates in the segmentation results [22]. Moreover, densely packed nuclei are often present in an embryo which requires overlapping nuclei detection strategies [23]. Wavelet transforms are common in signal processing and are applied in many fields where the method is used to isolate the object from a noisy signal. The 2D WTMM (2D wavelet transform modulus maxima) method was first introduced for multiscale edge detection and segmentation of regions of interest in cell nuclei [24-25] and CT detection [26].

Several microscopy image analysis software packages have been established, such as ImageJ [27], Fiji [28], and instrument-bundled packages: ZEN and cellSens. However, commercial bundled packages lack flexibility and may lack affordable options for on-the-fly segmentation, and open source tools are either limited in specific applications, not robust in segmentation results, or not convenient to use. For nuclei segmentation tools, widely used software CellProfiler [29-30], provides pre-processing, object identification, and cell counting modules, but it is limited to only 2D images. Additional related software includes Matlab toolbox OML that is capable of automatic cell segmentation. Spatzcells [31], FISH-quant [32] and MINS [33] are Matlab executable codes that can generate segmentation cell masks and spot recognition for fluorescent 2D or 3D images. LSDCAS (The Large Scale Digital Cell Analysis System) [34-35] provides the ability for live cell image segmentation and tracking of cell trajectories. RACE [36] was designed for automated 3D cell segmentation and cell reconstruction for embryos, based on a watershed segmentation approach that, in our hands is more sensitive to non-uniform of variable quality images (Table. 1.1).

In this chapter we present a wavelet-based automatic 3D nuclei segmentation method that is object size-dependent, robust to noise and intensity attenuation, and can divide overlapping nuclei. It contains six steps, including a 2D continuous wavelet transform, multi-scale object identification, 3D alignment, first and second division steps and outlier removal. We tested the model against manually curated 3D embryonic data and found that it has robust and stable segmentation performance with respect to True Positive Rate (TP rate), Precision and segmentation accuracy compared to other methods. We integrate wavelet-based nuclei segmentation, image registration, topology and geometry for structural analysis of the embryonic development on an automatic interactive analysis platform named WaveletSEG for embryological developmental studies in

Zebrafish. Additionally, we developed a 3D ground truth annotation tool, synthetic image generator, and a segmented training dataset export tool and data visualization interface for additional data analysis and data validation in WaveletSEG. We quantified 3D nuclei and intensity spatial distributions, topology features and shape classification results at 4.7, 5.7, and 6.5 hours post fertilization (hpf) Zebrafish using WaveletSEG.

Table 1.1. Lists of nuclei segmentation embryonic imaging analysis tools. In this table, we compare the embryonic-analysis-related functionality of common nuclei segmentation embryonic imaging software.

	WaveletSEG	DS	Point-wise	RACE	generic	MINS	ilastik	CellSegmentation3D	BioEmergences	CellSegm	MLS
Reference		38	41	36	51	33	62	63	64	65	66
Nuclei segmentation	●	●	●	●	●	●	●	●	●	●	●
Nuclei feature analysis	●	●		●	●						●
Time lapse/Tracking		●		●		●	●		●		
Graphical user interface	●			●		●	●				
Membrane SEG				●					●		
Synthetic dataset	●	●									
Quantification	●		●								
Nuclei shape analysis	●				●						
Cell fate/Cell lineage	●			●	●	●			●		●
Nuclei separation	●									●	●
Aberration calibration	●										
Data visualization	●					●		●	●		●
Count number	●		●			●					
Ground truth labeling	●										
Intensity calibration	●										
Cell registration	●										

1.2 Results

1.2.1 Overview of wavelet-based nuclei segmentation methods

Most existing nuclei segmentation methods are based on pixel intensity thresholding or region growing methods after multiple imaging preprocessing steps such as smoothing, denoising or morphology filtering (Figure 1.1a). Preprocessing steps are crucial for most segmentation methods; hence, preprocessing types, settings and parameters vary depending on different microscopy types and operation, and professional knowledge is needed to set the parameters to obtain good enough image quality for segmentation [37]. Unlike many segmentation methods, the wavelet-based method has a strong ability to segment objects in noisy images without any preprocessing steps.

The method herein combines a 2D continuous wavelet transform (Figure 1.1b), and the only parameter needed for object determination is the wavelet scale factor determined by target object size, and no other preprocessing steps are needed.

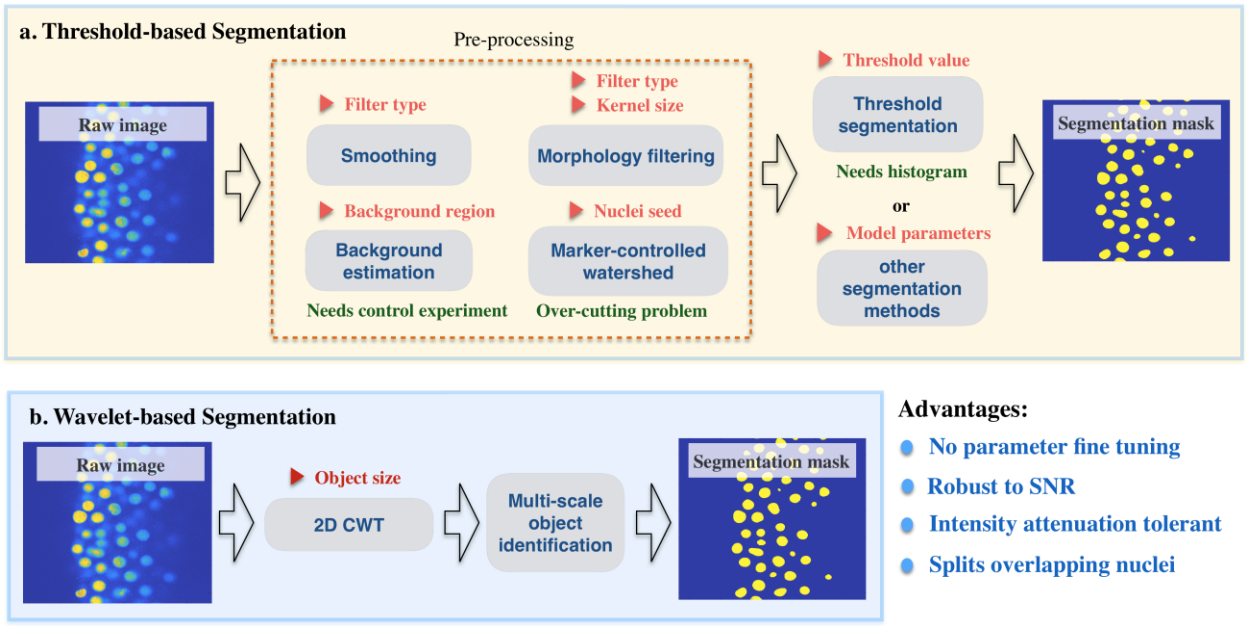


Figure 1.1. Comparison between threshold-based segmentation and wavelet-based segmentation methods. Preprocessing steps are crucial for most threshold-based segmentation methods and determine the quality of segmentation performance. Increased parameter number in preprocessing steps and segmentation method means the increased result complexity and decreased robustness. Wavelet-based segmentation we developed need no preprocessing step and need only one intrinsic parameter – object size such that we need no parameter fine tuning. The wavelet-based segmentation started from applying 2D continuous wavelet transform with appropriate scale factors, and identify object by comparing multi-scale identification. Unlike threshold-based segmentation which use threshold value to distinguish between fore and background, wavelet-based segmentation extract typical size peak from image which can overcome the intensity attenuation and robust to SNR. Red triangle symbol indicate parameters required for each step and may need parameter fine tuning.

The idea of the wavelet-based nuclei segmentation approach for embryo quantification arose from the comparison of the elevation-like plots that form as level-sets from different scale factor wavelet coefficient matrices (Figure 1.2). Figure 1.2b shows the wavelet coefficient maps with scale factors matching the object size we want to identify. By comparing these maps, we find a similar contour pattern between wavelet coefficient scales. Other regions that are not object containing show discordance and randomness between these wavelet coefficient maps. Through comparison

between several wavelet coefficient maps, wavelet-based segmentation can easily distinguish the object region from the background in images and it can do this without any preprocessing or thresholding steps. The coincidence of objects between multiple wave numbers reduces the impact of intensity attenuation (Figure 1.3), and by selecting different wave number ranges, it is capable of both nuclei and RNA spot segmentation (Figure 1.4).

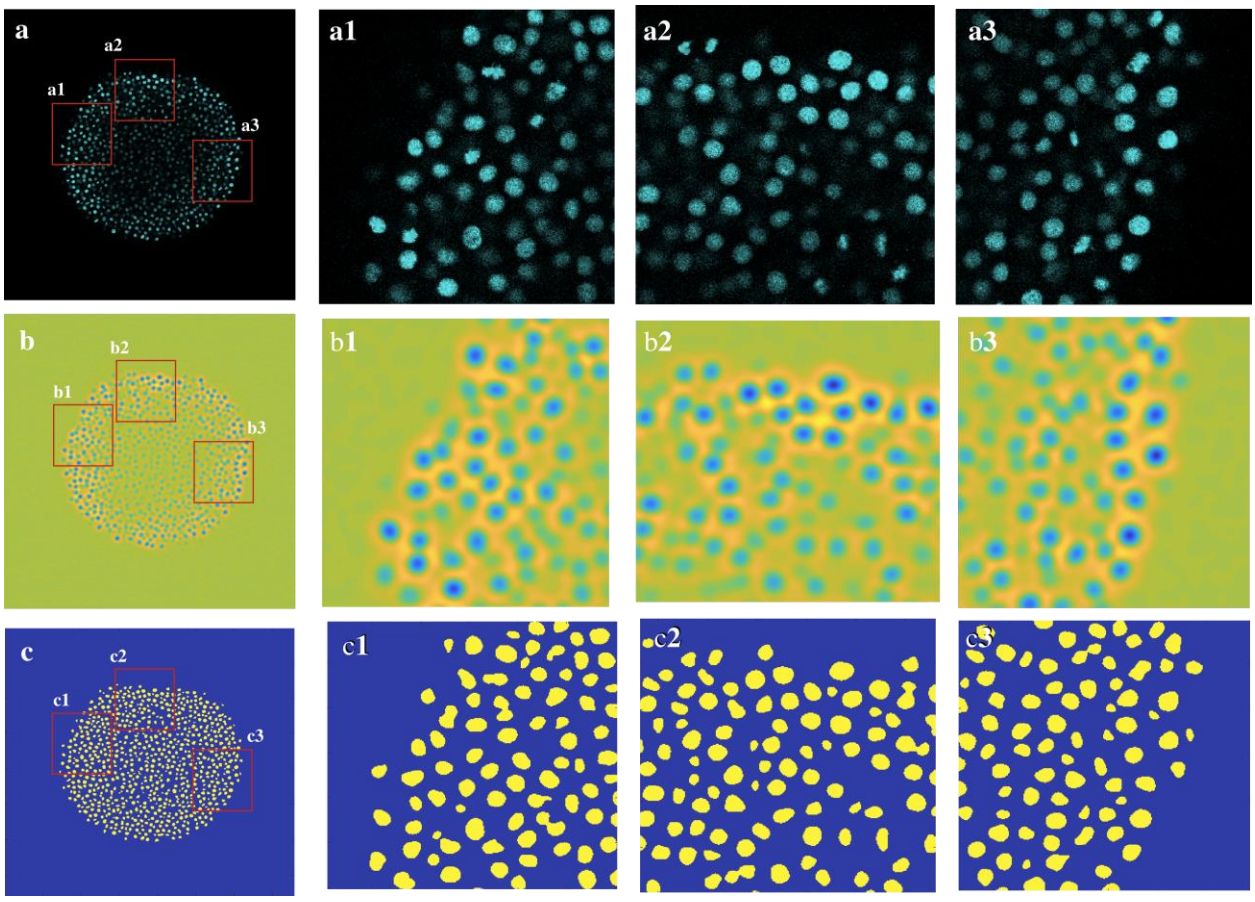


Figure 1.2. Illustration of wavelet-based segmentation method in an optical section of a whole-mount zebrafish embryo. (a, a1-3) Whole-mount embryo z-slice nuclear DAPI-staining raw images of germ ring-stage (5.7 hpf) zebrafish and three red-boxed regions. Image data were acquired with a Zeiss LSM 800 confocal microscope with 20X water lens. **(b, b1-3)** Wavelet coefficient matrixes after applying 2D CWT on raw images in **(a)**. The blue color corresponds to the negative coefficient value region, and the yellow color corresponds to the positive coefficient value region. 2D Mexican hat function was selected as the mother wavelet function. **(c, c1-3)** 2D segmentation masks (yellow regions) were obtained after applying multi-scale object identification on wavelet coefficient matrixes in **(b)**.

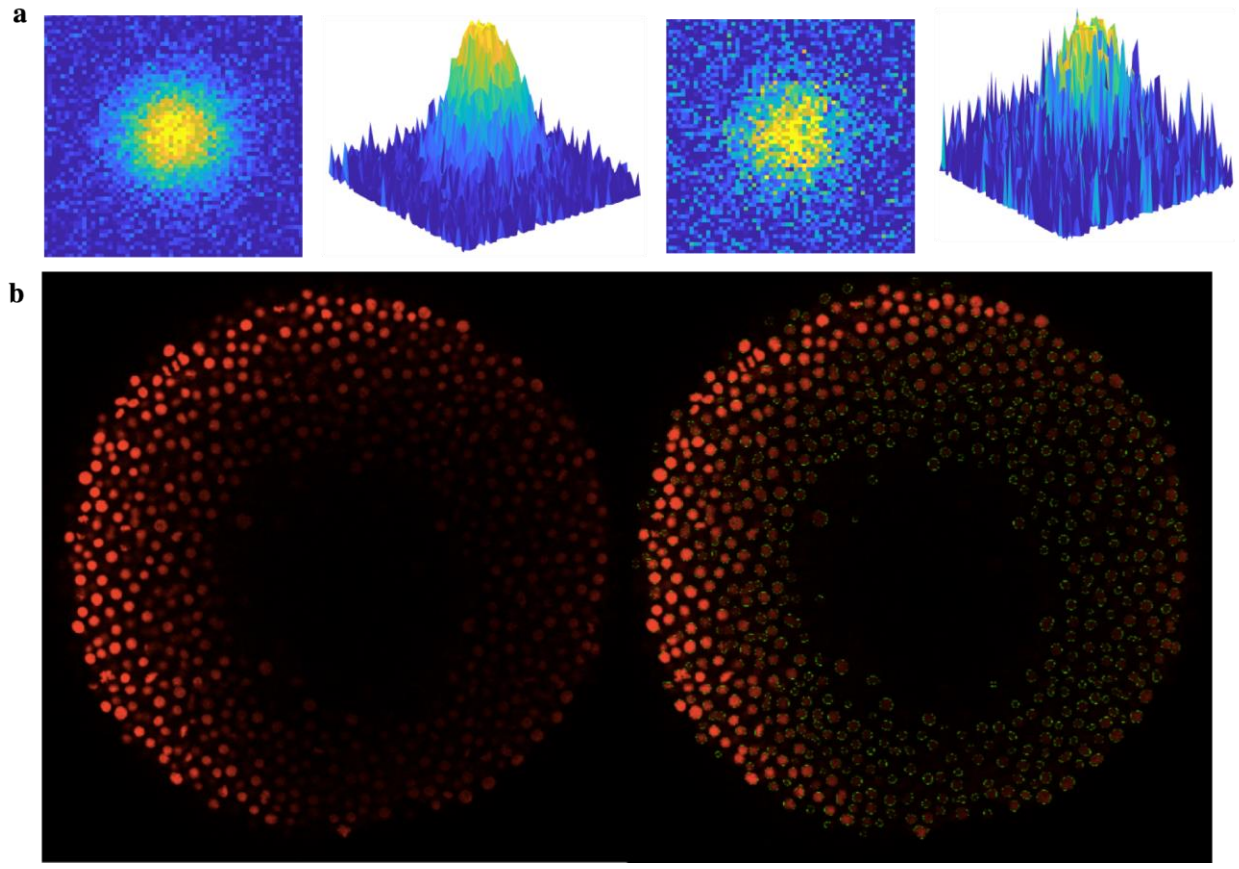


Figure 1.3. Limitations for nuclei segmentation on whole embryo confocal imaging. (a) Nucleus with clear boundaries and high contrast (left) and nucleus with inhomogeneous background intensity and low contrast (right). **(b)** Whole zebrafish embryo z-slice nuclear DAPI staining with intensity attenuation effect and nuclei segmentation results using wavelet-based segmentation method (right). Green boundaries show nuclear segmentation masks, and nuclei with inhomogeneous intensity can be detected.

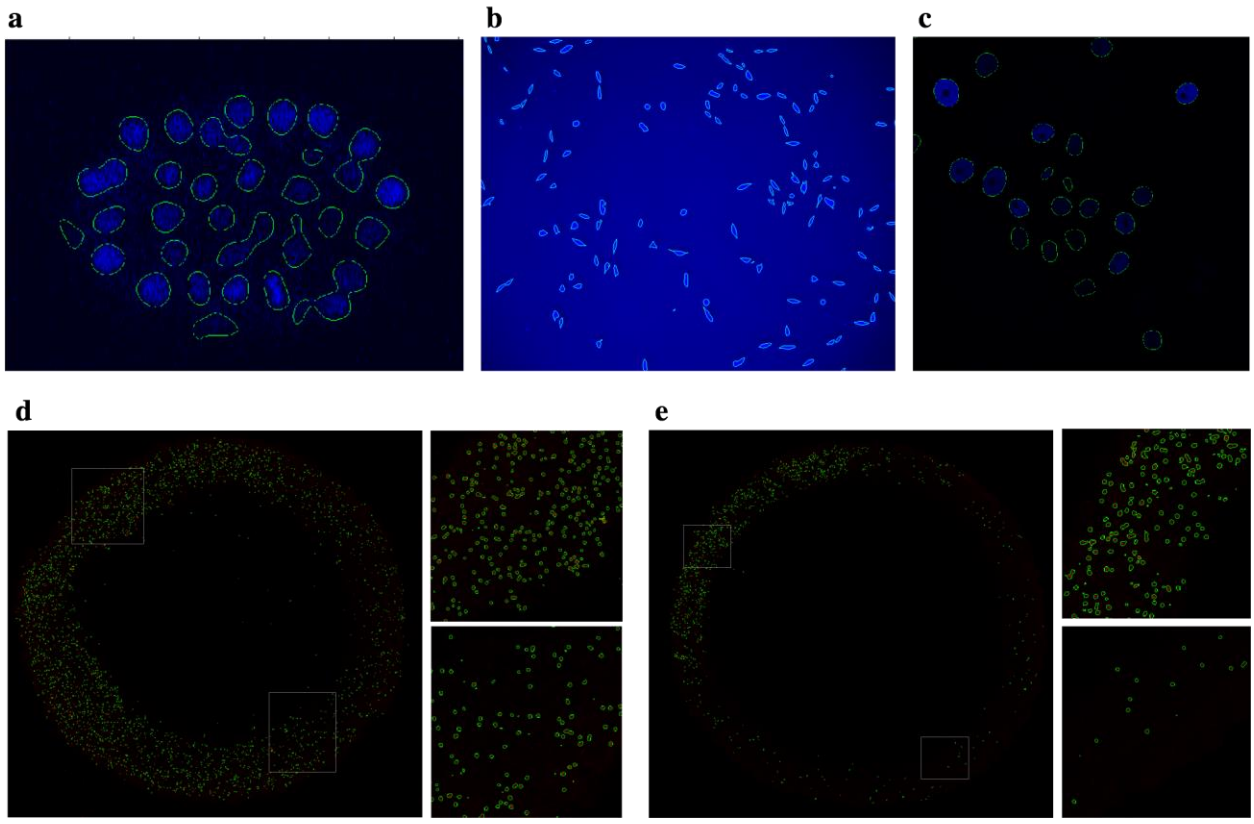


Figure 1.4. Nuclei segmentation and RNA segmentation results using wavelet-based segmentation method. **(a)** C.elegans developing embryo nuclear DAPI staining by light-sheet microscope with segmentation masks marked by green boundaries **(b)** Pancreatic Stem Cells on a Polystyrene substrate, **(c)** GFP-GOWT1 mouse stem cells, and **(d)(e)** Whole mount zebrafish embryo z-slice bmp2b mRNA expression at 5.7hpf, and bmp RNA spot segmentation results using wavelet-based segmentation method in different z planes(left and right). Insets: enlarged views of two while box regions.

1.2.2 Overview of wavelet-based nuclei segmentation methods

The six main steps of the wavelet-based nuclei segmentation method (Figure 1.5) are: (1) application of a 2D continuous wavelet transform (2D CWT), (2) multi-scale object identification, (3) 3D object alignment, (4) first round of nuclei division based on xy plane wavelet coefficients, (5) second division of nuclei based on z direction, and (6) deletion of outliers and small objects. In step one, a two-dimensional continuous wavelet transform is applied to each Z-stack slice of 3D sections common as output from confocal imaging (Figure 1.5b). While a number of wavelet functions were tested, we find that the 2D Mexican hat function as the mother wavelet function provides the ideal function to resolve point-wise or nuclear-sized structures. The wavelet scale factor range applied after the transform is then based on the object size we wanted to identify. In

line with this, we designed an efficient multi-scale search algorithm that compares different scales of the CWT coefficients and groups shape-similar regions such as the nuclei masks based on contour line similarities in the CWT coefficients diagram (Figure 1.5c, and Figure 1.6). Through the comparison between wavelet coefficient matrices over different scales we determined that a similarity index of 0.9 for between contour line similarity routinely identified nuclei without additional prepossessing or thresholding steps. Regions that meet the criteria are masked and labeled as nuclei in each stack for further processing to identify nuclei properties including progression through the cell cycle.

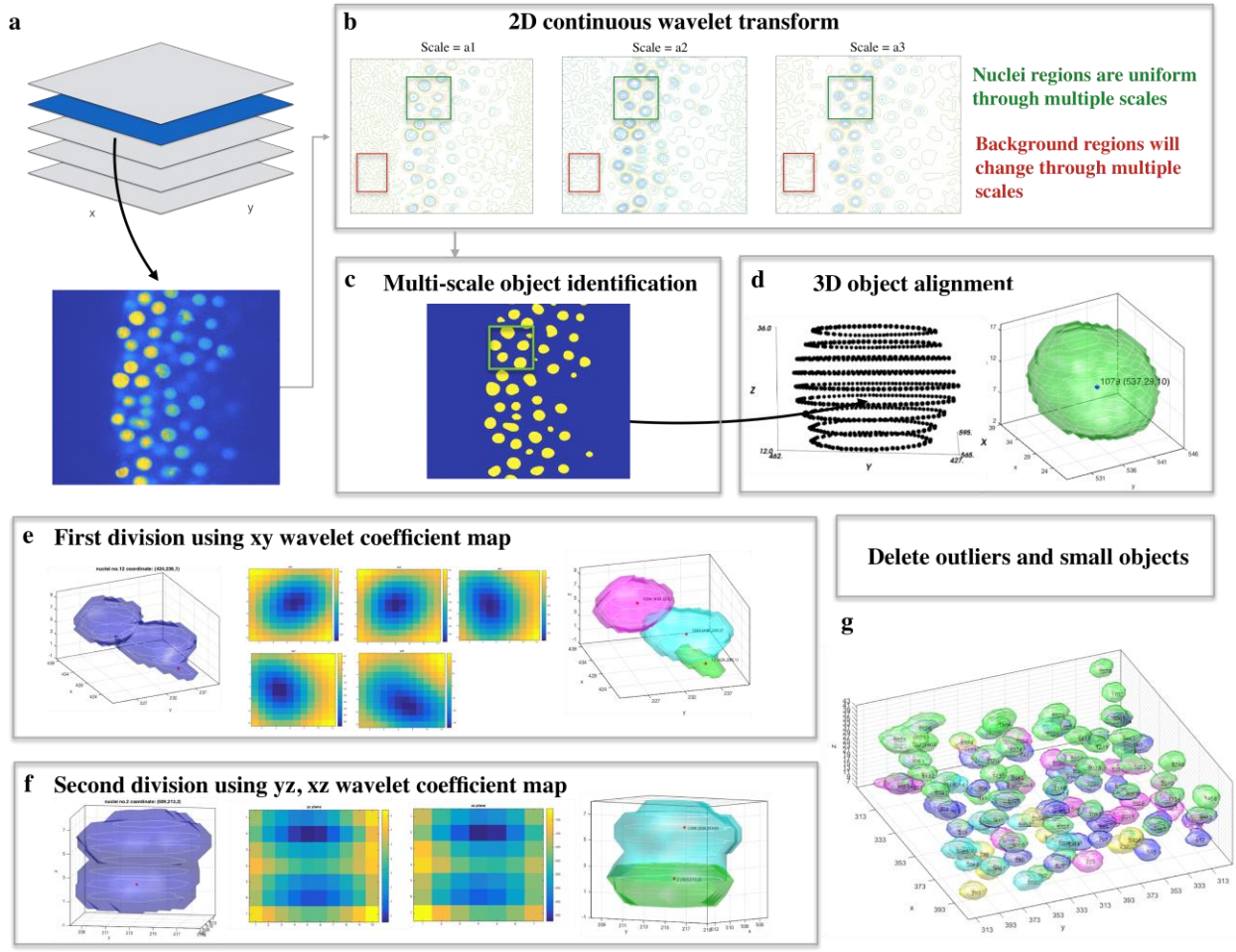


Figure 1.5. Overview of wavelet-based nuclei segmentation method. (a) Import 3D raw image files. **(b)** After importing 3D images, 2D continuous wavelet transform with appropriate wavelet scale factors are applying to each 2D z-stack. **(c)** 2D segmentation mask are obtained by applying multi-scale object identification to wavelet coefficient maps from previous step. **(d)** In 3D object alignment step, connected identified 2D masks from neighboring slices formed 3D foreground object. **(e)** First division step if more than one center or center shift on neighboring z slice wavelet coefficient maps. **(f)** Second division step in z direction if multiple centers found on yz or xz wavelet coefficient maps. **(g)** Delete 3D nuclei voxels smaller then nuclei size × 0.2.

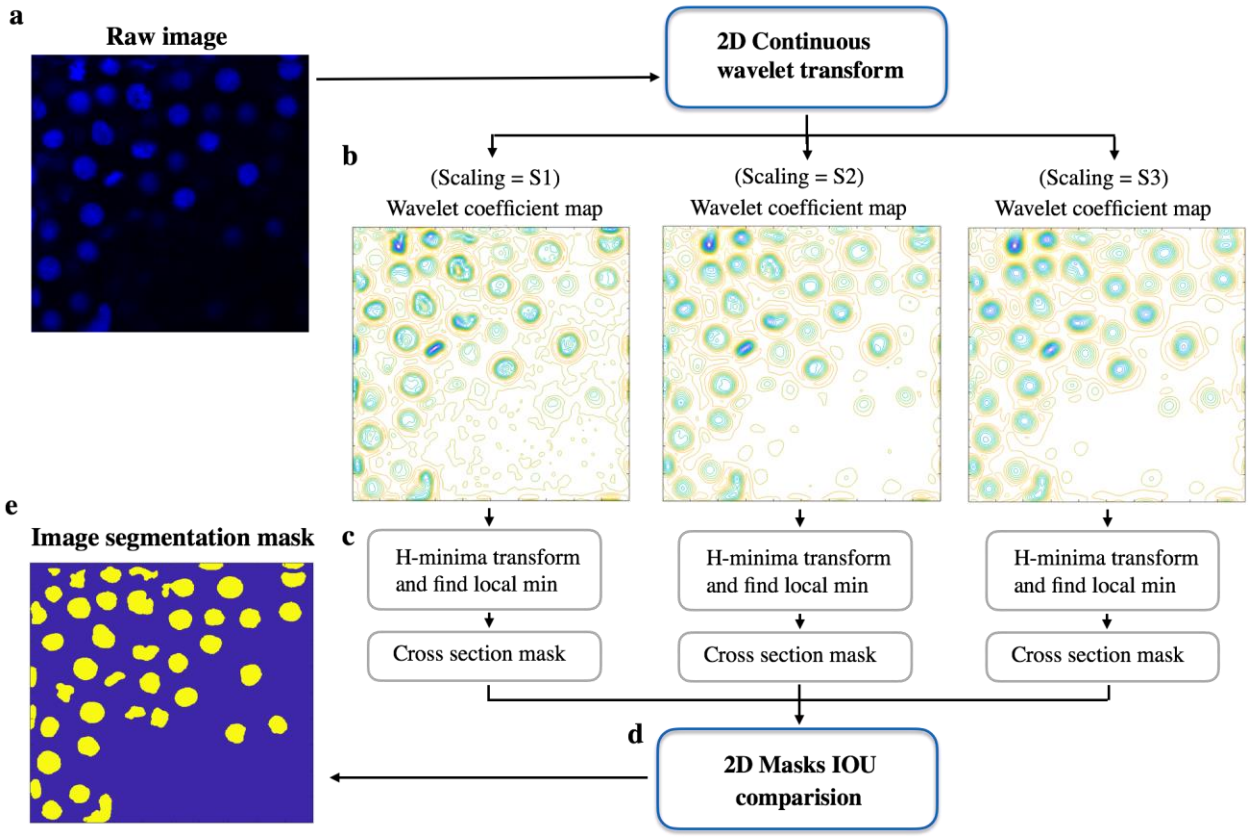


Figure 1.6. Workflow of multi-scale object identification step in wavelet-based segmentation method. (a) Zebrafish embryo nuclear DAPI staining raw images (5.7hpf). We apply 2D continuous wavelet transform (2D CWT) on this raw image with three wavelet scale factors s1, s2, s3. **(b)** Three wavelet coefficient maps with wavelet scale factors s1, s2, s3. Contour color represents intensity of the corresponding wavelet coefficient. The green intensive contoured ridges region in three wavelet coefficient map are almost identify and can be correlated with nuclei region in raw image. Besides that, the spatial pattern of lower right part of three wavelet coefficient maps have significant differences and can be correlated with background region. **(c)** Local minimums were determined on three wavelet coefficients map after the H-minima transform. If local minimums appeared in the same positions on all wavelet coefficient maps, they were identified as potential nuclei peak locations and the corresponding zero-value cross sections are calculated. **(d)** In this step we calculated the 2D masks intersection over Union (IOU) of all zero-value cross sections from all wavelet coefficient maps. If the difference in cross section areas is less than 10%, they averaged and were identified as nuclei 2D masks. **(e)** The final segmentation mask after multi-scale object identification step.

Following 2D identification, nuclei are reconstructed by 3D object alignment that aligns 2D nuclei cross-sections to recreate 3D nuclei shapes by comparing consecutive nuclei segmentation results from the preceding step (Figure 1.5d). The method computes the intersection of the mask coverage area of the neighboring z-plane. If the intersection area among adjacent z-planes is greater than

80% and if the max intensity is not lower than 20% nuclei center intensity, it is inferred that the adjacent mask coverage area should belong to the same 3D nuclei object.

In threshold and other segmentation approaches, a major challenge is errors in nuclei identification due to imaging conditions that lead to irregular or poor contrast boundaries or intensity attenuation that systematically distort measured fluorescence intensity. These factors compound in 3D and additional steps are needed for identification. First, the wavelet coefficient local minimum on each z-plane is determined because the spatial center of the object will produce a local minimum on the wavelet coefficient z-plane. If there is a shift in the wavelet coefficient local minimum on two consecutive z-planes and the shift distance is larger than 1/3 nuclei radius, this is indicative of two different nuclei objects, and it will be marked as likely belonging to different nuclei (Figure 1.5e, and Figure 1.7). For identification of potential separate, yet overlapping nuclei in the YZ-plane and XZ-plane a similar center point detection is used (Figure 1.5f and Figure 1.8). If multiple wavelet coefficient local minimum points are found on both the YZ-plane and the XZ-plane it is likely they come from two nearby nuclei. Lastly, noise or false positive nuclei after the additional segmentation are removed based on object size filtering (Figure 1.5g).

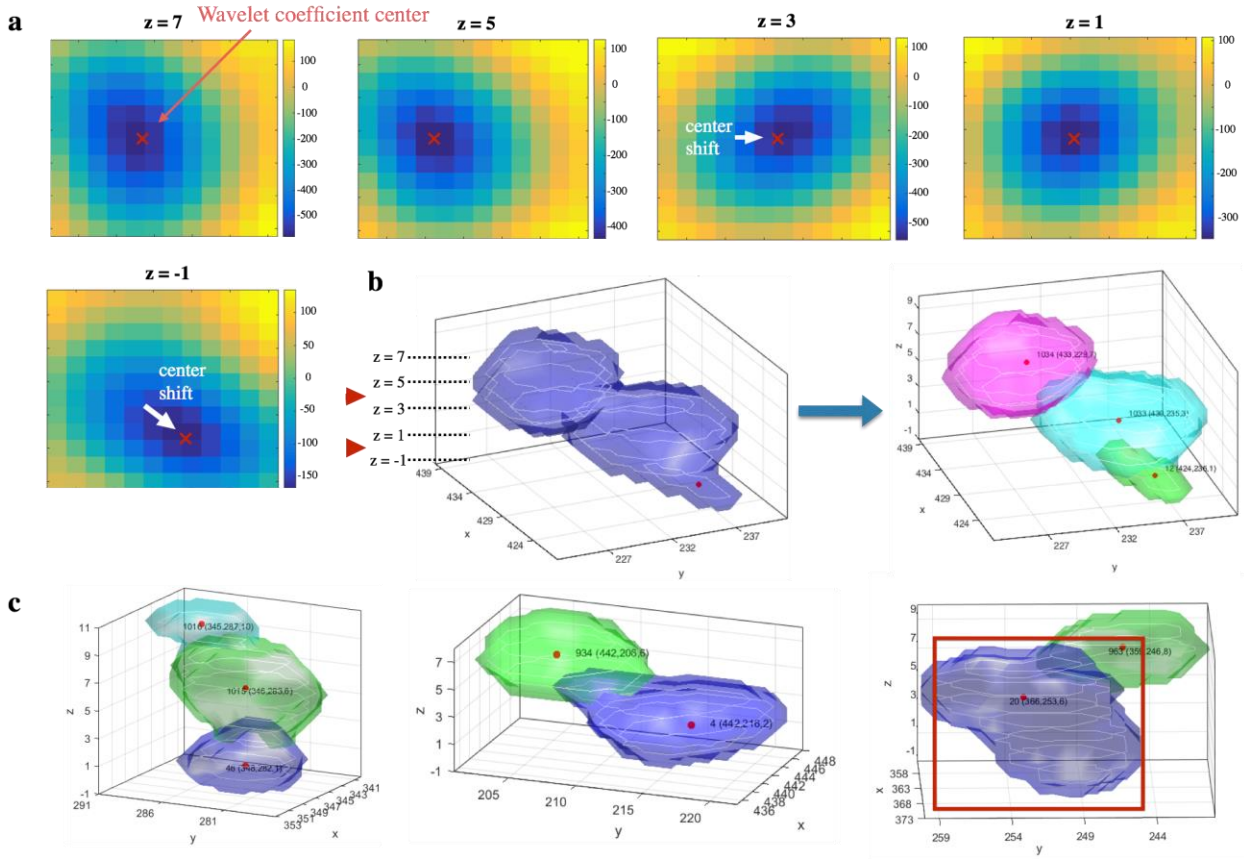


Figure 1.7. First division step based on wavelet coefficient center positions on z-slice of nuclei 3D blob. (a) Five z-slice wavelet coefficient maps of segmented nuclei 3D blob from top (z = 7) to button (z = -1). Color represents intensity of the corresponding wavelet coefficient from positive value (yellow) to negative value (deep blue). Red cross marks the center position (local minimum position) on each wavelet coefficient map z-slice in 3D blob. Here the z value came from z axis of whole embryo before 3D rotation. (b) If there are more than one center positions found or the shift of the center positions between neighboring Z slices is bigger than nuclei radius, we assigned a new nuclei object by cutting the median line between two centers, or assigned new nuclei in the case of center displacement. In (a) we found two center position shift between z=5 to z=3 and z=1 to z=-1, and marked using red arrow. So we divided the original 3D nuclei blob (left) into three nuclei 3D blobs (right) with three new nuclei centers (red dot) and displayed using different color. (c) Three examples show the 3D nuclei blobs after the first division step. Red box on the third 3D nuclei blob (right) shows the nuclei overlapping in z direction cannot be divided in the first division step.

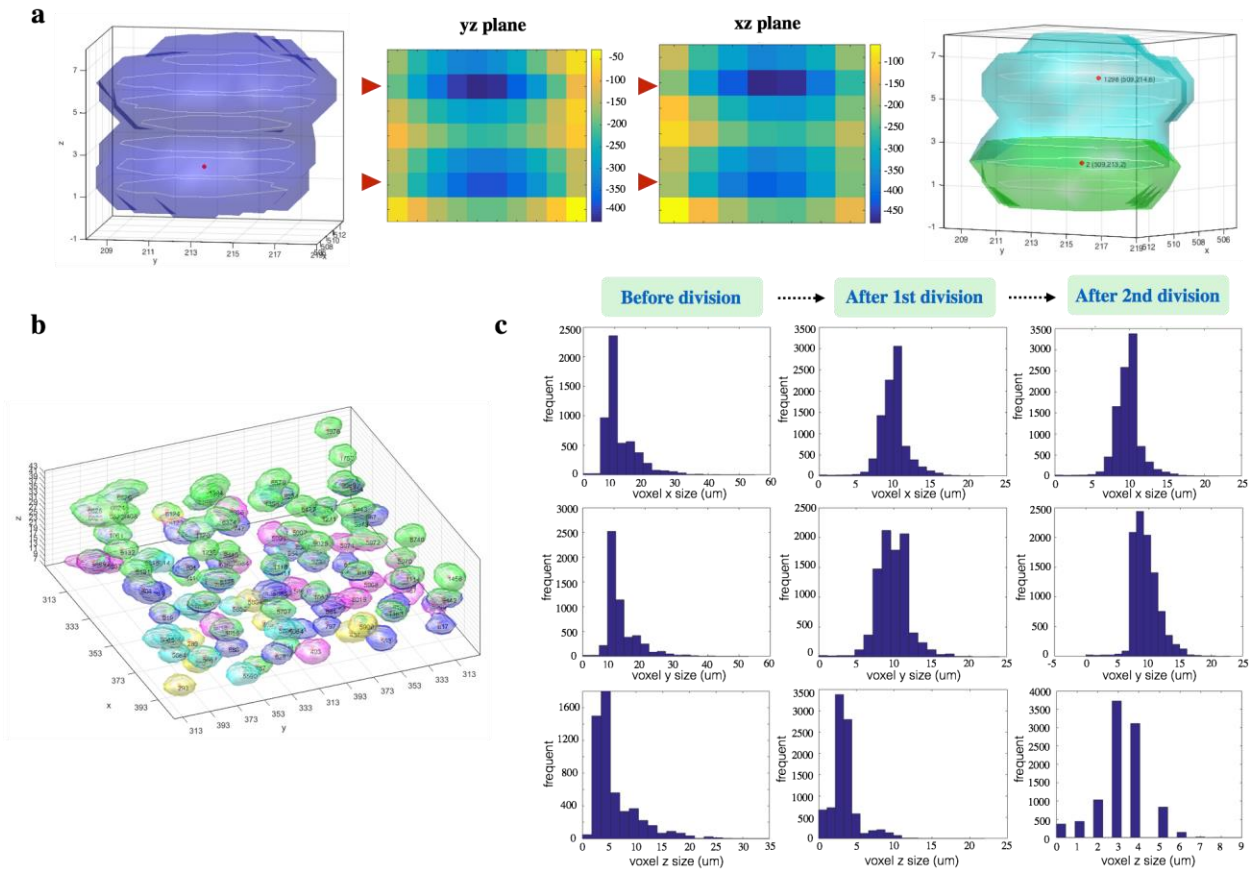


Figure 1.8. Second division step in z direction and the summary of blob size distribution before and after division steps. (a) The 3D nuclei blob (left) shows the nuclei overlapping in z direction which didn't divide appropriately in the first division step because there is no center position shift between neighboring Z slices. In the second division step, wavelet coefficient yz and xz plane are exanimated and search for center positions for every 3D nuclei. If two center positions are found on both yz and xz planes (center left and center right), original 3D nuclei blob will be divided into two new blobs from the middle plane of two center positions (right). **(b)** Remove the 3D nuclei blobs smaller then nuclei size × 0.2. **(c)** Blob size distributions in three axis directions (upper: x-direction, center: y-direction, lower: z-direction) before the first division step (left column), and after the first division step (center column), and after the second division step (right column) for all nuclei segmented using wavelet-based segmentation method in whole embryo. The final size distributions in three directions are all similar to normal distributions (right column).

1.2.3 Evaluation of segmentation protocol with synthetic and real images

To benchmark the wavelet and other methods that are coded into the WaveletSeg program, we developed tunable test data as well as a ground-truth data graphical user interface to manually segment 3D images for validation. In addition to the wavelet method, the following methods are included: point wise method [38], Otsu method [6], and Derivatives Sum (DS) method [39]. The validation datasets we used to evaluate segmentation performance included synthetic image datasets that were created by using a synthetic data generator (Figure 2.2), publicly open image datasets [40], and our ground truth image dataset determined by multiple rounds of manual segmentation. We also generated synthetic nuclei overlapping datasets to test the ability to divide overlapping nuclei.

Three evaluation criteria were applied to compare the accuracy of four segmentation methods. If a representative point was nearest-neighbor of a point in ground truth dataset and visa-versa, the object was regarded as a true positive (TP). If only the former condition was met, the segmented result is considered a false negative (FN). If only the latter condition was met, the object was regarded as a false positive (FP). Then the true positive rate (TP rate) is equal to the true positive number divided by the ground truth number, and Precision is defined as the TP number divided by the sum of TP number with FP number. TP rate defined in equation (1) represents the ratio of successful nuclei segmentation within all ground truth nuclei data, and Precision defined in equation (2) is the correct segmentation ratio in the nuclei segmentation result using a specific method. False negative rate is the false negative number divided by ground truth number which shows the missing segmentation results against ground truth data. We also use segmentation accuracy to estimate pixel-wise segmentation performance, which was defined as the overlapping region between ground truth segmentation region and segmentation region in pixel level.

$$TP\ rate = TP/P = TP/(TP + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

We use the Jaccard similarity index J (equation (3)) to define segmentation accuracy [40] which means the amount of overlap between the segmentation results S and ground truth annotation results R.

$$J(R,S) = \frac{|R \cap S|}{|R \cup S|} \tag{3}$$

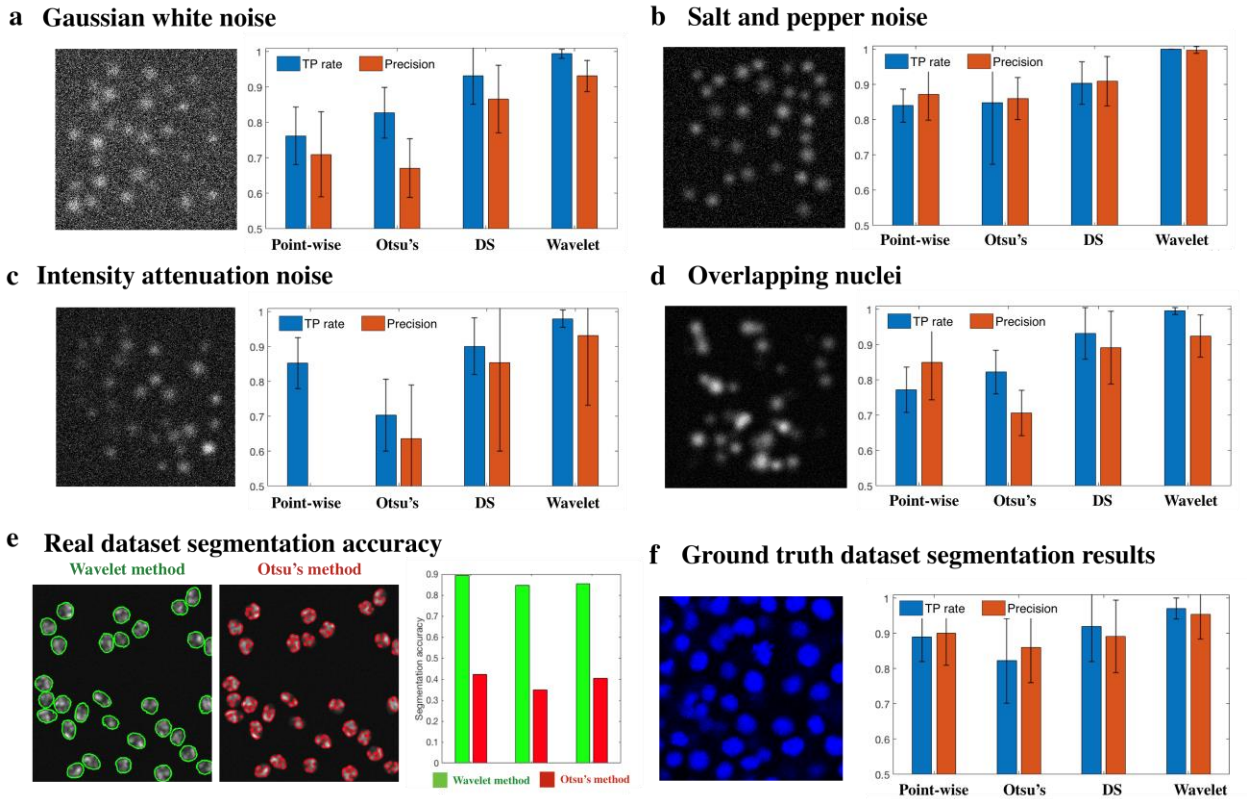


Figure 1.9. Performance criteria and segmentation accuracy of four segmentation methods on high level noisy synthetic images and ground truth image datasets. (a-d) True positive (TP) rate and Precision on four kinds of noisy synthetic images using three segmentation methods and wavelet method. Error bars indicate standard deviation. (a) Samples of synthetic images (N=20) Gaussian white noise and TP rate and Precision when applying point-wise method, Otsu method, DS method and wavelet-based segmentation method. (b) Samples of salt and pepper noisy synthetic images (N=20) and TP rate and Precision for four segmentation methods. (c) Intensity attenuation noisy synthetic image samples (N=20) and TP rate and Precision for four segmentation methods. (d) Samples of overlapping synthetic image samples (N=20) and TP rate and Precision for four segmentation methods. (e) Segmentation accuracy when applying Otsu method and wavelet-based segmentation method on three Fluo-N2DH-GOWT1 sample images. (f) True positive rate and Precision on five ground truth images when using point-wise method, Otsu method, DS method and wavelet-based segmentation method.

In test data (with test data generator described below) we added multiple different types of image noise that would normally be encountered in an imaging environment. For Gaussian white noise (Figure 1.9a), the wavelet-based segmentation method has the best TP rate and Precision followed by DS method. Otsu method obtained the lowest Precision and the point-wise method performed the poorest in TP rate. In salt and pepper noise cases (Figure 1.9b), both TP rate and Precision were equal to one when segmented by the wavelet-based method, and all other method results are lower than 0.9 in both TP rate and Precision. In images with Intensity attenuation (Figure 1.9c), the point-wise method has very high false positive rates and Otsu method performed the poorest in TP rate. In this test, the wavelet-based segmentation method had the highest TP rate and Precision. In the next step we examined the ability to divide nuclei in overlapping images for four segmentation methods (Figure 1.9d). Compared to three other segmentation methods, the wavelet-based method has the best ability to divide overlapping nuclei in both TP rate and Precision.

Next, we compared the pixel-level segmentation accuracy on publicly open image datasets using wavelet-based method and Otsu's method. Figure 1.9e shows that segmentation accuracy is above 0.85 for wavelet-based method, and smaller than 0.4 for Otsu's method. To further test the segmentation ability in real image datasets, we chose five zebrafish embryo images and selected sub-regions that we manually curated for our ground truth dataset and labeled the nuclei positions using our ground truth annotation tool (Figure 2.1). Both the TP rate and Precision of the wavelet method got best results (Figure 1.9f).

1.2.4 Outline of main structure and GUI of WaveletSEG research platform

Here we introduce WaveletSEG, the first automatic user-friendly integrated zebrafish embryo image processing analysis platform that can perform single nuclei scale to whole embryo scale quantification, topology and shape feature analysis during embryo development (Figure 1.10). WaveletSEG is an open-source Matlab-based imaging research platform and runs on Windows, Mac and Linux systems with complete GUI (Figure 1.11). The users can directly import either raw image files or data files with calculated results which helps them to check results or rerun processing with a different setting. 3D segmentation and quantification or other results such as embryo topology features can be easily displayed or saved in the WaveletSEG data visualization system. Users can also create scatter plots by selecting menu options to explore the relationship

between them. For the examples used herein, we show the fluorescent distributions of the BMP signaling component phosphorylated SMAD (pSMAD) through embryo DV axis, or the nuclei size spatial distribution from outer embryo layer to inside.

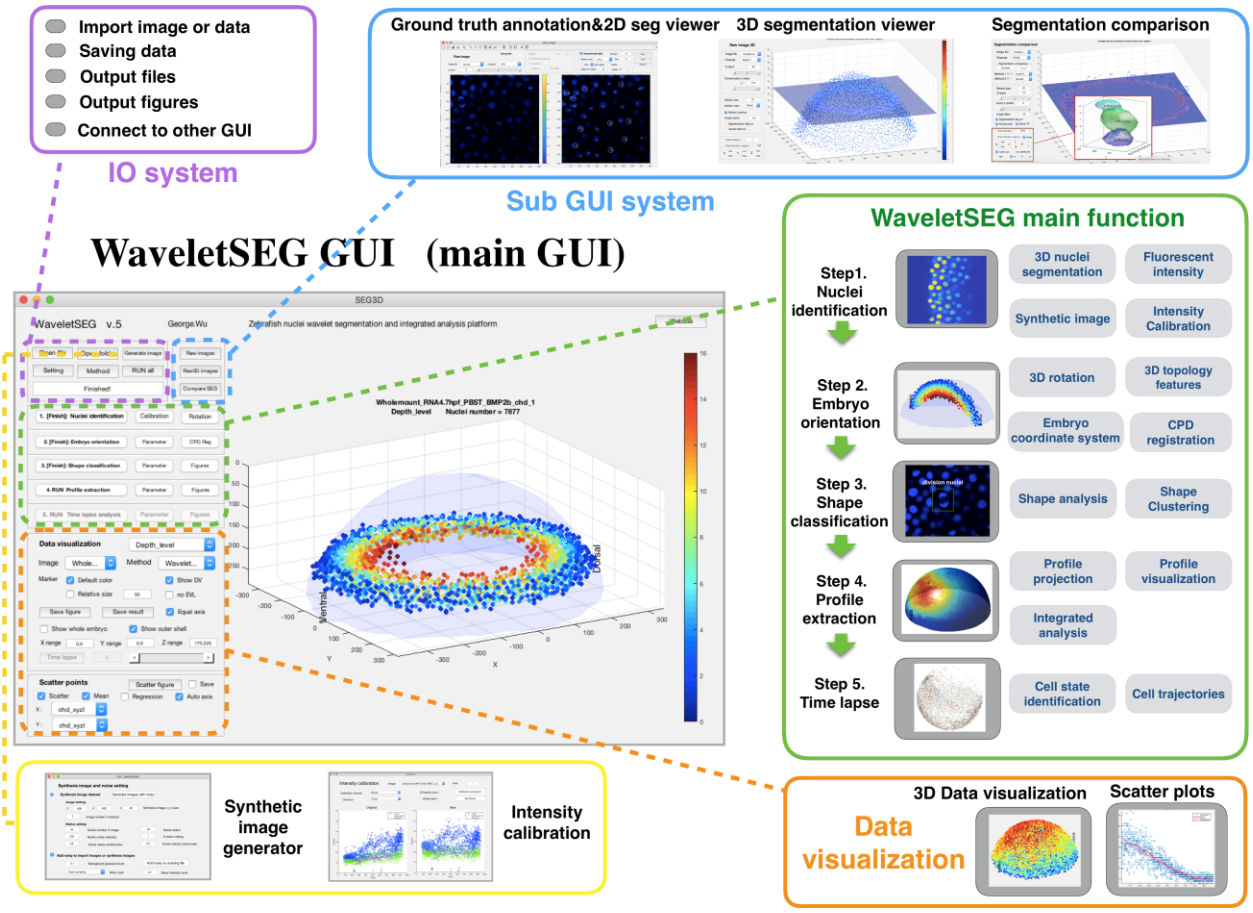


Figure 1.10. Graph show of WaveletSEG software main function structure and GUI. WaveletSEG is the fully automatically user-friendly image processing analysis platform which integrated all main and extension functions including IO system, extension GUI and data visualization system in one GUI. WaveletSEG main function block consists five main steps which can run independently including 1. Nuclei identification, 2. Embryo orientation, 3. Shape classification, 4. Profile extraction, and 5. Time lapse steps. In the nuclei identification step, user can do intensity calibration by using nuclei nearby intensity or do 3D after wavelet-based segmentation. In IO system block, User can directly import microscope image files or intermediate data files into WaveletSEG, and save or output data results or figures directly from the GUI. We also developed a set of segmentation validation tools in WavetSEG including synthetic data generator, 3D ground truth labeling system, 2D and 3D segmentation viewer sub-GUI, and segmentation method comparison extension GUI. In WaveletSEG data visualization block, 3D segmentation and quantification or other results such as embryo topology features can be easily displayed or saved in WaveletSEG data visualization system. User can also create scatter plots by selecting menu options to explore the relationship between them.

WaveletSEG consists of five main steps which can run independently: 1. Nuclei identification, 2. Embryo orientation, 3. Shape classification, 4. Profile extraction, and 5. Time lapse steps (Figure 1.11-12). In the nuclei identification step, users can do intensity calibration by using nearby nuclei intensity or do 3D rotation to adjust the embryo DV and AP axis in the right direction after wavelet-based segmentation. After the nuclei segmentation step, the embryo coordinate system is automatically created and each nuclei is given a coordinate list value, and the 3D topology feature for each nuclei was also calculated. CPD registration was instrumental in combining multiple image results into one representative result. In step 3, we can apply nuclei shape classification to analyze the 3D nuclei shape type and features which resulted from step 1 and 2. In the profile extraction step, we project the nuclei quantification and features into one averaged distributed sphere plane to enhance accuracy distribution. If an imported image is a 4D (3D + time) image, the user can run time-lapse analysis in step 5. In step 5, nuclei tracking is performed by segmenting all the nuclei in a time-lapse image sequence using wavelet-based segmentation. This is achieved one step at a time, and the spatiotemporal overlap between corresponding segmented regions in consecutive time steps is compared, providing information about coordinate values, topology features and shape classification.

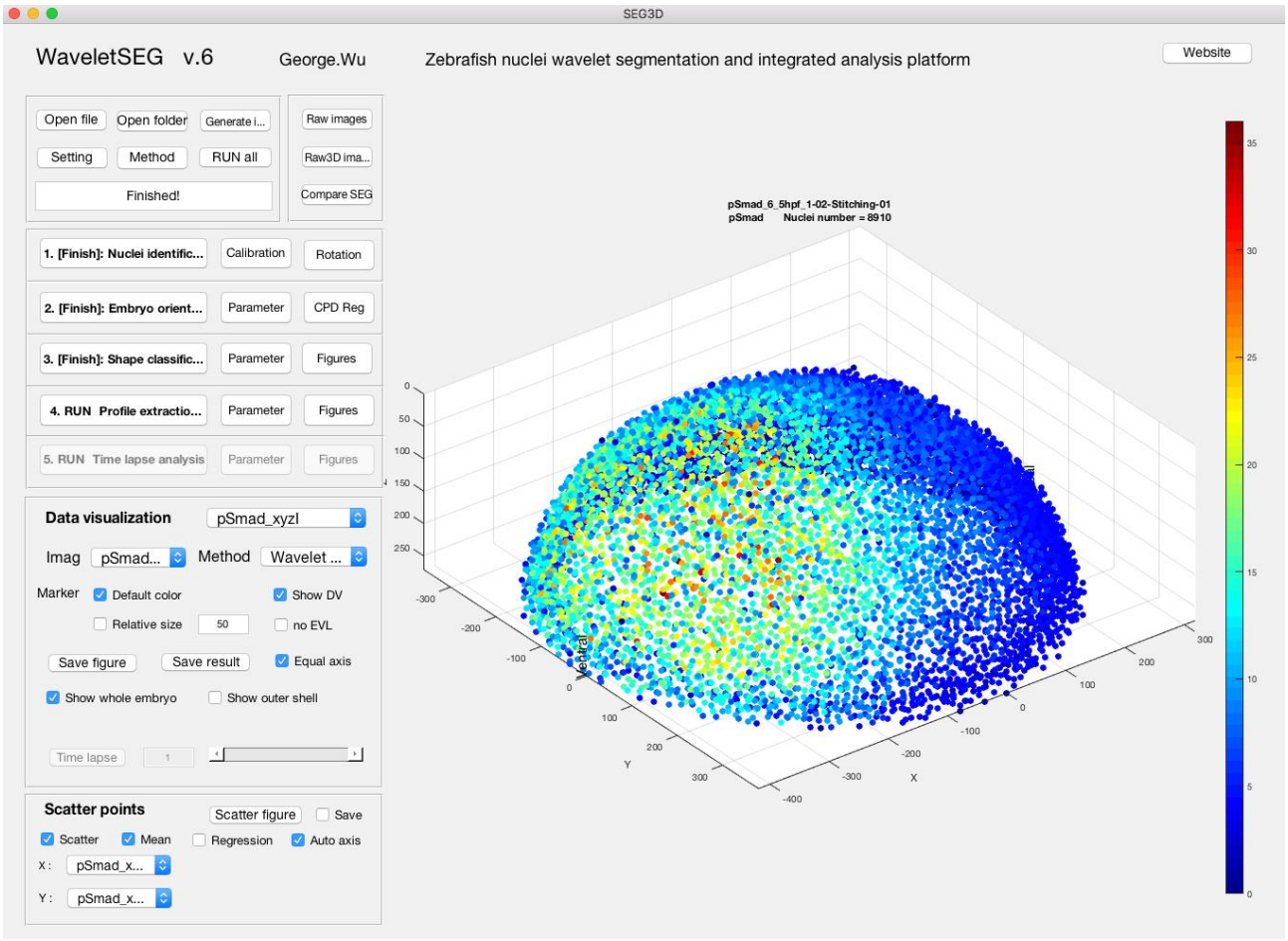


Figure 1.11. The interface of software WaveletSEG and data visualization. On the left side of WaveletSEG is the control panel, which can execute the five main steps or control the data visualization and output data and figures. Once the main steps are completed, "[Finish]" will appear inside the main step button. Users can execute this main step individually without starting from the beginning. Result data files and well-organized figures can be generated by clicking the "save result button" or "save figure button" on the control panel. The data visualization window visualizes most the results and enables 3D rotation and in and out zooming.

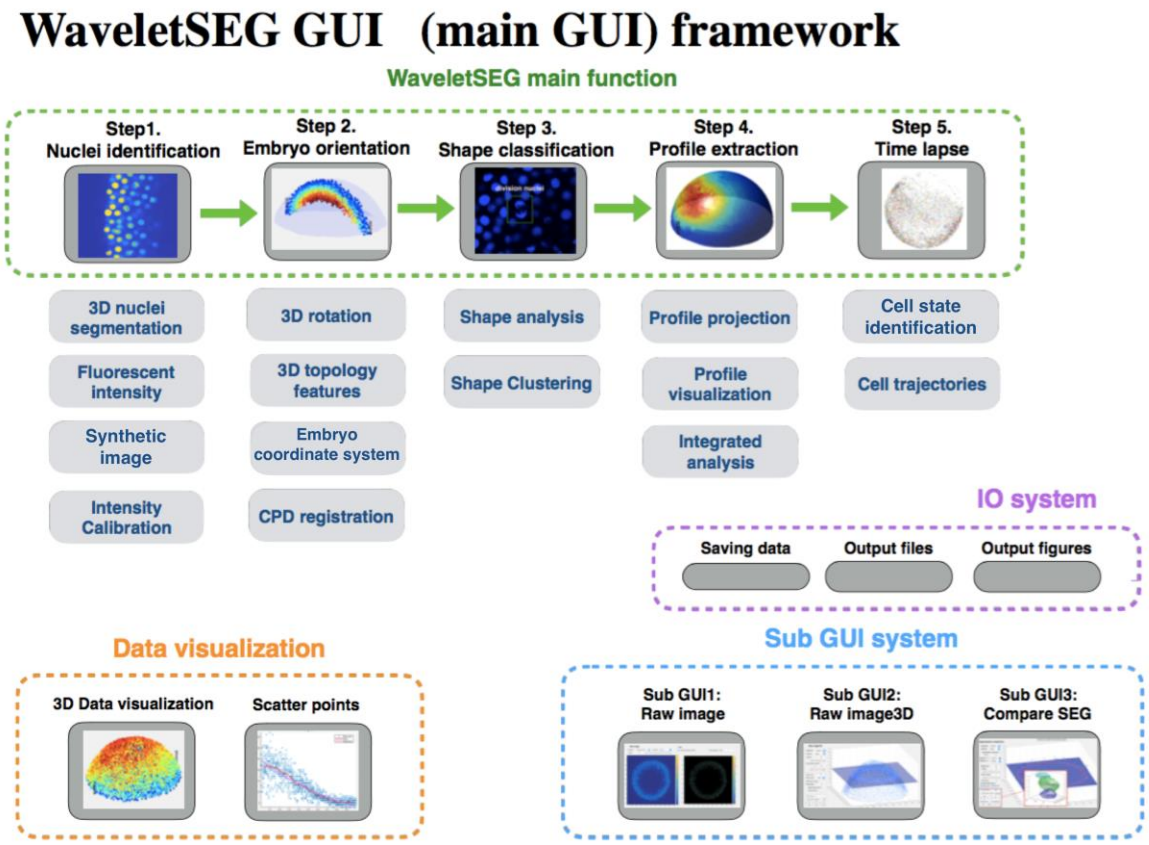


Figure 1.12. WaveletSEG main functions. WaveletSEG main function block consists five main steps which can run independently including 1. Nuclei identification, 2. Embryo orientation, 3. Shape classification, 4. Profile extraction, and 5. Time lapse steps. In the nuclei identification step, whole embryo nuclei are segmented using wavelet-based segmentation method, and averaged nuclear fluorescent intensity is also calculated. User can do intensity calibration by using nuclei nearby intensity or do 3D after wavelet-based segmentation. In IO system block, User can directly import microscope image files or intermediate data files into WaveletSEG, and save or output data results or figures directly from the GUI. In the embryo orientation step, individual embryo is rotated in 3D automatically according to some specific chemical such as chd mRNA. Embryo coordination value and 3D topology features for each nuclei is also calculated after rotation to provide positional information. To integrate multiple embryos, CPD registration is applied here and created one representative embryo. In shape classification step, we defined shape-based rules to classify nuclei into spherical nuclei, irregular shape nuclei, elliptical nuclei, and division nuclei to estimate cell proliferation and cell cycle phases. Unsupervised learning clustering such as SOM network is also included to cluster nuclei based on nuclei size and shape. In profile extraction step, we project the nuclei quantification and features into one averaged distributed sphere plane to enhance accuracy distribution. We also developed a set of segmentation validation tools in WavetSEG including synthetic data generator, 3D ground truth labeling system, 2D and 3D segmentation viewer sub-GUI, and segmentation method comparison extension GUI. In WaveletSEG data visualization block, 3D segmentation and quantification or other results such as embryo topology features can be easily displayed or saved in WaveletSEG data visualization system. User can also create scatter plots by selecting menu options to explore the relationship between them.

We also developed a set of segmentation validation tools in WaveletSEG including a synthetic data generator, a 3D ground truth labeling system, 2D and 3D segmentation viewer sub-GUI, and segmentation method comparison sub-GUI. It is challenging and time-consuming to label the 3D nuclei center as ground truth image because the user may have to check repeatedly to find a nuclei center among several Z-planes, and distinguish it from other nuclei. To overcome this challenge, we developed a convenient interactive 3D ground truth labeling interface that will both mark nuclei center labeling and non-center labeling by mouse clicking in jump-out GUI amongst a range of Z planes. In 2D and 3D segmentation viewer, users can easily validate the nuclei center position and mask with raw images, or display a single nuclei 3D shape. Using the segmentation comparison sub-GUI, users can also easily validate segmentation results by utilizing different methods and display TF, NP, or NF nuclei directly. In WaveletSEG, user can also export 2D or 3D segmentation masks as the machine learning training datasets for further object detection or semantic segmentation studies.

1.2.5 Zebrafish embryo quantification and registration and coordination system, and 3D topology features through time

Development of an animal embryo involves molecular signaling to coordinate cellular rearrangement and proliferation. To quantify these processes in zebrafish, we calculated zebrafish embryonic properties along the DV axis, AP axis, embryo dimensions including thickness, and cell level properties including cell division state. These properties provide a complete embryo picture (Figure 1.13-14) including 3D embryo geometry and quantitative signaling information including the levels of pSMAD in each segmented nucleus [42-43]. After segmentation, 3D rotation and deletion of the EVL cell later, we generated 2D hemispherical surfaces and projected every nuclei onto the 2D surface to obtain intensity distributions of nuclear properties (Figure 1.15). Whole embryo, sub region quantification results, and DV axis pSMAD intensity distribution in this sub region, or coherent point drift (CPD) registration to multiple embryos are also computed in WaveletSEG (Figure 1.16-17).

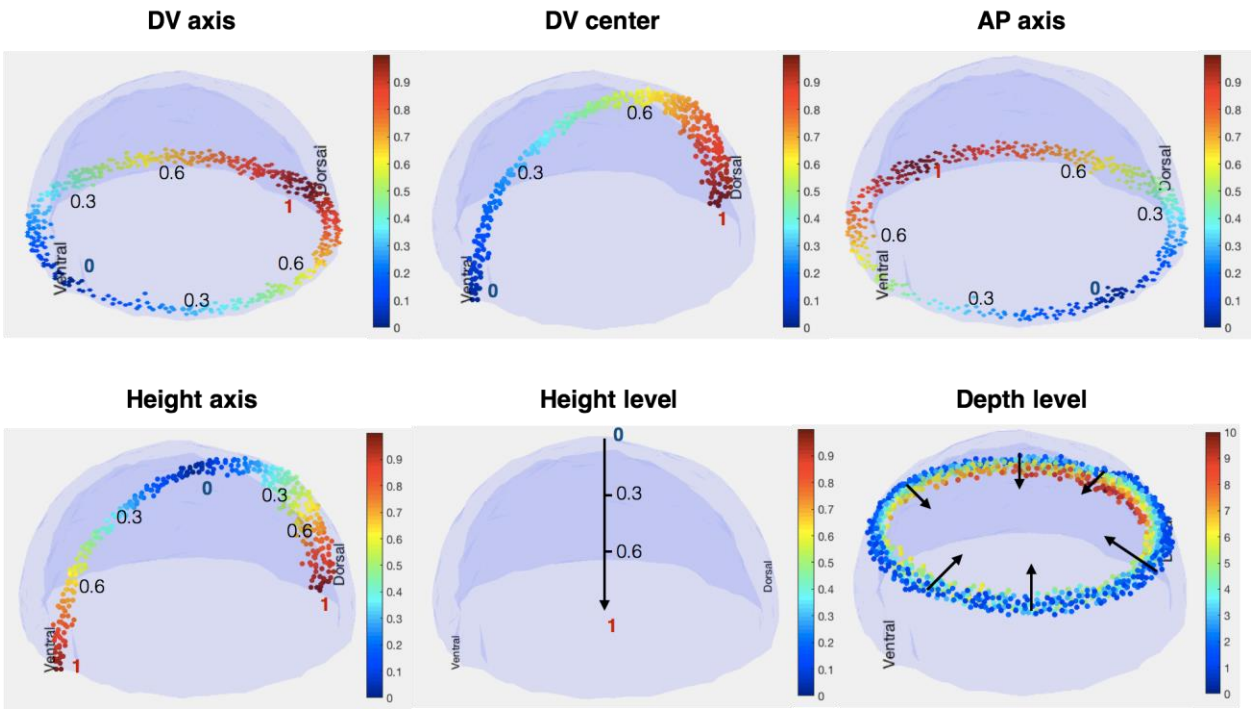


Figure 1.13. Zebrafish embryo coordinate system in ZebEmbIM. We developed the zebrafish embryo coordination system including embryonic axes such as DV axis, AP axis, and DV center axis, Height axis, Height level axis, Depth level axis. If time-lapse images are available, there is additional axis: time stage. The first five coordinates are values between zero to one, Depth level is the layer number counting from the outer shell, and the Time stage started from 1. DV center axis is defined by connecting ventral point to dorsal point through the top of embryo, angle is calculated and DV center value is from 0 to 1. Unlike the Height level axis which described the vertical distance from top to button of Embryo, we defined Height axis through the line connecting ventral point to dorsal point through the top of embryo to describe the embryo developmental process relative to which nuclei will grow from the top of the embryo to the embryo button.

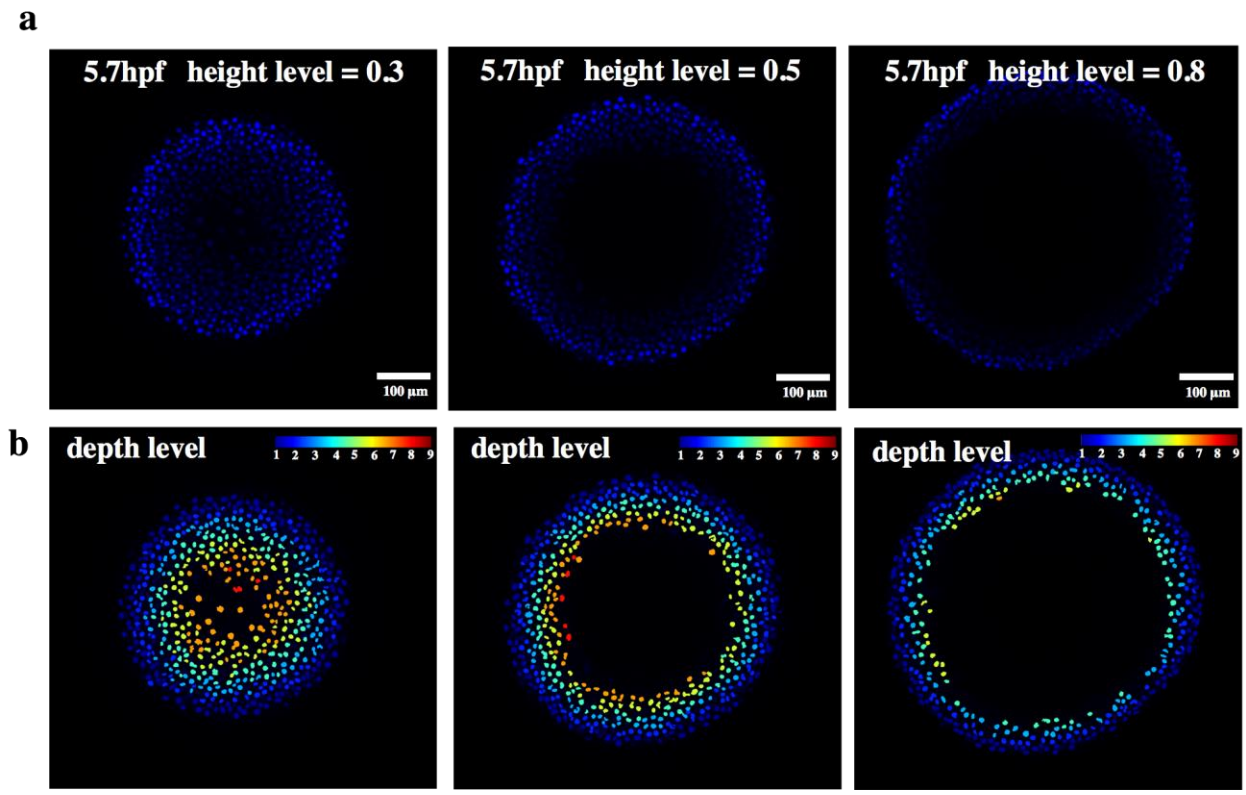


Figure 1.14. Whole embryo z-slice nuclear DAPI staining and radial depth level. (a) Z-slice DAPI nuclear stain on height level = 0.3 (left), 0.5 (center), 0.8 (right) in 5.7 hpf zebrafish embryo. Image data were acquired with Zeiss LSM 800 confocal microscope with 20X water lens. **(b)** Z-slice radial depth level for each nucleus on height level = 0.3, 0.5, 0.8 in 5.7 hpf Zebrafish embryo.

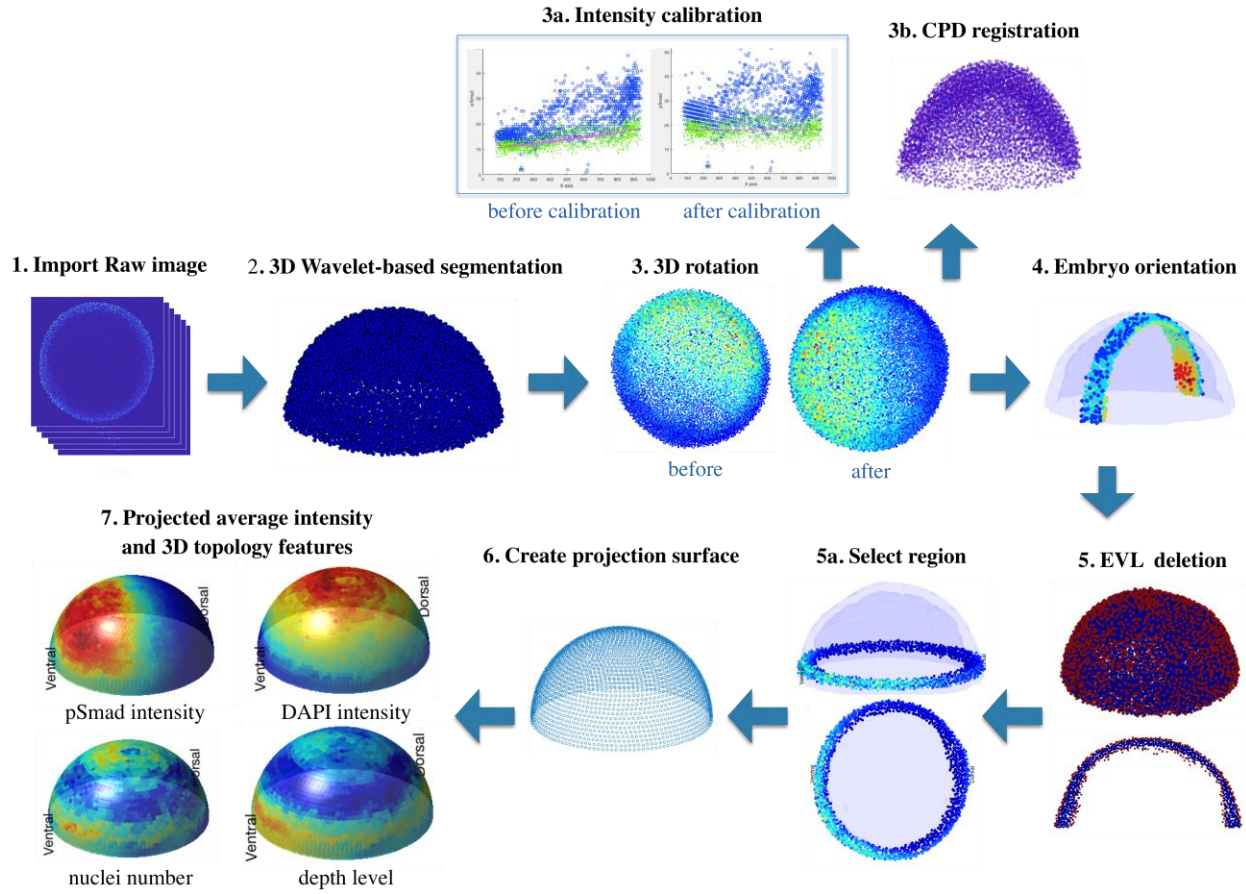


Figure 1.15. Data workflow from 3D whole embryo raw image to projected average intensity distribution in WaveletSEG. Step1: Import 3D raw image with both DAPI stained channel and pSMAD channel. Step2: Segmented nuclei point cloud by using wavelet-based segmentation method. Step3: Automatically rotation of embryo by using chd or pSMAD channel embryonic spatial distribution. Step3a (optional): Apply intensity calibration by using nearby nuclei intensity or do 3D rotation to adjust the embryo DV and AP axis in the right direction after wavelet-based segmentation. Step3b (optional): To merge embryos in the same stage into a representative embryo one by one using CPD. Step4: After 3D rotation and embryo center point is moved to position zero, 3D embryonic coordination system including DV-axis, AP-axis to provide embryo coordination axes for geometry features and quantification results. Step5: Automatically remove the outermost layer and the innermost layer from embryo using depth level information. Step5a (optional): data analysis and data visualization in specific sub-region of whole embryo. For example: zebrafish embryo marginal region. Step6: We generated a hemispherical surface with averaged distributed reference point using regular placement method, and projected each nucleus into this 2D hemispherical surface (projection surface). Step7: We projected the nuclei into reference points with the closest spherical coordination system angle (θ, φ) and create projected average intensity and 3D topology.

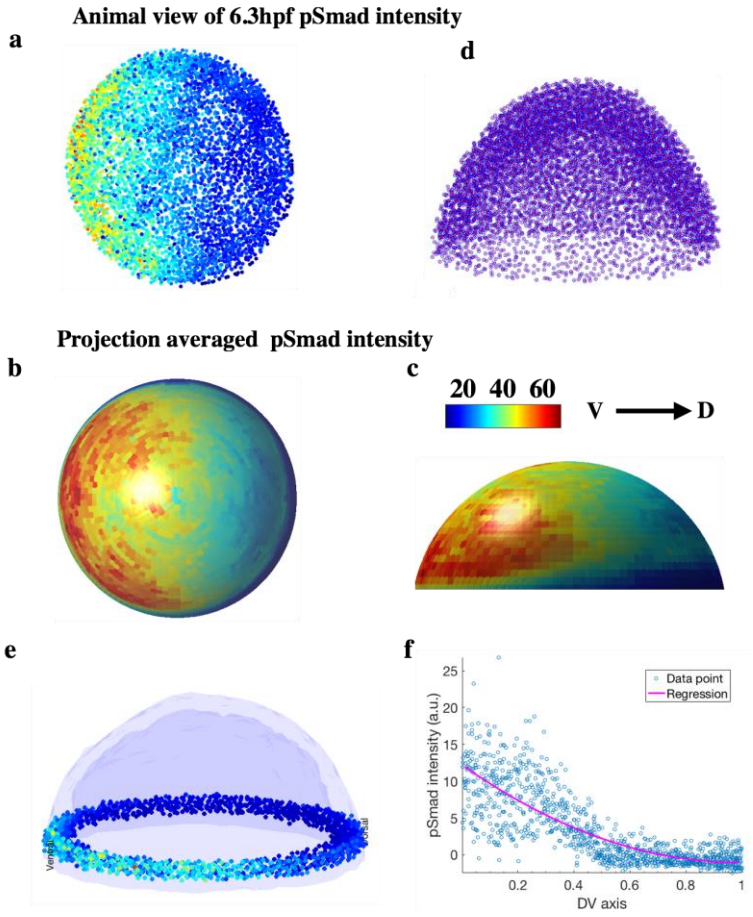


Figure 1.16. Quantification of nuclear P-Smad gradient in the zebrafish embryo. (a) Animal view of nuclear pSMAD intensity of all nuclei from the 6.3hpf embryo (left), and animal and lateral view of averaged pSMAD intensity by projecting to uniform distributed semisphere surface using WaveletSEG **(b, c)**. **(d)** Embryo alignment using coherent point drift (CPD) method in WaveletSEG. **(e)** Data visualization of nuclear pSMAD intensity for 6.5 hpf Zebrafish embryo in marginal region using WaveletSEG. Light blue and dark blue transparent shells indicate inner and outlier boundary surfaces of all nuclei in embryo. **(f)** Distribution of nuclear pSMAD intensity vs DV axis in embryo marginal region in **(c).** Blue dot and red line indicate individual nuclei intensity and averaged intensity along DV axis.

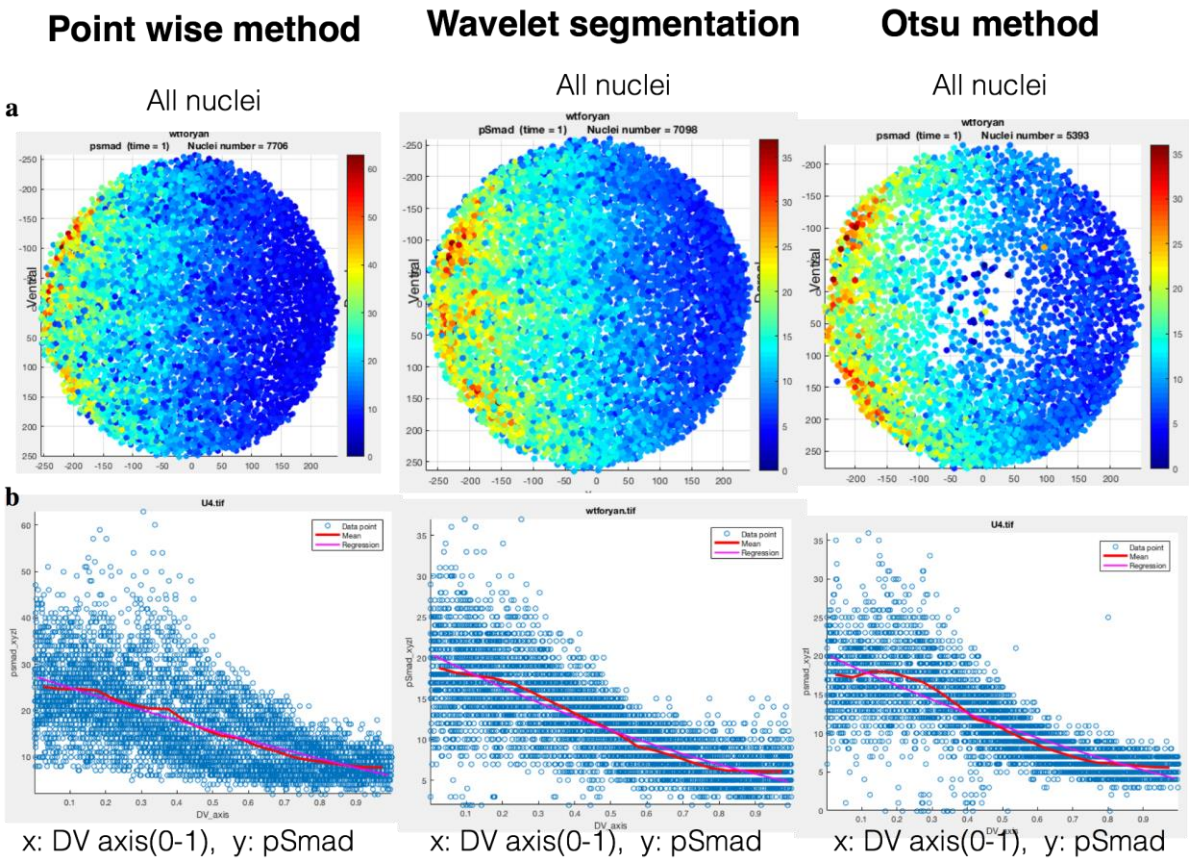


Figure 1.17. Segmented nuclear P-Smad expression spatial distribution for individual zebrafish embryos. **(a)** Animal pole views of segmented nuclear P-Smad signal spatial distribution for a whole-mount zebrafish embryo using pointwise method (left), wavelet segmentation (center), and Otsu's method (right). **(b)** Segmented nuclear P-Smad expression on the DV axis, mean value (red line), and regression line (pink line) based on the three segmentation methods.

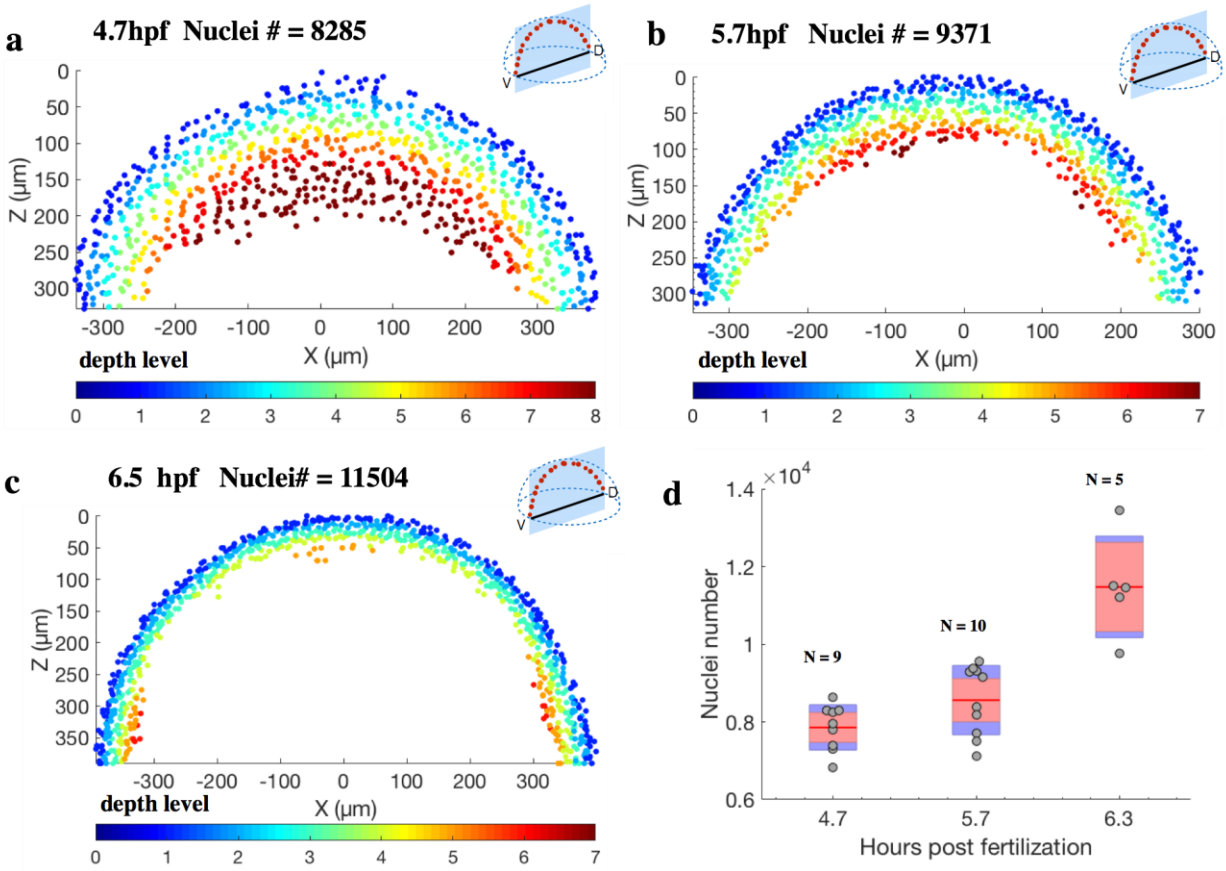


Figure 1.18. Spatial analysis of embryo coordinate system in zebrafish embryo between 4.7, 5.7, 6.5 hpf. (a, b, c) Depth level spatial distribution for nuclei on DV direction cross section at 4.7 hpf **(a)**, 5.7 hpf **(b)**, and 6.5 hpf zebrafish embryo **(c)**. **(d)** Nuclei number counts of embryo at different developmental times (4.7 hpf: N = 9, 5.7 hpf: N = 10, 6.5 hpf: N = 5). Gray dots are nuclear counts for individual embryos. The red lines, red-boxed regions, and blue regions show the mean nuclei numbers, one standard deviation, and 95% confidence intervals, respectively.

In early embryo development of the zebrafish, cell rearrangements guide the establishment of embryonic axes and layers [44-45] and embryonic cells spread over the yolk mass while the blastoderm thins. The radial intercalation movement between deep cell layer (DCL) and enveloping layer (EVL) is regarded as the main driving force for epiboly, and the cell density increase near the EVL plays an important role in this process [46]. Figure 1.18 shows the Depth level spatial distribution for nuclei in the DV direction cross section of 4.7 hpf (N=9)(left), 5.7 hpf (N=10)(center) and 6.5hpf Zebrafish embryo (N=8)(right) (Figure 1.18d). They have significant differences not only in nuclei number, but also in the depth level spatial distribution, nuclei density

and thickness. In figure 1.18c, we observed the cell accumulation (Hypoblast) and involution movement near the blastoderm margin.

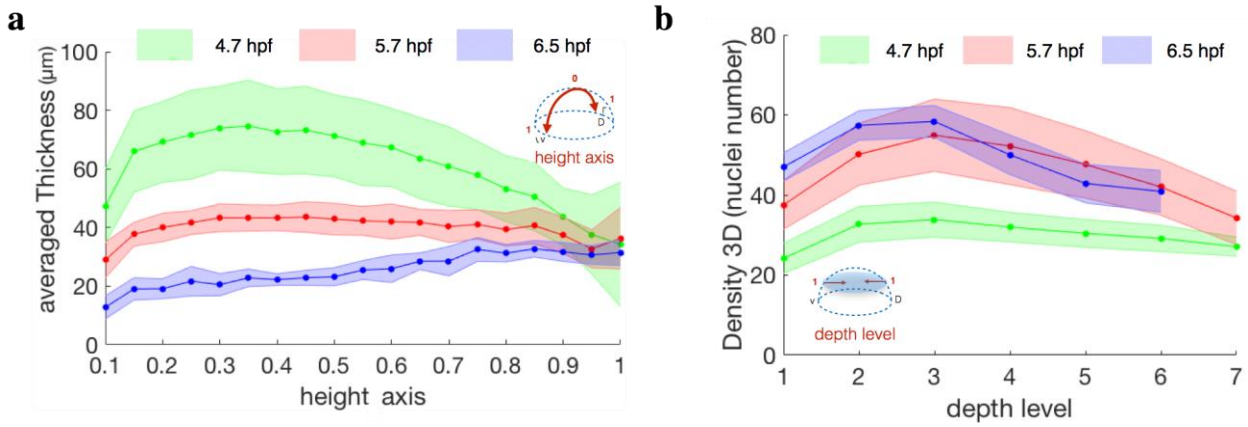


Figure 1.19. Spatial analysis of 3D topology features between 4.7, 5.7, 6.5 hpf zebrafish embryo. (a) Difference between threads of average thickness of all nuclei along the embryo height axis in 4.7, 5.7, and 6.5 hpf embryo developmental time. **(b)** Difference between threads of 3D density of all nuclei along the embryo height axis in 4.7, 5.7, and 6.5 embryo developmental time. Green, red, and blue lines indicate mean, and light green, light red, and light blue regions indicate one standard deviation region.

Emboly is another principle coordinated cell movement to form hypoblast which contains the involution and intercalation movements in the early gastrulation stage. Involuting cells move and accumulate in the blastoderm margin to narrow and elongate the embryonic axis, eventually form endoderm and mesoderm that characterizes the gastrulation stage in Zebrafish [47-48]. Convergence and extension of epiblast deep cells result from intercellular space decrease and cell density contributed in the internalization of hypoblast cells from margin. For further investigation of spatial and temporal analysis of embryonic structure distribution, we introduced eight 3D topology features including Size, Thickness, Density2D, Density3D, Neighbor2D, Neighbor3D, H sorting, EVL (Figure 1.20) which can be directly calculated in WaveletSEG. For example, figure 1.19a can be used to describe the averaged thickness thread in epiboly process in 4.7, 5.7 and 6.5 embryo developmental time. We also found that the thickness is increased near margin position in 6.5hpf, which is the signal for emboly process. In figure 1.19b, averaged cell density are decreased after depth level 2 (one layer EVL cell). The gradual separation of the EVL lineage correlates with the flattening of the blastoderm on the yolk cell and is accompanied by both an increase in tension

and cell shape changes within the EVL [49]. It has been suggested that the separation of the EVL as a lineage might be the consequence of increased tension that causes EVL cell divisions to occur preferentially within the plane of the EVL. At the end of gastrula stage the cell density is decreased at the ventral side of embryo.

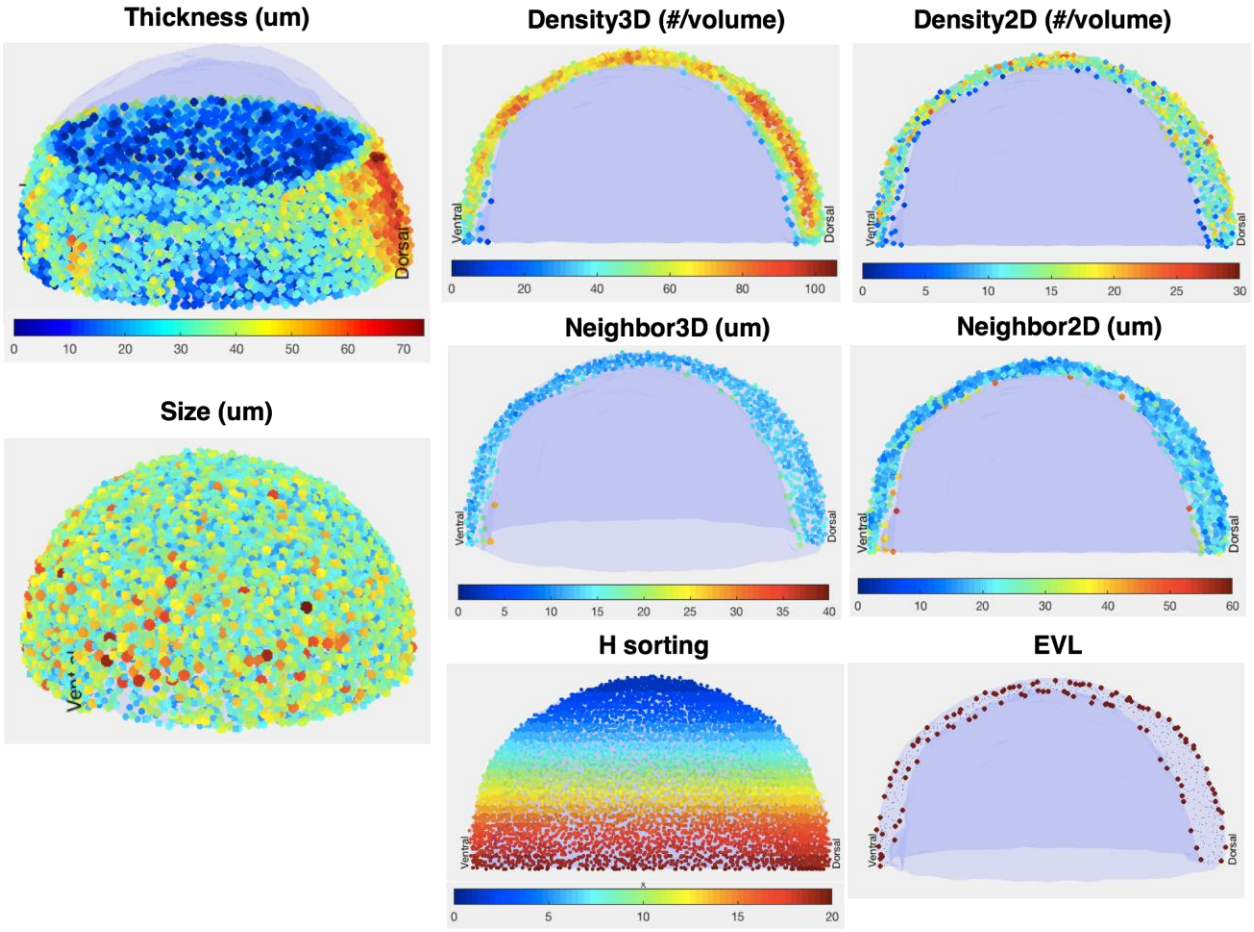


Figure 1.20. 3D topology features used in WaveletSEG. Eight 3D topology features are defined in WaveletSEG software to quantitative describe topology changing through embryo developing process. Nuclei size is the total pixel numbers for every 3D segmented nuclei. We used four indexes to describe the spatial nuclei density distribution for the embryo, including density 3D, density 2D, neighbor 3D and neighbor 2D distance. Density is the total nuclei number with distance smaller than 30 pixel around one nuclei, and Neighbor distance is the average distance of the nuclei to the closest two nuclei. The difference between 3D and 2D is that we only calculate density in the same layer (depth level) for density 2D and neighbor 2D distance. Thickness is defined as the vertical distance from the innermost layer of embryo. The outer epithelial monolayer of a zebrafish embryo is called the enveloping layer (EVL). Here we defined the inner EVL and outlier EVL as the layer with depth level equal to one or the maximum values. H_sorting is the index to describe the dynamic process of nuclei movement by dividing every 500 nuclei from the top of the embryo to the embryo button.

1.2.6 Spatial and temporal analysis of cell cycle phase pattern results

Cell proliferation and division spatiotemporal pattern plays a key role in embryo morphogenesis to regulate the embryo shape and growth. Additionally, morphogenesis also effects the cell division rate and orientation [50-51]. Experimental and theoretical evidence shows that cell proliferation is associated with tissue cell density, and both cell division and cell rearrangement is suppressed in densely packed embryonic tissue [51]. Nuclei segmentation and shape classification results have been used to estimate cell proliferation and cell cycle phases [52] because nuclei are spherical during early prophase, irregular shape with nuclear envelope break down, and become ellipsoidal shape in metaphase [53]. Here we created the cell-cycle phases pattern in zebrafish embryo based on nuclei shape features (Figure 1.21) and analysis the spatial distribution for whole embryo in different time stages.

Figure 1.21b and figure 1.21c show the prophase and prometaphase nuclei count ratios at 4.7, 5.7, and 6.5hpf. We found that in all time stages, the ratio of prophase nuclei decreases from the top of the embryo to the bottom, and the prometaphase nuclei ratio increases from between 0.2 and 0.3 at 4.7 hpf to between 0.3 and 0.5 at 6.5 hpf. It showed that cell proliferation rates are low in the animal region of the embryo and higher in the zebrafish margin. These results are consistent with reference [50] that shows abundant mitotic cells at 50% epiboly and especially during the shield stage.

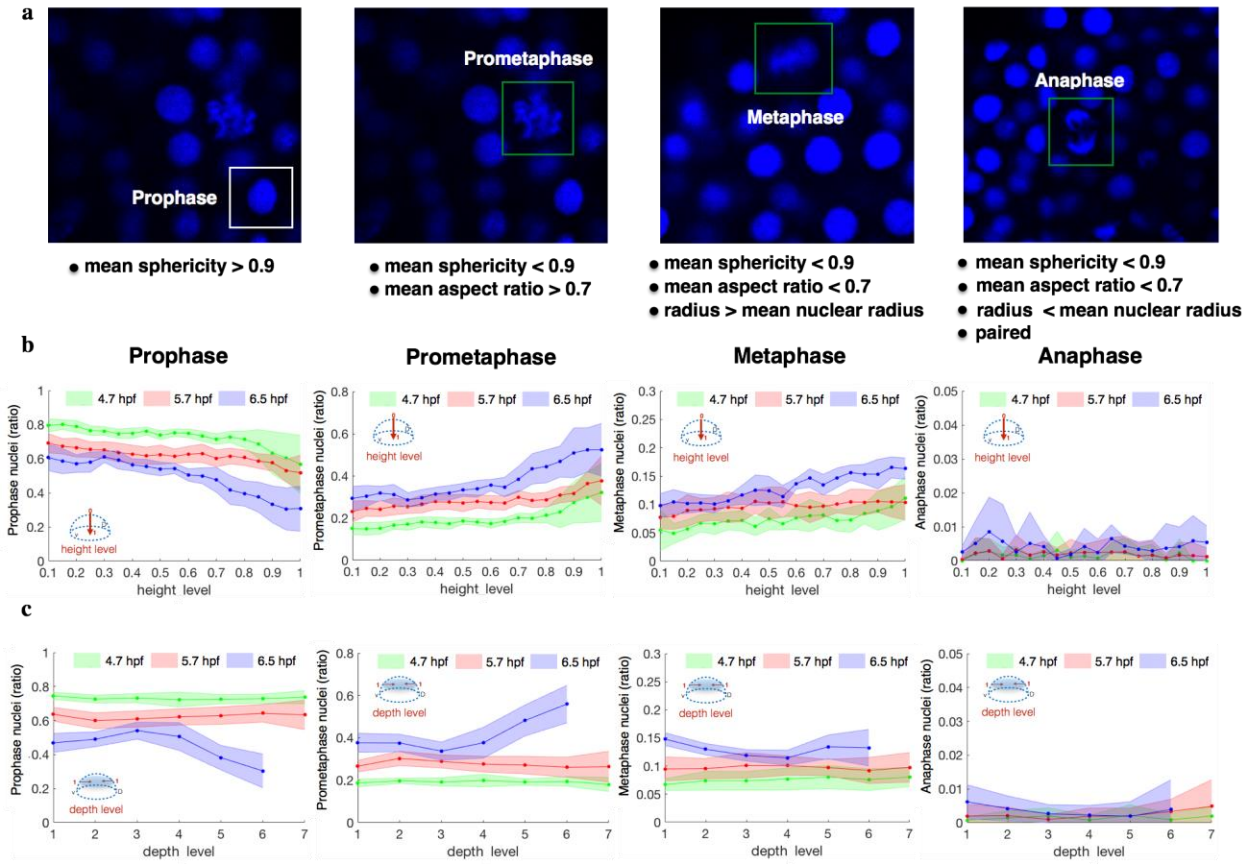


Figure 1.21. Cell cycle phase patterns based on nuclei shape classification results. (a) Examples of cell cycle phase in mitosis based on nuclei shape classification results, and the defined shape-based rules for cell cycle phase. **(b)** Nuclei count ratios for prophase, prometaphase, metaphase, and anaphase nuclei along the embryo height level in 4.7, 5.7, and 6.5 embryo developmental time. **(c)** Nuclei count ratios for prophase, prometaphase, metaphase, and anaphase nuclei along the embryo depth level in 4.7, 5.7, and 6.5 embryo developmental time. Green, red, and dark blue lines indicate mean value, and light green, light red, and light blue regions indicate one standard deviation region.

1.3 Methods

1.3.1 2D Wavelet transform

1D discrete wavelet transforms (1D DWTs) and 1D continuous wavelet transforms (1D CWTs) are the most commonly applied wavelet tools in 1D signal processing and time series analysis, and are used in signal decomposition in both time domain and frequency domain [54-55]. 1D CWT lies in decomposition in discrete frequency domain and continuous time domain, which is different from 1D DWT (Figure 1.22). Assuming that f(t) is a continuous and integrable time-frequency

signal, the 1D continuous wavelet transform (1D CWT) for f(t) can be represented by the following integral formula (equation (4),(5)):

$$\mathrm{K}_{1D}(s,a) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} f(t)\psi(\frac{t-a}{s})dt \quad (4)$$

$$\int_{-\infty}^{\infty}|\psi(t)|dt < \infty \; and \; \int_{-\infty}^{\infty}|\psi(t)|^2 dt < \infty \quad (5)$$

Where ψ(t) is the wavelet mother function which must satisfy the continuous and differentiable condition in all points with zero mean and square norm. Commonly used wavelet mother functions include the Mexican hat function, Meyer function, and Morlet function. a and b are the wavelet translation factor and scale factors that represent the location shift of the mother wavelet function through time, and the dilation ratio of the mother wavelet. The output of the CWT is the wavelet coefficient matrix that is the wavelet coefficient value K_1D (s,a) in a specific translation factor a (horizontal axis) and scale factor b (vertical axis), that is also called a daughter wavelet subset.

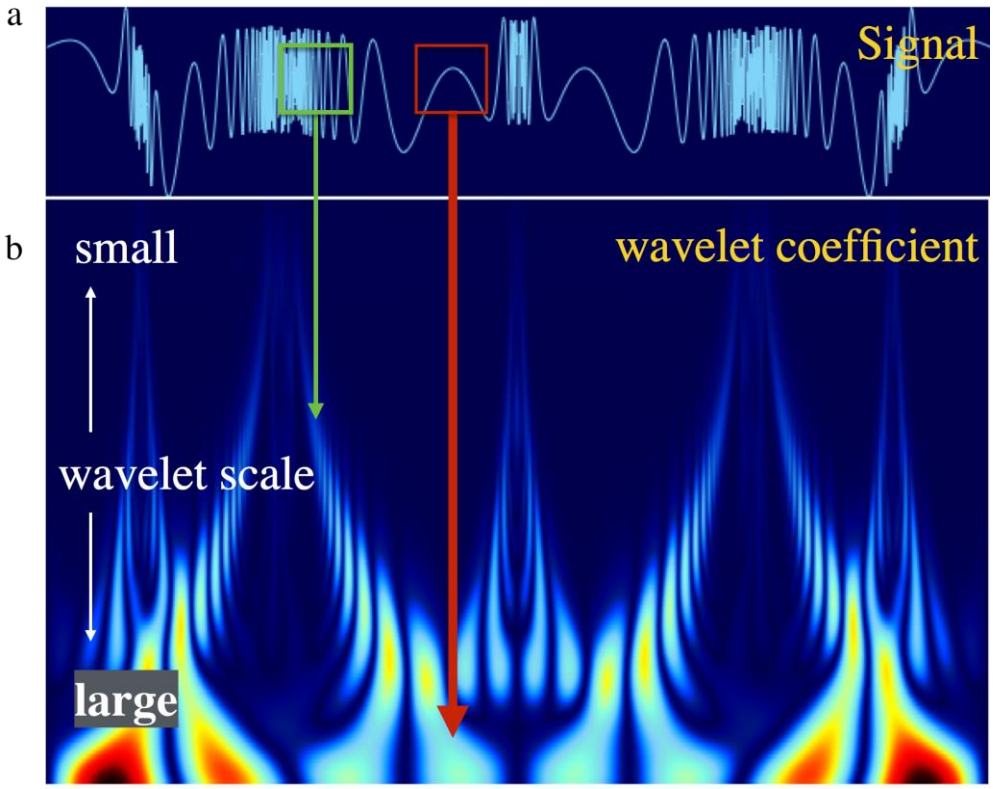


Figure 1.22. One-dimensional signal and its corresponding wavelet coefficient matrix. (a) One-dimensional signal includes a mix of low-frequency and high-frequency regions. The green and red boxes and arrows show high- and low-frequency regions of the signal, respectively. **(b)** Contour plot of wavelet coefficient matrix obtained from applying CWT with a series of wavelet scale factors. The horizontal axis represents the same axis as the one-dimensional signal, and the vertical axis represents wavelet scale factors from small (top) to large (bottom). Color represents the magnitude of the corresponding wavelet coefficient (blue: low, red: high). The ridges in the wavelet coefficient matrix can be correlated with low-frequency or high-frequency regions. For example, the identified ridge from the high-frequency regions (green box) corresponds to a smaller wavelet scale, and the identified ridge from the low-frequency regions (red box) corresponds to a larger wavelet scale. This provides an efficient way to isolate and to detect specific frequency regions.

We extend the 1D CWT to the two-dimensional continuous wavelet transform (2D CWT) and apply it to the spatial domain of the image56. Supposing that f(x,y) is continuous and differentiable 2D image data, we choose the 2D Mexican hat function as the wavelet mother function and perform 2D CWT with the translation factors a and b and the wavelet scale factor s. $\psi(x,y)$ is the 2D Mexican hat function and K_2D (s,a,b) is the wavelet coefficient matrix (equation (6),(7)):

$$\mathrm{K}_{2D}(s,a,b) = \frac{1}{\sqrt{s}}\iint_{-\infty}^{\infty} f(x,y)\psi(\frac{x-a}{s},\frac{y-b}{s})dxdy \quad (6)$$

$$\psi(x,y) = \frac{1}{\pi\sigma^2}(1-\frac{1}{2}(\frac{x^2+y^2}{\sigma^2}))e^{-\frac{x^2+y^2}{\sigma^2}} \quad (7)$$

Figure 1.23 shows the nuclei fluorescence microscopy raw image and wavelet coefficient matrix with different scale factors after the 2D continuous wavelet transform. Following application of the wavelet function, an efficient search algorithm that we developed compares different scales of CWT coefficients and marks shape-similar regions as nuclei masks in nuclei position(s) by looking for several contour lines in the CWT coefficients diagram. The similarities between maps are then calculated to determine the presence of a nuclei.

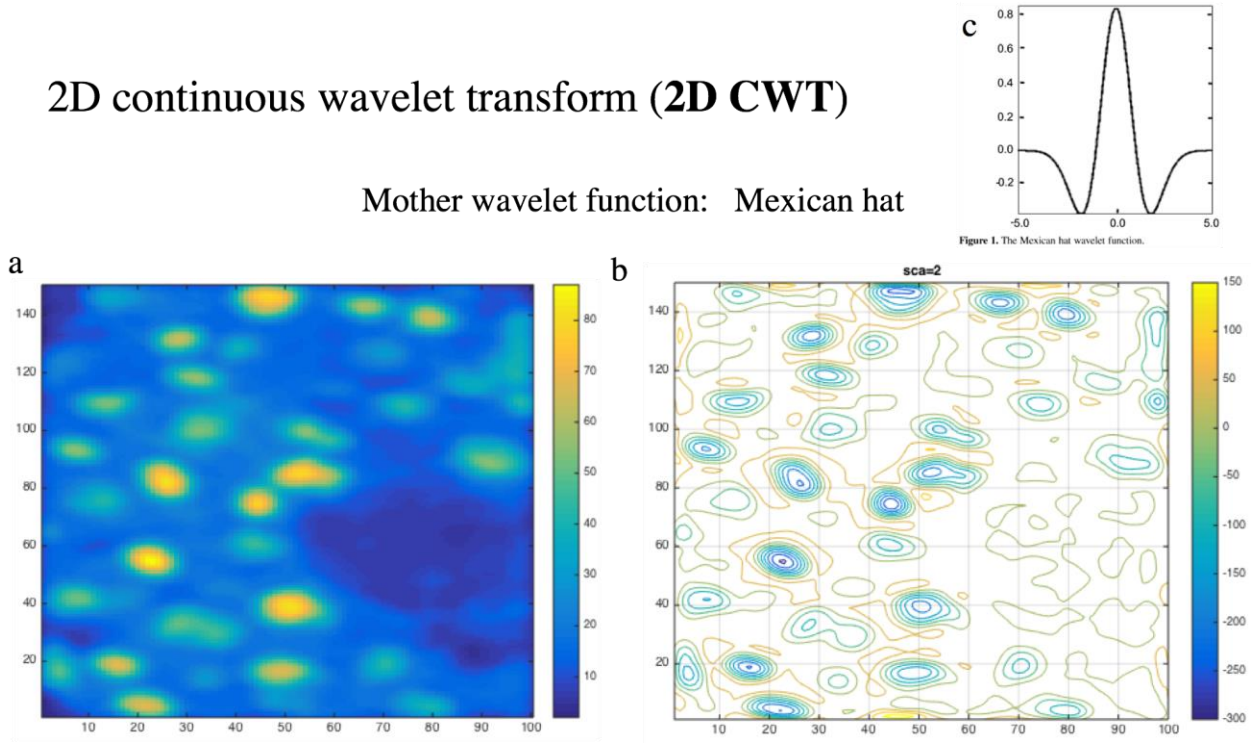


Figure 1.23. Two-dimensional image, its corresponding wavelet coefficient matrix, and mother wavelet function. (a) Two-dimensional nuclear DAPI staining raw image. **(b)** Contour plot of wavelet coefficient matrix from 2D CWT with specific scale factor on raw image. Color represents intensity of corresponding wavelet coefficient. Contour boundaries in this matrix can be correlated to nuclear boundaries. **(c)** We chose the most common Mexican hat function as the mother wavelet function for the wavelet-based segmentation method.

1.3.2 First and second division on XY wavelet coefficient plane, XZ and YZ plane and delete small object

For this purpose, we designed division steps in two stages to divide overlapping nuclei or overlapping 2D masks using wavelet coefficient maps. For each 3D nuclei object that we obtained from the previous step, we compared Z-slice wavelet coefficient planes and found the center of each plane which is local minimum. If there is more than one wavelet coefficient local minimum or the displacement of center is bigger than the nuclei radius on neighboring Z slices wavelet coefficients, we assigned a new nuclei object by cutting the median line between two centers, or assigned new nuclei in the case of center displacement. Using the center of Z-plane wavelet coefficients, most y or z direction overlapping can be divided.

In the second-stage division step, we divided z-direction overlapping by checking both the xz plane and the yz plane wavelet coefficient. If there were more than local minimum points present in both xz and yz plane, a wavelet coefficient division plane was set in the center plane between the two centers, and new 3D nuclei object was created. In the last step of the wavelet-based segmentation protocol, we removed the 3D nuclei voxels where the size was smaller than nuclei size × 0.2.

1.3.3 Image synthesis, noise adding process and nuclei overlap

To provide a dataset of synthesized images in validating segmentation methods, we developed a convenient interface to generate a synthesized image. Users can define synthesized image's 3D size and the number of image and synthetic nuclei signals based on the nuclei number, nuclei radius and intensity, and the randomness of the nuclei radius and intensity between 0 to 1. If the radius randomness is 0.2, this means the generated nuclei radius = defined nuclei radius r ± uniform (0, 1) random variable × 0.2 × r. Nuclei are generated in the following order: Synthetic nuclei with random 3D positions are generated inside the image. If the distance of this nuclei relative to the previously generated nuclei, or the distance to the image boundary is smaller than nuclei r × 1.5 we regenerate the synthetic nuclei. A 3D Gaussian function with nuclei intensity and standard deviation equal to nuclei radius r × 0.4 is added to the synthetic image. We also created a synthetic overlapping nuclei dataset to evaluate the segmentation method's ability to isolate overlapping nuclei. Subsequently, half of the nuclei were randomly created, and additional nuclei were iteratively generated and added to the dataset once its distance from the previous nuclei set was

smaller than a nuclei radius r × 1.5. After initial nuclei are distributed throughout the image, we add both white noise and user-defined noise types with user-defined noise level noise density (ND) onto the synthetic image that we generated in the previous step. White noise is white Gaussian noise with user-defined variance. User-defined noise types include Gaussian white noise and salt-and-pepper noise with ND and intensity attenuations with image gradient ND, in which pixel intensity will increase continuously from right to left with intensity ND times on the left side of the image.

1.3.4 Image synthesis, noise adding process and nuclei overlap

To evaluate the segmentation performance of the wavelet-based segmentation method, we compared it with three commonly used segmentation methods, including the point-wise method, the Otsu method and the DS algorithm. The point-wise method is the simplest method to estimate the intensity of nuclei center points i. Raw images were firstly smoothed using 9 × 9 × 3 kernel, and followed by h-maxima and h-minima transform to suppress all background signals. After combining nearby local maxima closer than 6 pixels, the remaining local maxima were assumed to be nuclei center points and intensity was calculated after applying 6 x 6 x 3 spherical kernel.

The Otsu method is the most popular threshold-based segmentation method to separate foreground and background pixels by finding the optimum threshold to satisfy minimum intra-class variance. We applied a symmetric Gaussian low-pass filter and sharpened the filter with specific radius and amount on the raw image. We then used Wiener filter or pixel-wise adaptive low-pass Wiener filter to deblur the image and remove noise. To this end, we used the multilevel thresholding Otsu method to decide threshold values.

Derivatives Sum algorithm (DS algorithm) started with applying denoising filters such as a Gaussian filter or a non-linear isotropic diffusion filter on the image, and the 2D spatial derivatives were computed to get the image Gauss gradient, Laplacian determinant, and Hessian determinant. A mask function F was calculated by combining the first and second spatial derivatives with weight parameters, and the Otsu method was applied here to obtain binary image slices. 3D nuclei segmentation was obtained by connecting the neighboring 2D masks, and we used a genetic algorithm for the parameter screening in this method57.

1.3.5 Ground truth labeling GUI

Creating embryonic 3D nuclei segmentation ground truth data is challenging and time-consuming. Here we introduced GRU3D, a 3D nuclei annotation tool that provides a convenient and efficient way to label 3D nuclei, and can also be used to evaluate nuclei segmentation results.

In GRU3D the user can directly label embryo nuclei by simple mouse clicking. There are two main types of labeling: nuclei center labeling that shows the nuclei index, and non-center nuclei labeling that is marked as a cross symbol. Non-center nuclei labeling can help to prevent double-counting. After clicking on the screen, one sub-GUI will jump up with the specific region and a range of Z-slack planes (Figure 2.1). The user can directly assign the labeling by clicking. GRU3D can also be used to check segmentation results by either comparing segmentation masks with raw images or checking single nuclei masks.

1.3.6 CPD registration method

The coherent point drift (CPD) algorithm is a probabilistic-based point set registration method widely used in the field of pattern recognition58. The main goal of point registration is to merge two or more point sets into one representative set. CPD treats the registration process as a probability density estimation problem and fits the Gaussian mixture model (GMM) of one point set to a reference point set using maximum likelihood method, and the expectation maximization algorithm is used to make the most effective use of the optimization function. We use CPD to merge embryos in the same stage into a representative embryo one by one. The final GMM centroids are the representative embryo nuclei sets.

1.3.7 3D topology features

We designed eight 3D topological features to describe the topology and geometric structural changes for embryological development studies in Zebrafish. Nuclei size is the total pixel numbers for every 3D segmented nuclei. We used four indexes to describe the spatial nuclei density distribution for the embryo, including density 3D, density 2D, neighbor 3D and neighbor 2D distance. Density 3D is the total nuclei number with distance smaller than 30 pixel around one nuclei. The problem of using density 3D is the underestimation near the outermost or innermost layer. In this instance, we define density 2D as the total nuclei number closer within 30 pixel

distance and also in the same layer (depth level). Neighbor distance is the average distance of the nuclei to the closest two nuclei which is suitable to find the outlier point in point set. The difference between neighbor 2D and neighbor only considers nuclei in the same layer.

Another important topological feature of the embryo is the thickness. We define thickness here as the vertical distance from the innermost layer of embryo. The innermost shell was divided into 360 sub-regions, and the vertical distance was calculated from nucleus to its corresponding plane. The outer epithelial monolayer of a zebrafish embryo is called the enveloping layer (EVL). Here we defined the inner EVL and outlier EVL as the layer with depth level equal to one or the maximum values. H_sorting is the index to describe the dynamic process of nuclei movement by dividing every 500 nuclei from the top of the embryo to the embryo button.

1.3.8 Cell cycle phase pattern based on nuclei shape classification

To study the spatial and temporal shape distribution of embryo nuclei, we defined some rules relative to dividing them into spherical nuclei, irregular shape nuclei, elliptical nuclei and dividing nuclei. We calculated the sphericity value using 4 × pi × area divided by squared perimeter in three planes (xy, yz and xz planes), and averaged them as the averaged sphericity value. The average aspect ratio is the mean of aspect ratio xy, yz, and xz. Aspect ratio xy is the proportional relationship between x axis range divided by y axis range. Here we defined spherical nuclei if the mean sphericity value is larger than 0.9 which means the shape of nuclei is similar to a circle with the same area. If the mean sphericity value is smaller than 0.9 and the mean aspect ratio is larger than 0.7, we identified it as an irregular shaped nuclei. A nucleus is determined as either elliptical or dividing if the mean sphericity value is smaller than 0.9 and the mean aspect ratio is also smaller than 0.7. In addition, if the radius of this nuclei is smaller than the embryo mean radius and they are paired, they are considered dividing nuclei.

1.3.9 Embryo profile projection and data visualization

To indicate a better concentration or feature distribution on a three- dimensional multi-layer embryo, and also allow for the integration of similar stage embryos, we projected features of each nuclei into one reference two-dimensional surface. We generated a hemispherical surface with averaged distributed reference point which we created in using regular placement method60. We

projected the nuclei into reference points with the closest spherical coordination system angle (θ, φ). Thereafter, we summarized the concentration of nuclei in this reference point, as well as average topological features such as nuclei size, nuclei density or thickness.

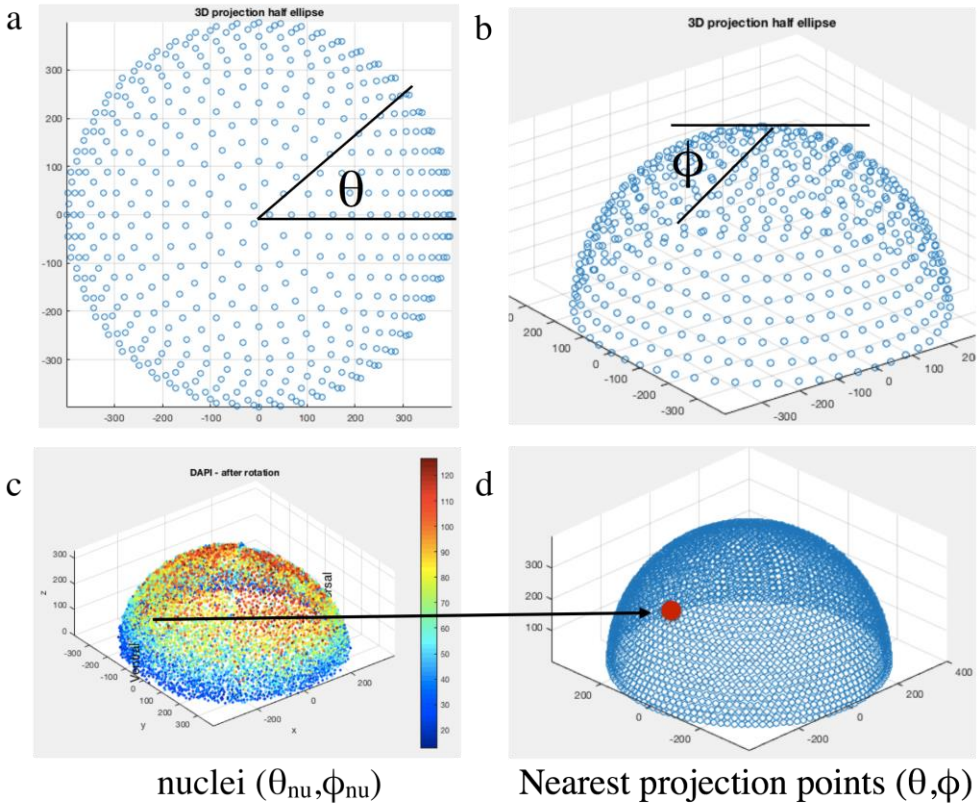


Figure 1.24. Embryo profile projection onto 2D average distributed reference surface. (a)(b) We generated a 2D averaged distributed hemispherical surface using a regular placement method to place reference points such that their distance in two orthogonal directions was always the same. Reference point numbers were assigned, and polar angle θ and azimuthal angle φ for each reference were calculated and recorded. **(c)** Polar angle and azimuthal angle for each segmented nucleus for the whole embryo was calculated after 3D rotation and coordinate system base point resetting in WaveletSEG. **(d)** We projected the nuclei onto reference points with the closest spherical coordinate system angle (θ, φ).

We designed the interactive graphical user interface to visualize segmentation and analyze the results of the embryo directly. Users can also choose a particular region such as the embryo margin region by choosing z region range, or DV axis concentration distribution. In addition, users can remove EVL nuclei directly by clicking the no EVL icon on the data visualization board. In a

scatter plot board, users can choose one coordinate system index in the first pop-up menu and one feature in second pop-up menu to analyze the spatial distribution on the embryo. Alternatively, users can also plot the relationship between two features using the scatterplot board.

1.4 Discussion

As illustrated above, we have developed a novel wavelet-based image segmentation algorithm for robust 3D nuclei segmentation that demonstrates good segmentation performance in many image datasets. Recent technological advances in biological imaging are focused on developing fully automated and large-scale imaging algorithms, and wavelet-based segmentation will be a good fit because of its reliability in the face of noise and intensity attenuation and the only necessary parameter is the wavelet scale factor which is chosen based on property object size and no preprocessing or fine-tuning steps are needed. We also presented a new way to divide overlapping nuclei according to the self-similarity between multiple wavelet coefficient maps.

We developed WaveletSEG to determine embryonic patterning information for molecular signaling, genetic regulation, and 3D positional data for nuclei at various stages of the cell cycle. This integrated image analysis platform helps define embryo coordinate systems to provide embryo axes for geometry features and quantification. After 3D nuclei segmentation for whole embryos, the post processing in WaveletSEG provides convenient tools to rotate the embryo, remove EVL layer nuclei and select subregions for additional quantification. Overall, this quantification and post-processing program is designed to support converting imaging data into information that can be used to infer biological mechanism, gain quantitative insight into development and develop quantitative data sets of pattern formation and signal regulation by morphogens.

1.5 References

1. Keller, P. J., Schmidt, A. D., Wittbrodt, J., & Stelzer, E. H. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. science, 322(5904), 1065-1069 (2008).

2. McMahon, A., Supatto, W., Fraser, S. E., & Stathopoulos, A. Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration. Science, 322(5907), 1546-1550 (2008).

3. Meilhac, S. M., Adams, R. J., Morris, S. A., Danckaert, A., Le Garrec, J. F., & Zernicka-Goetz, M. Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. Developmental biology, 331(2), 210-221 (2009).

4. Stegmaier, J., Otte, J. C., Kobitski, A., Bartschat, A., Garcia, A., Nienhaus, G. U., ... & Mikut, R. Fast segmentation of stained nuclei in terabyte-scale, time resolved 3D microscopy image stacks. PloS one, 9(2), e90036 (2014).

5. Win, K. Y., Choomchuay, S., Hamamoto, K., & Raveesunthornkiat, M. (2018). Comparative study on automated cell nuclei segmentation methods for cytology pleural effusion images. Journal of healthcare engineering, 2018.

6. Otsu, N. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66 (1979).

7. Cai, H., Yang, Z., Cao, X., Xia, W., & Xu, X. A new iterative triclass thresholding technique in image segmentation. IEEE transactions on image processing, 23(3), 1038-1046 (2014).

8. Vala, M. H. J., & Baxi, A. A review on Otsu image segmentation algorithm. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(2), pp-387 (2013).

9. Vincent, L., & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence, (6), 583-598 (1991).

10. Beucher, S. The watershed transformation applied to image segmentation. SCANNING MICROSCOPY-SUPPLEMENT-, 299-299 (1992).

11. Dufour, A., Shinin, V., Tajbakhsh, S., Guillén-Aghion, N., Olivo-Marin, J. C., & Zimmer, C. Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. IEEE Transactions on Image Processing, 14(9), 1396-1410 (2005).

12. Zhang, B., Zimmer, C., & Olivo-Marin, J. C. Tracking fluorescent cells with coupled geometric active contours. In 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821) (pp. 476-479). IEEE (2004).

13. Dzyubachyk, O., Van Cappellen, W. A., Essers, J., Niessen, W. J., & Meijering, E. Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. IEEE transactions on medical imaging, 29(3), 852-867 (2010).

14. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 2010;57:841–852 (2010).

15. Santella, A., Du, Z., Nowotschin, S., Hadjantonakis, A. K., & Bao, Z. A hybrid blob-slice model for accurate and efficient detection of fluorescence labeled nuclei in 3D. BMC bioinformatics, 11(1), 580 (2010).

16. Gudla, P. R., Nandy, K., Collins, J., Meaburn, K. J., Misteli, T., & Lockett, S. J. "A high-throughput system for segmenting nuclei using multiscale techniques. Cytometry A, 73(5):451–466 (2008).

17. Ambühl, M. E., Brepsant, C., Meister, J. J., Verkhovsky, A. B., & Sbalzarini, I. F. High‐resolution cell outline segmentation and tracking from phase‐contrast microscopy images. Journal of microscopy, 245(2), 161-170 (2012).

18. Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering, 7, 97-114 (2014).

19. Yau, C., & Wakefield, J. Quantitative image analysis of chromosome dynamics in early Drosophila embryos. In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 264-267). IEEE (2007).

20. Chinta, R., & Wasser, M. Three‐dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live Drosophila embryos. Cytometry Part A, 81(1), 52-64 (2012).

21. Rizzi, B., & Sarti, A. Region-based PDEs for cells counting and segmentation in 3D+ time images of vertebrate early embryogenesis. Journal of Biomedical Imaging, 2009, 20 (2009).

22. Xu, H., Sepúlveda, L. A., Figard, L., Sokac, A. M., & Golding, I. Combining protein and mRNA quantification to decipher transcriptional regulation. Nature methods, 12(8), 739 (2015).

23. Toyoshima, Y., Tokunaga, T., Hirose, O., Kanamori, M., Teramoto, T., Jang, M. S., ... & Iino, Y. Accurate Automatic Detection of Densely Distributed Cell Nuclei in 3D Space. PLoS computational biology, 12(6), e1004970 (2016).

24. Arneodo A, Decoster N, Roux S. A wavelet-based method for multifractal image analysis. I. Methodology and test applications on isotropic and anisotropic random rough surfaces. Eur. Phys. J. B. 15:567-600 (2000).

25. Arneodo A, Decoster N, Kestener P, Roux SG. A wavelet-based method for multifractal image analysis: from theoretical concepts to experimental applications. Advances in Imaging and Electron Physics. 126:1-92 (2003).

26. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLOS Biology. 4:e138 (2006).

27. Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. Image processing with ImageJ. Biophotonics international, 11(7), 36-42 (2004).

28. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Tinevez, J. Y. Fiji: an open-source platform for biological-image analysis. Nature methods, 9(7), 676 (2012).

29. Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., ... & Golland, P. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome biology, 7(10), R100 (2006).

30. McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A., Karhohs, K. W., ... & Wiegraebe, W. CellProfiler 3.0: Next-generation image processing for biology. PLoS biology, 16(7), e2005970 (2018).

31. Skinner, S. O., Sepúlveda, L. A., Xu, H., & Golding, I. Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. Nature protocols, 8(6), 1100 (2013).

32. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., ... & Zimmer, C. FISH-quant: automatic counting of transcripts in 3D FISH images. Nature methods, 10(4), 277 (2013).

33. Lou, X., Kang, M., Xenopoulos, P., Munoz-Descalzo, S., & Hadjantonakis, A. K. A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data. Stem cell reports, 2(3), 382-397 (2014).

34. Ianzini, F., & A. Mackey, M. Development of the large scale digital cell analysis system. Radiation protection dosimetry, 99(1-4), 289-293 (2002).

35. Davis, P. J., Kosmacek, E. A., Sun, Y., Ianzini, F., & Mackey, M. A. The Large‐Scale Digital Cell Analysis System: an open system for nonperturbing live cell imaging. Journal of microscopy, 228(3), 296-308 (2007).

36. Stegmaier, J., Amat, F., Lemon, W. C., McDole, K., Wan, Y., Teodoro, G., ... & Keller, P. J. Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. Developmental cell, 36(2), 225-240 (2016).

37. Sánchez, C. I., García, M., Mayo, A., López, M. I., & Hornero, R. (2009). Retinal image analysis based on mixture models to detect hard exudates. Medical Image Analysis, 13(4), 650-658 (2009).

38. Zinski, J., Bu, Y., Wang, X., Dou, W., Umulis, D., & Mullins, M. C. Systems biology derived source-sink mechanism of BMP gradient formation. Elife, 6, e22199 (2017).

39. Rajasekaran, B., Uriu, K., Valentin, G., Tinevez, J. Y., & Oates, A. C. Object segmentation and ground truth in 3D embryonic imaging. PloS one, 11(6), e0150853 (2016).

40. Ulman, V., Maška, M., Magnusson, K. E., Ronneberger, O., Haubold, C., Harder, N., ... & Smal, I. An objective comparison of cell-tracking algorithms. Nature methods, 14(12), 1141 (2017).

41. Solnica-Krezel, L., Stemple, D. L., & Driever, W. Transparent things: cell fates and cell movements during early embryogenesis of zebrafish. Bioessays, 17(11), 931-939 (1995).

42. Li, L., Wang, X., Mullins, M. C., & Umulis, D. M. Evaluation of BMP-mediated patterning in zebrafish embryos using a growing finite difference embryo model. bioRxiv, 585471 (2019).

43. Huang, Y., & Umulis, D. M. Scale invariance of BMP signaling gradients in zebrafish. Scientific reports, 9(1), 5440 (2019).

44. Lepage, S. E., & Bruce, A. E. Zebrafish epiboly: mechanics and mechanisms. International Journal of Developmental Biology, 54(8-9), 1213-1228 (2010).

45. Morita, H., Grigolon, S., Bock, M., Krens, S. G., Salbreux, G., & Heisenberg, C. P. The physical basis of coordinated tissue spreading in zebrafish gastrulation. Developmental cell, 40(4), 354-366 (2017).

46. Concha, M. L., & Adams, R. J. (1998). Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. Development, 125(6), 983-994 (1998).

47. Bensch, R., Song, S., Ronneberger, O., & Driever, W. Non-directional radial intercalation dominates deep cell behavior during zebrafish epiboly. Biology open, 2(8), 845-854 (2013).

48. Warga, R. M., & Kimmel, C. B. Cell movements during epiboly and gastrulation in zebrafish. Development, 108(4), 569-580 (1990).

49. Bruce, A. E. (2016). Zebrafish epiboly: spreading thin over the yolk. Developmental Dynamics, 245(3), 244-258 (2016).

50. Mendieta-Serrano, M. A., Schnabel, D., Lomelí, H., & Salas-Vidal, E. Cell proliferation patterns in early zebrafish development. The Anatomical Record, 296(5), 759-773 (2013).

51. Godard, B. G., & Heisenberg, C. P. Cell division and tissue mechanics. Current opinion in cell biology, 60, 114-120 (2019).

52. Gul-Mohammed, J., Arganda-Carreras, I., Andrey, P., Galy, V., & Boudier, T. A generic classification-based method for segmentation of nuclei in 3D images of early embryos. BMC bioinformatics, 15(1), 9 (2014).

53. Chan, C. J., Heisenberg, C. P., & Hiiragi, T. Coordination of morphogenesis and cell-fate specification in development. Current Biology, 27(18), R1024-R1035 (2017).

54. Lio, P. Wavelets in bioinformatics and computational biology: state of art and perspectives. Bioinformatics, 19(1), 2-9 (2003).

55. Du, P., Kibbe, W. A., & Lin, S. M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics, 22(17), 2059-2065 (2006).

56. Antoine, J. P., Carrette, P., Murenzi, R., & Piette, B. Image analysis with two-dimensional continuous wavelet transform. Signal processing, 31(3), 241-272 (1993).

57. Bäck, T., & Schwefel, H. P. An overview of evolutionary algorithms for parameter optimization. Evolutionary computation, 1(1), 1-23 (1993).

58. Myronenko, A., & Song, X. Point set registration: Coherent point drift. IEEE transactions on pattern analysis and machine intelligence, 32(12), 2262-2275 (2010).

59. Kohonen, T. Exploration of very large databases by self-organizing maps. In Proceedings of International Conference on Neural Networks (ICNN'97) (Vol. 1, pp. PL1-PL6). IEEE (1997).

60. Deserno, M. How to generate equidistributed points on the surface of a sphere. If Polymerforshung (Ed.), 99 (2004).

CHAPTER 2. GROUND TRUTH ANNOTATOR AND 3D DATASET GENERATOR FOR VALIDATION OF NUCLEI SEGMENTATION PROGRAMS

This material covered in this chapter has been published in the 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2019).

Wu, T. C., Wang, X., & Umulis, D. M. Ground truth annotator and 3D dataset generator for validation of nuclei segmentation programs. 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2019.

2.1 Introduction

Several nuclei segmentation methods exist to track cells in fixed or living tissues [1-4]. Most nuclei segmentation algorithms need image preprocessing steps before the segmentation, and the settings in preprocessing are highly dependent on the microscopy imaging environment [5]. Many microscopy image analysis software packages have also been established, such as ImageJ [6], Fiji [7], and CellProfiler- a widely used software [8], that provides preprocessing,object identification, and cell counting modules. Additional related software includes Spatzcells [9], FISH-quant [10], and MINS [11] are MATLAB executable codes that can generate segmentation cell masks and spot recognition for fluorescent 2D or 3D images.

Ground truth dataset can be used to evaluate results from segmentation methods and software, and modify and improve the segmentation algorithm. However, the lack of ground truth expert delineations of fluorescent microscopy imaging data is a serious challenge to evaluate and to improve the training and validation of microscopy image analysis techniques. Some manual nuclei labeling imaging datasets have been built [12], but few convenient ground truth annotation tools have been developed [13-15].

We developed a benchmarking dataset generator, a segmentation validation tool, and a 3D ground truth annotation tool as a resource to the imaging and annotation community. It is challenging and time-consuming to label the 3D nuclei centers in a ground truth image. To overcome this challenge, we developed a convenient interactive 3D ground truth annotation interface (GRU3D) that provides 3D nuclei labeling by a mouse clicking interface. When the user clicks specific location

inside the software window, another interface will pop up and display every z planes for 3D nuclei blob, and the user can directly label the nuclei center position. In SYNIMAGE, users define the synthetic image's 3D size and number of image and synthetic nuclei signals, and add both white noise and user-defined noise types with user-defined noise density (ND) onto the synthetic image. Using SEG3D, users can also easily validate segmentation results by utilizing different segmentation methods and determine the True Positive (TP), False Negative (FN) and False Positive (FP) nuclei directly.

2.2 System overview

2.2.1 GRU3D

Annotating a nuclei segmentation ground truth dataset is challenging and time-consuming. Here, we introduce GRU3D, a 3D nuclei annotation tool that provides a convenient and efficient way to label 3D nuclei that can also be used to evaluate nuclei segmentation results (Figure 2.1). In GRU3D, the user can directly label nuclei through simple mouse clicking. There are two main types of labeling: nuclei center labeling, which shows the nuclei index, and noncenter nuclei labeling, which is marked as a cross symbol. Noncenter nuclei labeling can help to prevent double counting. After clicking on the screen, one sub graphical user interface (GUI) will pop up in a specific region with a range of z-stack planes. The user can directly assign the labeling by clicking. GRU3D can also be used to check segmentation results by either comparing segmentation masks with raw images or checking single nuclei masks.

The left window of the GRU3D shows a raw image with a side bar for viewing different z-slices. The right window shows both raw image and segmentation labels and provides a very convenient way to label 3D nuclei. When the user clicks inside the right-side window, a GUI will pop up and display every z-slice in this position. Users can easily select the nuclei center and start labeling. Figure 2.1b are four 3D ground truth 400 × 400 × 24 voxel region nuclear dye 4'-6-diamidino-2-phenylindole (DAPI) staining raw images from 4.7 hours post fertilisation (hpf) zebrafish embryo (the first three images), and raw images of 400 × 400 × 24 voxel region of 4.7 hpf zebrafish embryo in pSMAD protein datasets.

2.2.2 SYNIMAGE

To provide a dataset of synthetic images in validating segmentation methods, we developed a convenient interface (SYNIMAGE) to generate synthetic image datasets (Figure 2.2). Users define the synthetic image's 3D size and number of image and synthetic nuclei signals based on the nuclei number, nuclei radius, and intensity, and the randomness or variability of nuclei radii and intensity between 0 to 1.

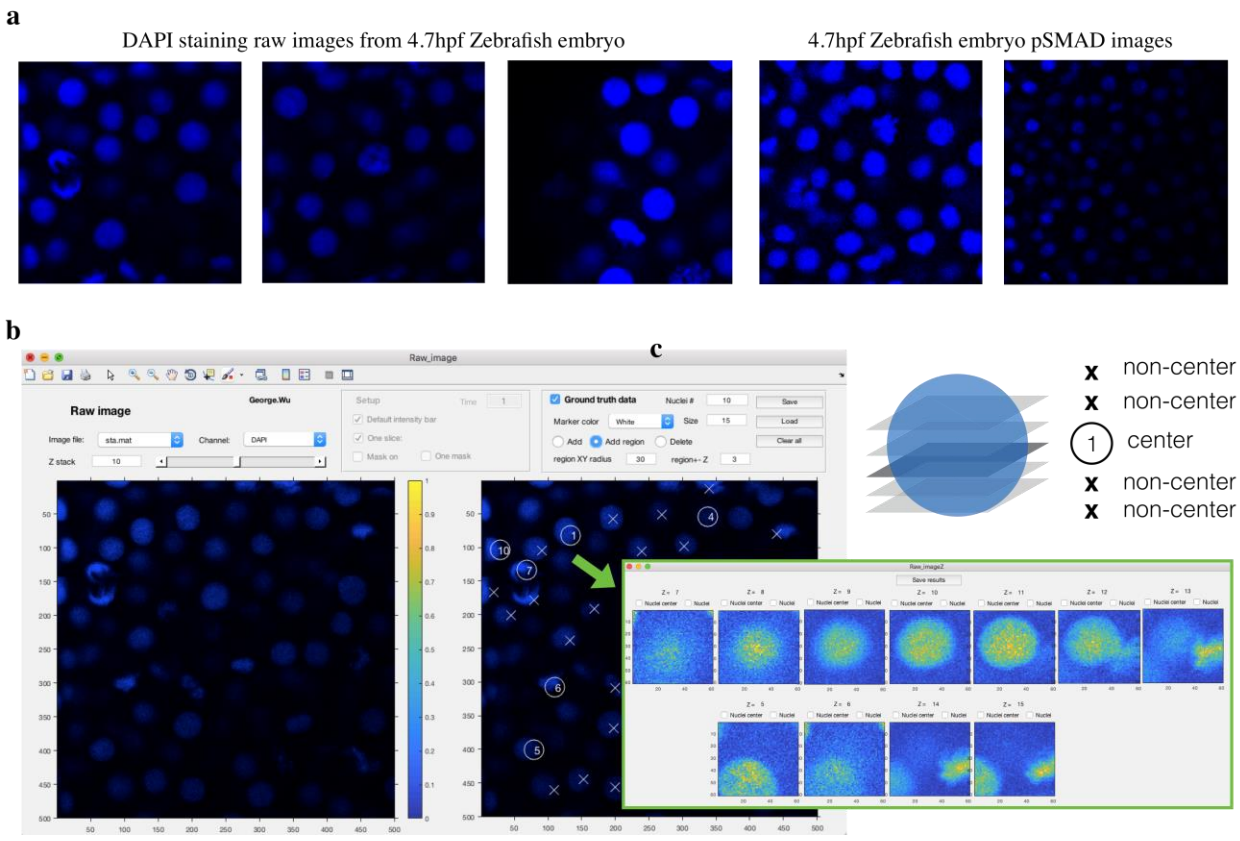
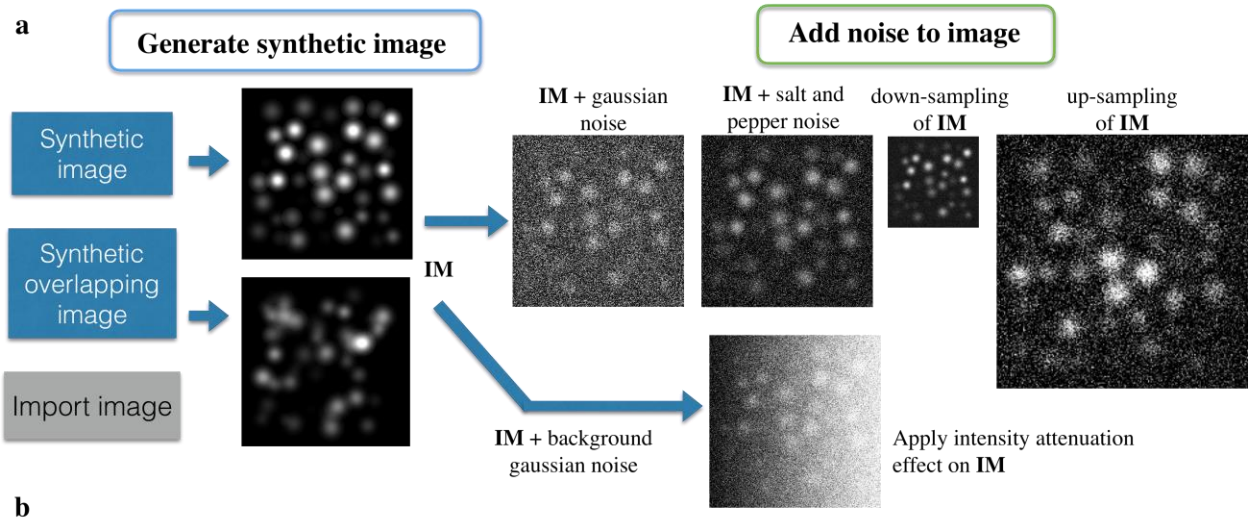


Figure 2.1. GRU3D: the 3D ground truth annotation GUI and segmentation result viewer. **(a)** Five 3D ground truth of 400 × 400 × 24 voxel region nuclear DAPI staining raw images from 4.7hpf Zebrafish embryo (first three images), and raw images 400 × 400 × 24 voxel region of 4.7hpf Zebrafish embryo in pSMAD datasets. **(b)** To build the ground truth dataset using our own images to validate the segmentation method, we introduced GRU3D, a 3D nuclei annotation tool which provides a convenient and efficient way to label 3D nuclei, and can also be used to evaluate nuclei segmentation results. The left window of the GRU3D shows the raw image with side bar to view different z-slice. **(c)** The right window shows both raw image and segmentation labels, and provides very convenient way to label 3D nuclei. When user clicking inside the right side window, pop-up GUI will show up and display every z-slice in this position. User can very easy to select the nuclei center and do labeling.

If the radius randomness is 0.2, this means the generated nuclei radius = defined nuclei radius r ± uniform (0, 1) random variable × 0.2 × r. Nuclei are generated in the following order: Synthetic nuclei with random 3D positions are generated inside the image. If the distance of this nuclei relative to the previously generated nuclei, or the distance to the image boundary is smaller than nuclei r × 1.5 we regenerate the synthetic nuclei. A 3D Gaussian function with nuclei intensity and standard deviation equal to nuclei radius r × 0.4 is added to the synthetic image. We also created a synthetic overlapping nuclei dataset to evaluate the segmentation method's ability to isolate overlapping nuclei. Subsequently, half of the nuclei were randomly created, and additional nuclei were iteratively generated and added to the dataset once its distance from the previous nuclei set was smaller than a nuclei radius r × 1.5.

After initial nuclei are distributed throughout the image, we add both white noise and user-defined noise types with user-defined noise level noise density (ND) onto the synthetic image that we generated in the previous step. White noise is white Gaussian noise with user-defined variance. User-defined noise types include Gaussian white noise and salt-and-pepper noise with ND and intensity attenuations with image gradient ND, in which pixel intensity will increase continuously from right to left with intensity ND times on the left side of the image. In this study, to validate different segmentation methods, we generated 20 synthetic images for each test with size 400 × 400 × 40 with 50 nuclei. The nuclei radius was equal to 15 with randomness = 0.2, and the intensity of the nuclei center was 250 with randomness = 0.8. We applied four kinds of noise, Gaussian white, salt-and-pepper, intensity attenuation effect, and overlapping nuclei with three levels (low, middle, and high).



b

Synthetic image property	Synthetic image size: 400x400x40, Image number = 20 Nuclei# = 50, Nuclei radius = 30, Nuclei center I = 250 Radius randomness = 0.2, Intensity randomness = 0.8
Gaussian noise	Noisy level = 0.01(low), = 0.1(mid), = 0.5 (high)
Salt and pepper noise	Noisy level = 0.01(low), = 0.1(mid), = 0.5 (high)
Optical aberration	Noisy level = 2(low), = 4(mid), = 7 (high)
Nuclei overlap	Nuclei number = 50(low), = 70(mid), = 100 (high)

c

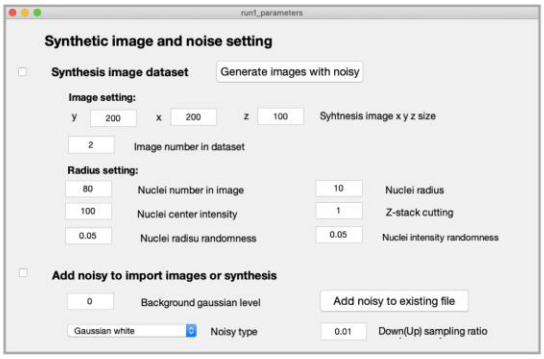


Figure 2.2. The workflow of the synthetic data generator and process to add noise in SYNIMAGE. The synthetic image generator in WaveletSEG can generate synthetic images or synthetic overlapping images by providing image number, image size, synthesis nuclei number, nuclei radius and intensity and their randomness. For overlapping images, we iteratively create pairing nuclei with distances between two nuclei are smaller than nuclei radius ×1.5. After initial synthetic image dataset are created, Gaussian noise, salt and pepper noise with different noise level can be added into images, or apply down-sampling or up-sampling to those images. When applying intensity attenuation effect to initial images, firstly we added background noise, and intensity attenuation with image gradient ND which pixel intensity will increase continuously from right to left with intensity ND times in the left side of the image.

2.2.3 SEG3D

We also provide a convenient segmentation method comparison GUI (SEG3D) to validate segmentation results or compare them with other segmentation methods (Figure 2.3). The right side of this GUI is the main data visualization window, where segmented nuclei positions are displayed as a 3D point cloud. The color of the 3D point cloud is fluorescent intensity of a specific microscopy channel. The horizontal plane is a raw image z-slice that enables the user to validate

and compare 3D segmented nuclei with original raw image z-slices. This data visualization region also enables 3D rotation zoom. The color of the 3D point cloud is fluorescent intensity of a specific microscopy channel.

On the left side of SEG3D is the main control panel with two dropdown menus (Method 1 menu and Method 2 menu). You can select different segmentation methods from two dropdown menus to compare their segmentation results. Additionally, if users generate synthetic image datasets using SYNIMAGE (Figure 2.2) or create ground truth datasets using GRU3D (Figure 2.1), they can select "ground truth data" from the dropdown menu to validate the segmentation results. For example, when users select "ground truth data" in the Method 1 menu and select "Otsu's method" in the Method 2 menu, TP, FP, and FN are calculated and displayed in the data visualization window. When comparing two methods or ground truth data, segmented nuclei will be marked as blue (TP), red (FN), or green (FP). TP rate and precision will automatically be saved in data files (Figure 2.4).

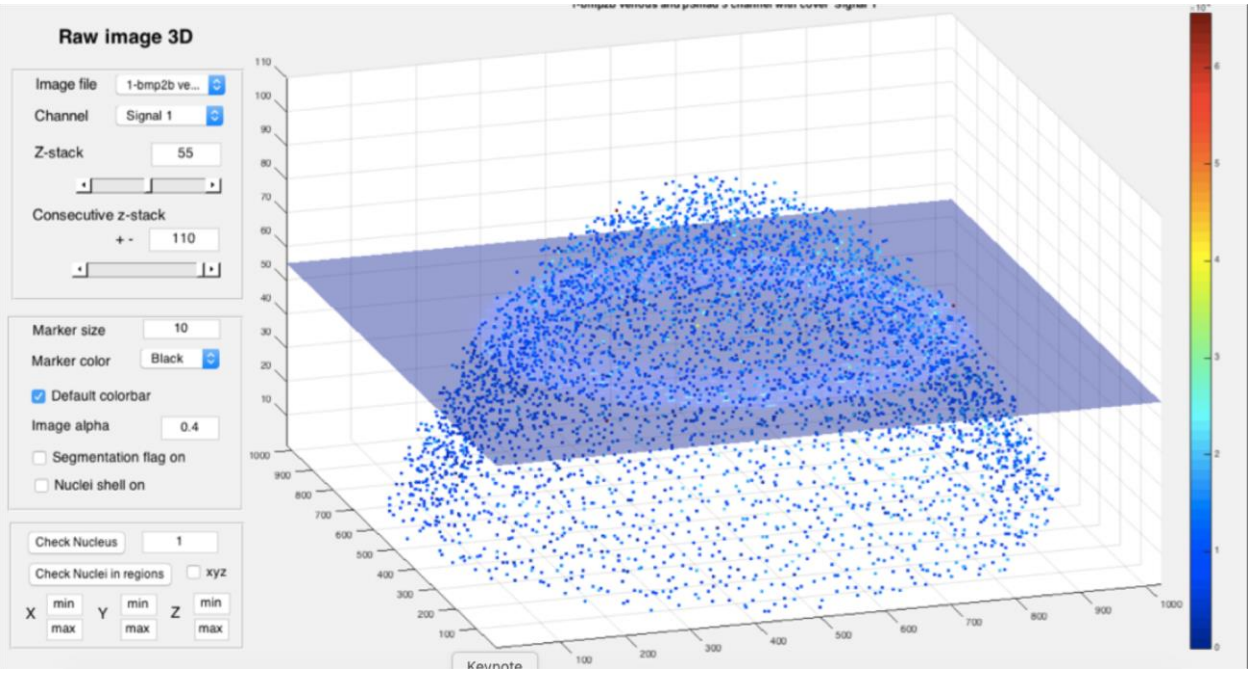


Figure 2.3. SEG3D: 3D segmentation method comparison GUI. SEG3D is the 3D segmentation data visualization GUI to check and validate the segmentation results. Blue points inside the main data visualization window (Right) are segmented nuclei positions are displayed as 3D point cloud, and horizontal plane is the raw image z-slice enable user to validate and compare 3D segmented nuclei with original raw image z-slice. Left side of this sub-GUI is the control panel to display regional nuclei or change into different images or channels

Further information including the 3D surface rendering of each segmented nuclei can be conducted by selecting a nuclei spot number or the xyz axis range after 3D reconstruction in the control panel in SEG3D. The red solid point inside the 3D rendering nuclei is the nuclei center position inside the nuclei marked with number and xyz positions. Color is used only to distinguish between different objects (Figure 2.5).

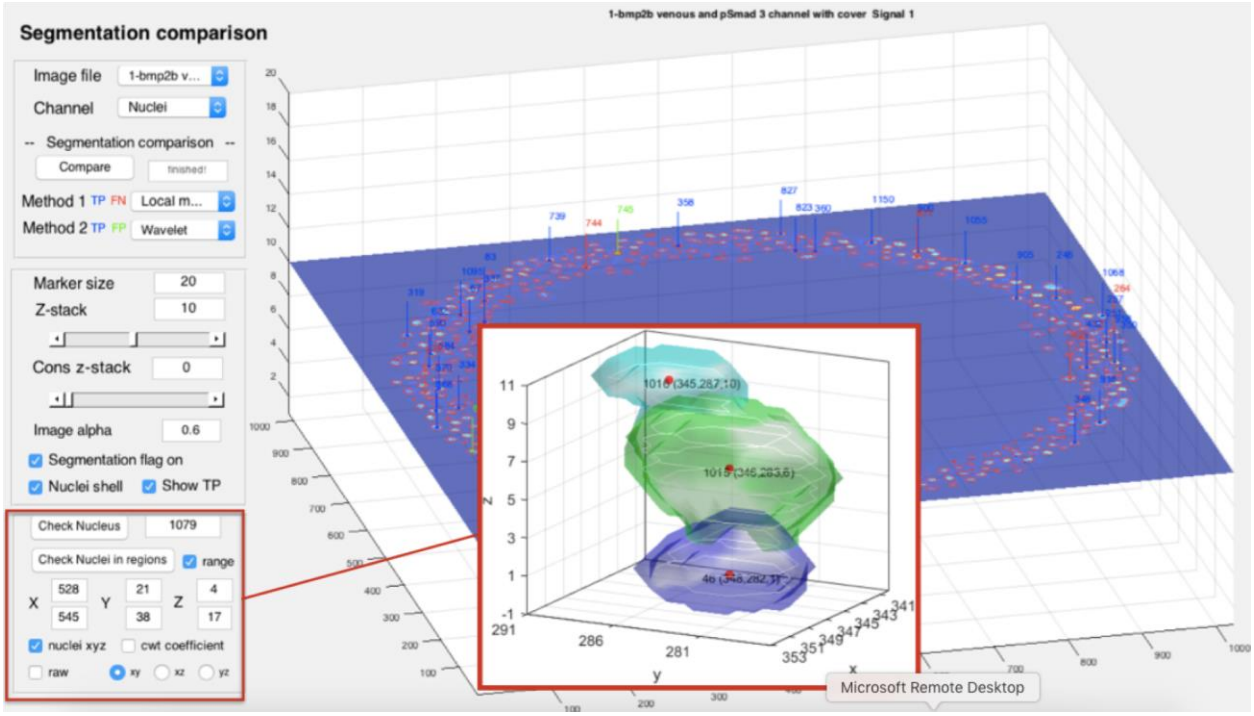


Figure 2.4. SEG3D: 3D segmentation method comparison GUI with 3D surface rendering. SEG3D is designed to validate the segmentation results or compare segmentation results using different segmentation methods. Right side of this sub-GUI is the main data visualization window and segmented nuclei positions are displayed as 3D point cloud. The horizontal plane is raw image z-slice enable user to validate and compare 3D segmented nuclei with original raw image z-slice. When comparing with two methods or with ground truth data, segmented nuclei will be marked as blue (TP), red (FN), or green color (FP). 3D surface rendering of segmented nuclei can display by choosing nuclei spot number or select the xyz axis range after 3D reconstruction.

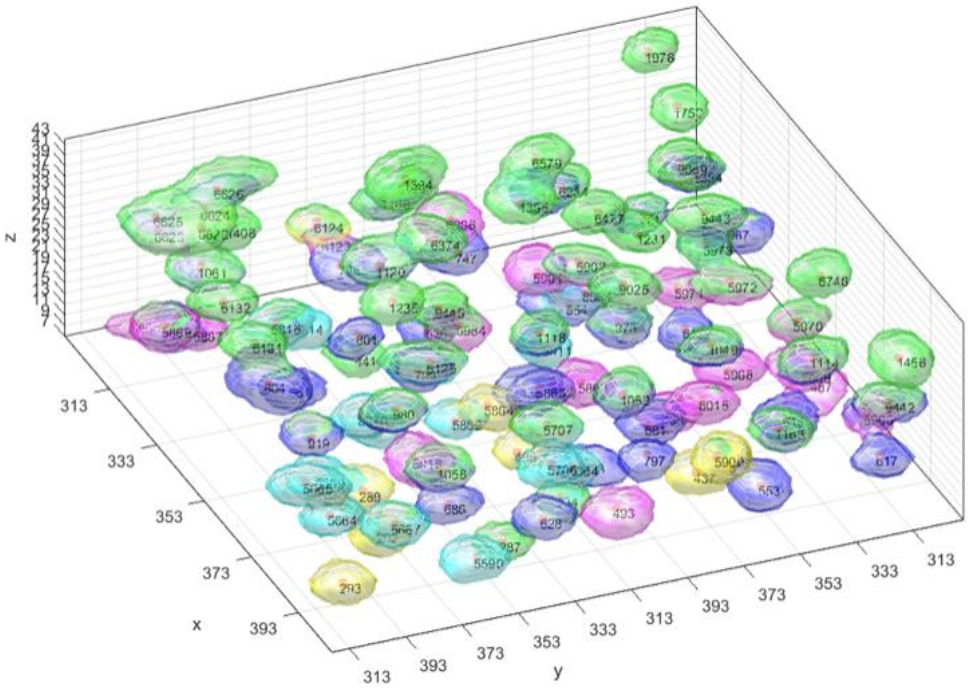


Figure 2.5. 3D reconstruction of surface rendering in SEG3D. 3D surface rendering of each segmented nuclei can also be conducted by selecting a nuclei spot number or the xyz axis range after 3D reconstruction in the control panel in Raw_image3D. The red solid point inside the 3D rendering nuclei is the nuclei center position inside the nuclei marked with number and xyz positions. Color is used only to distinguish between different objects.

GRU3D, SYNIMAGE and SEG3D were developed in MATLAB R2016 environment and are compatible across multiple MATLAB versions after R2012, and can support many different platforms, such as PCs, Linux and Mac OS X. Take GRU3D as an example-the FIG-file contains the layout of the GUI, and all executed functions and scripts are located in an M-file. To run this GUI code, both M-file and FIG-file must located in the same directory and the program begins by executing GRU3D at the MATLAB workspace prompt. All codes requires the image processing toolbox and statistics toolbox.

2.3 Results

In this section, we use SYNIMAGE and GRU3D to create synthetic image datasets and ground truth image datasets and demonstrate the nuclei segmentation performance comparison between a point-wise method and Otsu's method in SEG3D. We added three kinds of noise, Gaussian white noise, salt-and-pepper noise, and intensity attenuation effect with three levels (low-, mid-, and

high-level) to datasets in SYNIMAGE to validate the segmentation ability for noisy images. We also generated synthetic nuclei overlapping datasets to test the performance in dividing overlapping nuclei.

Three evaluation criteria were applied to compare the accuracy and robustness of the segmentation methods using ground truth dataset [16]. If a representative point was the nearest neighbor of a point in the ground truth dataset and visa-versa, the object was regarded as a TP. If only the former condition was met, the ground truth was regarded as a FN. If only the latter condition was met, the object was regarded as a FP. The TP rate is equal to TP number divided by ground truth number, and precision is TP number divided by the sum of TP number and FP number. TP rate represents the ratio of successful nuclei segmentation within all ground truth nuclei data, and precision is the correct segmentation ratio in the nuclei segmentation result obtained using a specific method. The FN rate is the FN number divided by ground truth number.

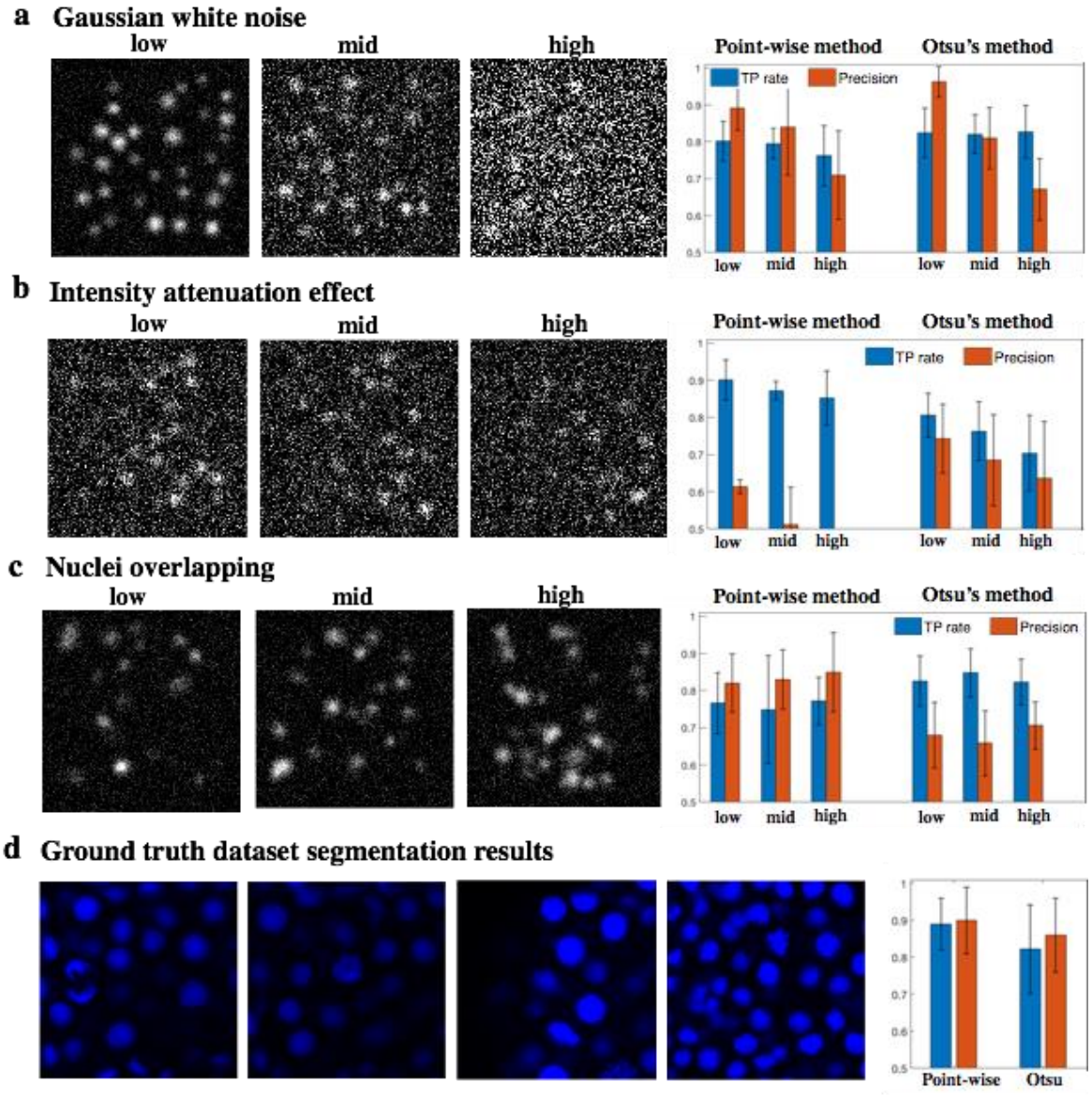


Figure 2.6. Performance criteria of noise in synthetic images. (a-c) Segmentation performance criteria of three levels of (low, mid, and high) noise in synthetic images created by SYNIMAGE and ground truth image dataset from GRU3D (d).

The point-wise method [17] is the simplest method to estimate the intensity of nuclei center points i. Raw images are first smoothed using 9 × 9 × 3 kernel and followed by h-maxima and h-minima transform to suppress all background signals. After combining nearby local maxima that were closer than 6 pixels to each other, the remaining local maxima were assumed as nuclei center points, and intensity was calculated after applying 6 × 6 × 3 spherical kernel. Otsu's method is the most

popular threshold-based segmentation method to separate foreground and background pixels by finding the optimum threshold to satisfy minimum intra-class variance. We applied a symmetric Gaussian low-pass filter and sharpened the filter with a specific radius and amount on the raw image. We then used Wiener filter or pixel-wise adaptive low-pass Wiener filter to deblur the image and remove noise. For this, we used multilevel thresholding Otsu's method to decide threshold values [17].

In datasets with Gaussian white noise (Figure 2.6a), Otsu's method provides better TP rate (above 0.8) than point-wise method but the precision rate decreased rapidly for Otsu's method which shows the high ratio of negative to positive results. In images with intensity attenuation effect (Figure 2.6b), the TP rate decreased from 0.8 to 0.72 and precision decreased from 0.77 to 0.63 when applying Otsu's method. Conversely, the point-wise method results show higher TP rate and very low precision compared to Otsu's method. This showed that the intensity attenuation effect will greatly affect the point-wise method. We also examined the ability of segmentation methods to divide nuclei in overlapping images (Figure 2.6c). Otsu's method results in a TP rate higher than the point-wise method and Otsu's method results in precision that is lower than point-wise method. This means that only part of the nuclei (TP rate: 0.85 ~ 0.7) are identified from overlapping nuclei.

2.4 Conclusion

The proposed 3D nuclei annotation tool provides a publicly available, convenient and efficient way to label 3D nuclei in ground truth datasets. This tool is expected to help researchers to improve their algorithms and also validate and compare between alternative nuclei segmentation methods as we show for a point-wise method and Otsu's method. Synthetic datasets are also important to evaluate segmentation methods with known nuclei positions and properties. Using GRU3D and SYNIMAGE, we can provide high quality, reliable and sufficient ground truth datasets and synthetic datasets to improve segmentation accuracy and robustness. The SEG3D we developed can also help researches to check segmented nuclei positions in the software screen, and judge segmentation comparison results with ground truth dataset or between segmentation methods.

2.5 References

1. Keller, P. J., Schmidt, A. D., Wittbrodt, J., & Stelzer, E. H. "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy," Science, 322(5904), pp. 1065-1069, 2008.

2. McMahon, A., Supatto, W., Fraser, S. E., & Stathopoulos, A. "Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration," Science, 322(5907), pp. 1546-1550, 2008.

3. Meilhac, S. M., Adams, R. J., Morris, S. A., Danckaert, A., Le Garrec, J. F., & Zernicka-Goetz, M. "Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst," Developmental biology, 331(2), pp. 210-221, 2009.

4. Stegmaier, J., Otte, J. C., Kobitski, A., Bartschat, A., Garcia, A., Nienhaus, G. U., ...& Mikut, R. "Fast segmentation of stained nuclei in terabyte-scale, time resolved 3D microscopy image stacks," PloS one, 9(2), e90036, 2014

5. Win, K. Y., Choomchuay, S., Hamamoto, K., & Raveesunthornkiat, M. "Comparative study on automated cell nuclei segmentation methods for cytology pleural effusion images," Journal of healthcare engineering, 2018.

6. Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. "Image processing with ImageJ," Biophotonics international, 11(7), pp. 36-42, 2004.

7. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Tinevez, J. Y. "Fiji: an open-source platform for biological-image analysis," Nature methods, 2012, 9(7), pp. 676, 2012.

8. Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., ... & Golland, P. "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," Genome biology, 7(10), R100, 2006.

9. Skinner, S. O., Sepúlveda, L. A., Xu, H., & Golding, I. "Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization," Nature protocols, 2013, 8(6), pp. 1100, 2013.

10. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., ... & Zimmer, C. "FISH-quant: automatic counting of transcripts in 3D FISH images," Nature methods, 10(4), pp. 277, 2013.

11. Lou, X., Kang, M., Xenopoulos, P., Munoz-Descalzo, S., & Hadjantonakis, A. K. "A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data," Stem cell reports, 2014, 2(3), pp. 382-397, 2014.

12. Gelasca, E. D., Byun, J., Obara, B., & Manjunath, B. S. “Evaluation and benchmark for biological image segmentation,” In 2008 15th IEEE International Conference on Image Processing, pp. 1816-1819. IEEE., October 2008.

13. Rajasekaran, B., Uriu, K., Valentin, G., Tinevez, J. Y., & Oates, A. C. “Object segmentation and ground truth in 3D embryonic imaging,” PloS one, 2016, 11(6), e0150853, 2016.

14. Hamarneh, G., Jassi, P., & Tang, L. “Simulation of ground-truth validation data via physically-and statistically-based warps,” In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, pp. 459-467, 2008.

15. Byun, J., Verardo, M. R., Sumengen, B., Lewis, G. P., Manjunath, B. S., & Fisher, S. K.. “Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images,” Mol Vis, 12(105-07), pp. 949-60, 2006.

16. Ulman, V., Maška, M., Magnusson, K. E., Ronneberger, O., Haubold, C., Harder, N., ... & Smal, I. “An objective comparison of cell-tracking algorithms,” Nature methods, pp. 14(12), 2017.

17. Zinski, J., Bu, Y., Wang, X., Dou, W., Umulis, D., & Mullins, M. C. “Systems biology derived source-sink mechanism of BMP gradient formation,” Elife, 6, e22199, 201

CHAPTER 3. PAVEMENT CELL SHAPE QUANTIFICATION WITH LOBEFINDER

This material covered in this chapter has been published in the Plant Physiology Journal [45].

Wu, T. C., Belteton, S. A., Pack, J., Szymanski, D. B., & Umulis, D. M. LobeFinder: a convex hull-based method for quantitative boundary analyses of lobed plant cells. Plant physiology, 171(4), 2331-2342 (2016).

3.1 Introduction

The size, shape, and angle of leaves are important adaptive traits in natural populations and key determinants of yield in agronomic settings [1]. Therefore it is important to understand the cellular events that collectively, at the levels of the tissues and organs, lead to the formation of durable, lightweight, and appropriately sized leaf blades for efficient light capture [2]. In Arabidopsis, the growth properties of the epidermis may have particular importance in terms of organ size control [3], and the growth behaviors of the sectors of the epidermis and individual cells can correlate with organ-level growth behaviors [4][5]. In dicots, the basic cellular unit of the epidermis is the jig-saw-puzzle piece shaped pavement cell, the division and expansion of which drive leaf expansion [6].

The biomechanics of pavement cell shape change are complicated [7-10]. Turgor pressure is the driving force for cell expansion. However, the magnitude and directions of cell wall tension forces are difficult to predict because of the presence of adjacent neighboring cells both in the plane of the epidermis and in the underlying mesophyll tissue [10]. The shape of the cell itself may also influence the stress patterns in the wall, and regions of high cell curvature are predicted to have increased wall stress compared to other domains of the cell [11]. The growth trajectory or strain response of the cell is also strongly influenced by heterogeneity in the cell wall, and a current challenge is to understand how differences in cell wall thickness and local cellulose-dependent cell wall anisotropy might contribute to polarized growth in this cell type [7, 10].

The developmental control of lobe initiation in cotyledons and leaves is also poorly understood. In one early model, lobe initiation was proposed to direct organ shape, with cell elongation and

lobe initiation occurring independently in populations of cells to influence organ growth in length and width, respectively [12,13]. Other studies, which relied on cell shape measurements from populations of unsynchronized cells, detected correlations between cell size and lobe number, implying a continuous process of cell expansion and lobe initiation [14,15]. Neither of these models appears to be correct based on several recent papers that employ either long-term time lapse imaging of pavement cell morphogenesis [4,16] or cell population analyses that analyzed cells from developmentally staged leaves over time intervals spanning days [17] or weeks [18]. The clear outcome for these studies is that the frequency of lobe initiation clearly depends on the developmental stage and location on the leaf. However, in many instances lobe initiation is unpredictable. For example, a given cell's anticlinal (perpendicular to the leaf surface) walls are in contact with several neighboring cells. New lobes can form along either one or several of these cell boundaries, and the factors that define the probability of forming a new lobe at a particular location are not known. Lobe initiation is therefore episodic, and morphogenesis appears to include both anisotropic growth during lobe initiation and lobe expansion, as well as extended phases of symmetrical cell expansion in which the cell size increases but the overall geometry of the cell remains essentially unchanged [4].

One major limitation in the field is the lack of a robust and objective method to identify new lobes. The discussion above on the cellular and developmental control of lobe formation is largely based on subjective evaluation of pavement cell segments as being either lobed or unlobed. This has generated confusion and variability in the literature with regard to detecting phenotypes and comparing the severity of phenotypes among different mutants. In some instances, the endpoints of a midline skeleton of individual pavement cells have been used to estimate lobe number [18,19]; however, this method is not very accurate and appears to underestimate lobe number. As an alternative, dimensionless shape descriptors like circularity (4π x cell area/perimeter^2), a ratio that approaches 1 for more circular cells and gets smaller as cells become more lobed, are used to test for differences among cells in the complexity of their cell shape [19-23]. The major weakness of this approach is that it does not directly reflect lobe number, and there are many equally plausible explanations in which reductions in either lobe initiation or lobe expansion could lead to similar differences in cell shape complexity. In this technical advance, we describe a highly useful convex hull-based Matlab program termed LobeFinder that operates on cell perimeter coordinates

extracted from images of pavement cells and returns an array of useful cell shape data including a value of lobe number and a map of their positions. Based on median scores of manually identified features from a diverse population of pavement cells, LobeFinder predictions outperformed the alternative method of binary image skeletonization and subjective human scoring. The development, validation, limitations, and uses of LobeFinder are described below.

3.2 Materials and Methods

3.2.1 Annotation and use of the LobeFinder program

To run LobeFinder, start an instance of Matlab ® on the workstation (PC, MAC, Linux) and change the working directory to the install location of LobeFinder. The script and all functions that make up LobeFinder are located in one Matlab ® m-file: 'LobeFinder_GUI.m'. To run LobeFinder, first create a directory to which all region-of-interest (ROI) of cell perimeter coordinates obtained by manual segmentation are saved. Start the LobeFinder Graphical User Interface (GUI) by typing 'LobeFinder_GUI' at the MATLAB workspace prompt and <Enter>. This will open an instance of LobeFinder GUI in a separate window (Figure 3.7). To import files, click on the "Open Folder" button to select the folder that contains the ROI files. At this point one can select the checkboxes for the types of data output files to be generated (CSV, Figures, Matlab File) as well as the resolution of the images from which the ROIs were extracted. Once the folder and options have been selected, click on the "Run" button to start the ROI processing. The total number of ROI files being processed will be shown in the image number box on the GUI. To view the results from the LobeFinder processing, select the image number from the image number box. This will populate the GUI with the measured parameters as well as an image of the refined hull, the cell boundary, identified lobe points, and DTRH plot. Moving the cursor over the perimeter of the cell will allow its corresponding position on the DTRH plot to be seen. Depending on the output options selected, a new folder in the directory of LobeFinder will be created ([Output] NameOfInputFolder) with up to 3 folders (CSV, FIG_cell, FIG_dtrh). CSV folder contains one Matlab mat-file 'Lobe_result.mat' which contains all results and geometric scalar properties for each ROI in the directory, a 'CellDescriptors.csv' file containing all single value measurements such as area, perimeter, etc for all ROIs, and individual 'DTRH_[nameOfROI] file containing xy-values for DTRH plots. FIG_cell and FIG_dtrh folders will contain images of cells and DTRH plots as

displayed in the LobeFinder GUI ROILobeFinder is available for download at the Dryad Digital Repository: http://doi.org/10.5061/dryad.cs78t.

3.2.2 Plant material and growth conditions.

Arabidopsis seeds were grown on 1/2X Murashige and Skoog medium with 1 percent sucrose and 0.8 percent Bacto agar under constant illumination at 220C. Seeds were treated with a 6 hr light pulse, cold-treated for 3 days, then placed in the growth chamber. Germination was checked 36 hrs after plating, and only seedlings with a barely visible radicle were used for further analysis.

3.2.3 Time-lapse imaging of lobe initiation

For time point imaging, cell outlines were detected using a tubulin:GFP marker for datasets 2 and 3 as previously described [4]. For dataset 1, from 38 to 55 HAG, 10 cells were analyzed. For dataset 2, from 48 to 120 HAG 12 cells were analyzed. For dataset 3, from 72 to 120 HAG 12 cell were analyzed. For dataset 1, the PIN7:GFP [24] plasma membrane marker was used. The seedlings were mounted in water using a petroleum jelly gasket to form a chambered microscope slide. After initial imaging, the slides were returned to the growth chamber until the next imaging session. Samples were imaged using a Bio-Rad 2100 laser scanning confocal microscope mounted on a Nikon eclipse E800 stand. Images were obtained with a 60X 1.2 NA water objective. Samples were excited with a 488 nm laser and fluorescence signal was collected using a nm long pass dichroic, and a 500-550 nm band-pass emission filter. Selected planes from confocal image stacks were converted to maximum intensity projects and were traced with the polygon selection tool using in FIJI 4.0 [25]. The coordinates from the ROIs from the manually segmented cells were used as the input for LobeFinder.

3.3 Results

Currently, quantification methods of lobe formation are often focused on the localization of specific factors related to cellular shape change such as the distribution of actin filaments, the presence of anticlinal microtubule bundles, and qualitative descriptions of cell shape [13,15]. However, because there is no known marker protein for lobe initiation, and because lobe counting results vary greatly between labs and among individuals (see below), there is a strong need for a standardized computational approach to measure the number and location of pavement cell lobes.

A number of commercial and open-source software applications are available to quantify the geometry of cell shapes. These methods can be broadly separated into two categories: quantification of descriptive scalar properties such as circularity, roughness, perimeter, area, etc. that describe the shape by descriptive parameters [26, 27]; and image-segmentation approaches that we broadly define here as methods that reduce the pixel information in the raw image into segments or a reduced set of data points that have greater biological meaning [28], such as converting an image of a cell into segmented regions for nucleus, cytoplasm, golgi, ER, etc. automatically. These approaches offer a reduction in the size of the data and a transformation of pixel intensity data into classifications that directly informs the biology of the problem. The Medial Axis Transform (MAT) [18] has been used to quantify pavement cell geometry. The MAT uses the midline points of cells to quantify cell shape differences by tracking the percent change in angles between linear segments of the branches along the central axes or skeleton. A similar method for tracking the midline of a cell is available as the FIJI plugin AnalyzeSkeleton method [25, 29].

In the analysis of pavement cell shape, the most widely used computational method to identify lobes is based on the AnalyzeSkeleton algorithm that detects the midline of irregularly shaped objects, categorizing the pixel properties of the surrounding area and choosing the best path to detect areas of image continuity. Based upon the number of neighboring pixels, some pixel points are ignored or favored over others and a skeletonized representation of the central axes of the cell's shape and structure is formed. In this method, individual cells are manually extracted from a confocal image of a field of pavement cells (Figure 3.1). A midline skeleton is calculated from the binary image and the skeleton end-points are extended to the cell perimeter, depending on the magnitude of the protrusion, to map positions of predicted lobes. As shown in Figure 3.1A and Figure 3.1B, the skeletonize method is not very accurate, and only about half of the lobes that would be identified by a trained scientist are accurately identified with this method. Therefore, this method is useful in determining generalized lobing events, usually well after a new lobe has formed, but is unable to detect slight variations in wall geometry that signify recent lobing events.

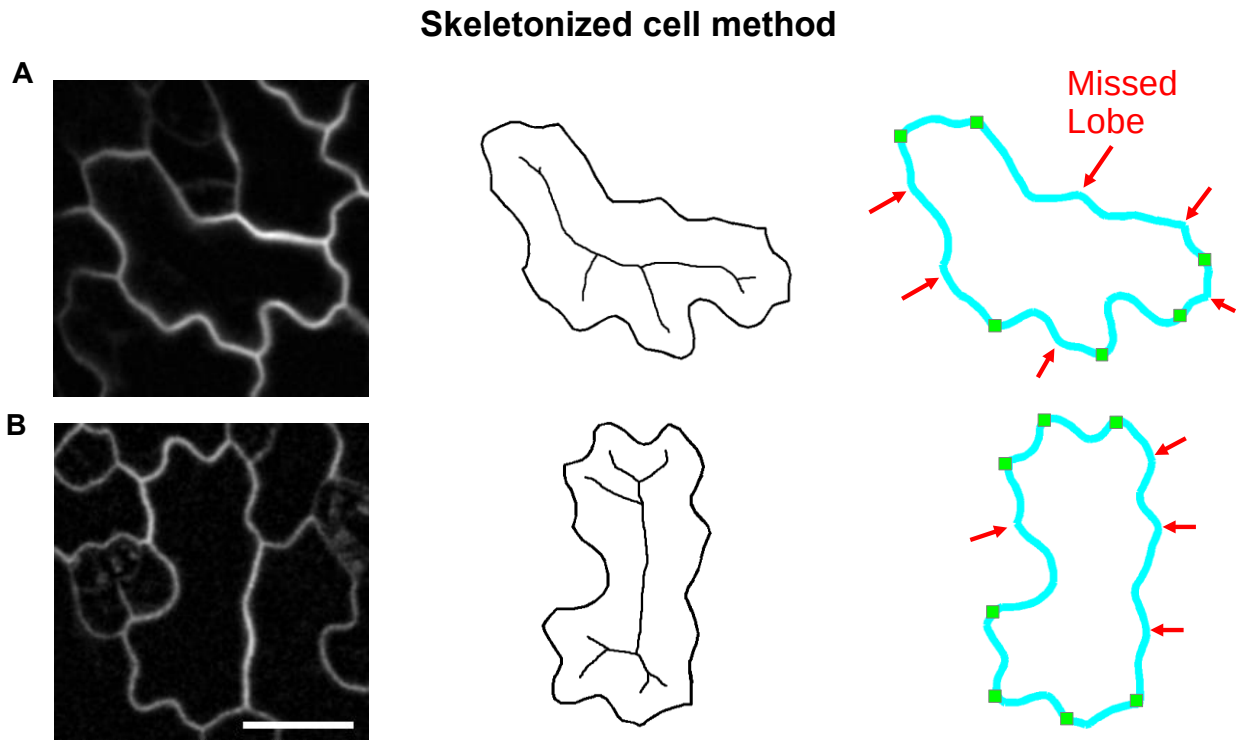


Figure 3.1. The AnalyzeSkeleton processing technique of lobe identification method does not accurately identify pavement cell lobes. (A) and (B), Left, representative confocal images of an early stage cotyledon pavement cells. Middle, calculated midline skeletons of the corresponding pavement cells. Right, summary of the accuracy of the AnalyzeSkeleton method. Green squares, correctly identified lobe points based on the extensions of the skeleton endpoints; red arrows, missed lobe points compared to voting results. Scale bar, 20μm.

3.3.1 Outline of LobeFinderualization

To overcome the limitations of the previous methods in identifying the position and number of lobes in pavement cells, we developed a new cell geometry analysis approach named LobeFinder. LobeFinder operates on user-supplied cell boundary coordinates that are extracted from high-resolution confocal images of pavement cells of various sizes and shapes. The algorithm is based on a multi-step process starting with a convex hull of the cell boundary and a sequence of processing events to robustly identify lobes [30]. First, cell boundaries are segmented from the original image. For our analysis of cell boundary variation and lobe detection, existing segmentation methods utilizing gradient vector field SNAKEs or a related approach [31, 32] were not sufficiently accurate, frequently merging cells or creating additional cells from an irregularly-shaped lobe. The recently published semi-automated method for pavement cell segmentation

termed CellECT improves the efficiency of 3D pavement cell segmentation and includes user input to reduce errors [33]. In the future, CellECT could be modified to output a single set of splined coordinates that accurately depict the boundary of the anticlinal cell wall. We anticipate that as cell segmentation methods improve, LobeFinder will be integrated into an image processing workflow to enable high-throughput cell phenotyping. However, at the present time manual segmentation is the only reliable method to extract cell coordinates, and this can be easily achieved using the polygon selection tool that is available in ImageJ. The ImageJ segmentation tool is advantageous because it allows the user to adjust the position of the cell boundary points and add or delete points as needed.

For this study, confocal images were at a resolution of 3.95 or 2.55 pixels/μm. After testing a range of sampling densities along the cell perimeter, we found that sampling frequencies of 0.5 to 1.5 points/μm were sufficient to yield accurate results for cell shape analyses using LobeFinder because lobe detection was consistent in this range. Sampling frequencies of 1 point every 2 μm or less led to obvious mismatches between the cell shapes in the raw image and the segmented cells. We recommend sampling cell perimeters at 1 point/μm and selecting the spline function within ImageJ to smooth the manual tracing and provide a high density of interpolated points.

Following extraction of the cell perimeter by segmentation, the center of mass of each cell is calculated and moved to the origin. The overall cell size is normalized and scaled by a constant factor to calculate lobe numbers (Figure 3.2A). Following analysis the outputs are rescaled back to microns for the outputs reported in the graphical user interface (GUI). This allows the use of the same relative metrics and LobeFinder settings to determine if a lobe is present for cells of different ages and sizes. The normalization step also allows raw images at multiple different resolutions to be processed in LobeFinder. To remove artifacts introduced by the uneven sampling of perimeter points during manual cell segmentation and reorientation, the cell perimeter data is approximated by a cubic spline interpolation.

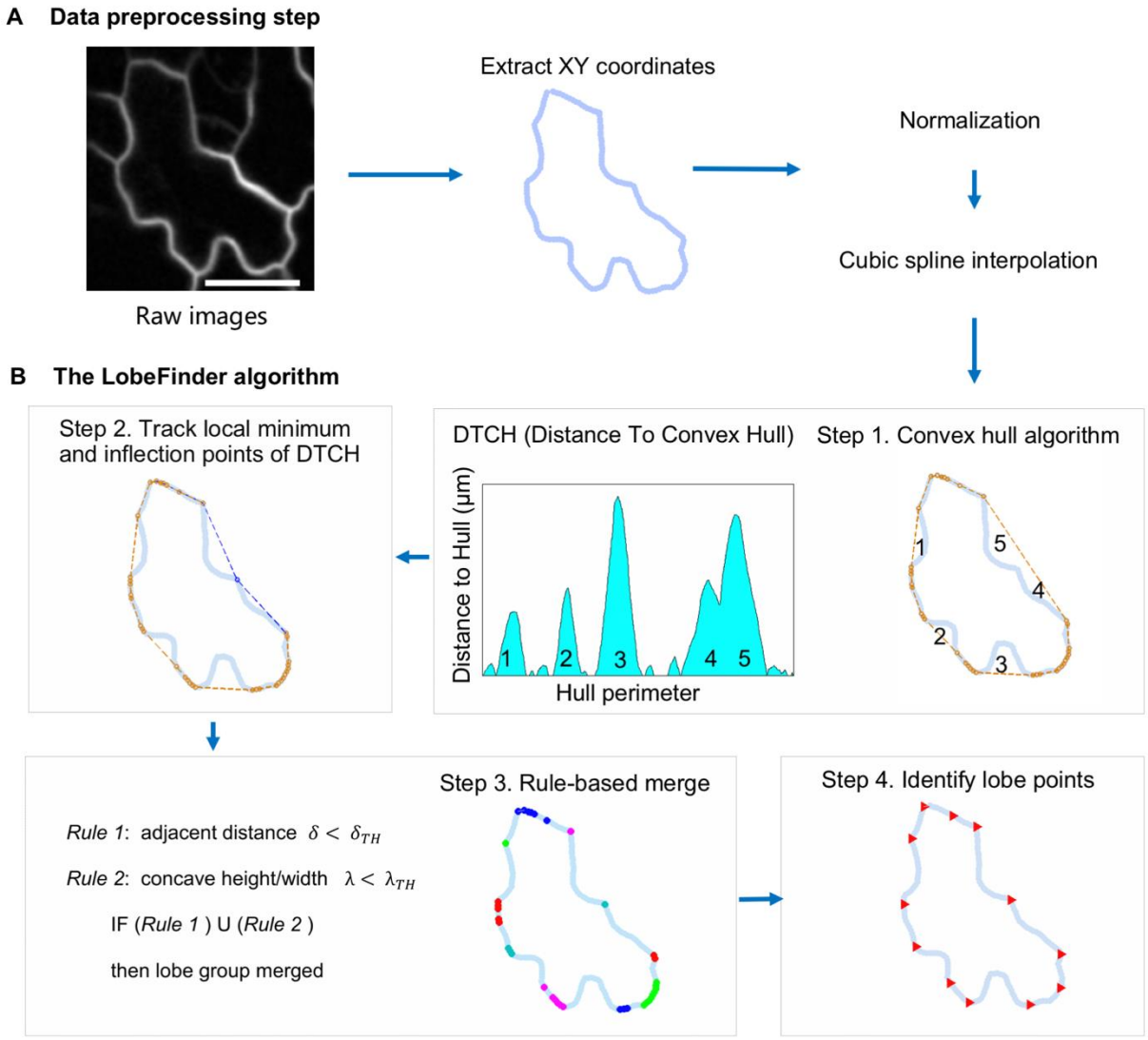


Figure 3.2. Overview of the LobeFinder logic and workflow. (A) Cell perimeter positions are manually segmented from raw images, scaled, and resampled. (B) A convex hull, defined as the as the minimal polygon which encloses the entire given cell perimeter is computed (step 1), then the perimeter is scanned for missed lobe points (the extrema between segments 4 and 5) using the PeakFinder algorithm within MatLab (step 2). The optimized values for thresholds (δTH and λTH) (TH: Threshold) for rule based lobe geometry and spacing (step 3) are used to identify putative lobe points (step 3), and last, groups of lobe points are merged and the final set of predicted lobe positions are extracted (step 4). Scale bar, 20μm.

The output of the preprocessing steps is a cell perimeter that is scaled, aligned with the center of mass, smoothed and resampled (cubic spline interpolation), and ready for further analysis. To acquire the minimal polygon that surrounds the entire set of coordinates that define the cell boundary (Figure 3.2B), we employ the MATLAB function ‘convhull’ that returns the coordinates

of the convex polygon (hull) that contains all the coordinate points of the cell set (MATLAB, 2013). The convex hull provides two important features for further analysis: first it provides information for the minimum convex set that encompasses the entire cell, and second it provides a convenient coordinate system onto which the cell boundary properties are easily mapped (Figure 3.2B, middle). Both of these outputs will serve to subsequently identify key points and structures.

Pavement cells do not typically produce an outline where all of the extrema at the lobe tips are located precisely on the hull. For example, in Figure 3.2B, the convex hull produces a line that does not separate regions 4 and 5 by the lobe that is located between the regions since the lobe does not land on the hull itself. To adjust the hull, the distance between the cell and the hull is calculated and plotted on an axis of position vs. distance (Figure 3.2B). Using the orthogonal Distance To the Convex Hull (DTCH) to the cell perimeter, the local minima are retrieved, and the convex hull is then refined to capture the interior local minima points. To determine whether there are interior lobe points between adjacent points on the convex hull, we use the program PeakFinder [34] to determine both local and absolute extrema between hull(i) and hull(j) (points on the cell periphery coincident with hull). Peakfinder identifies the location of the missed lobes and the hull used to encapsulate the cell is modified to contact the lobe point (Figure 3.2B, step 2). These additional processing steps capture the majority of interior lobes that would otherwise be missed since they do not lie on the hull surrounding each cell. The resulting hull is termed the "refined hull" because it no longer conforms to the strict definition of a convex hull. The Distance to the Refined Hull (DTRH) plots contain highly useful information on the local patterns of growth. Therefore, the cell and its refined hull are re-scaled back to their real dimensions, and the DTRH coordinates are available to be exported within the LobeFinder program. In rare instances, there are relatively large pavement cells in which a cell lobe is bulbous. In these instances the path of the cell perimeter doubles back on itself on one axis creating multiple solutions for the DTRH plot. In this sub-region of the cell, the peakfinder routine uses only the smallest distance value, and this can lead to erroneous hull refinements and lobe calls. This morphology is rare in our dataset, but a bulbous morphology is the default state in the crenulated boundary of many monocot leaf epidermal cells. For these species, LobeFinder would likely perform well in analyzing early events associated with lobe initiation, but would likely fail to accurately count the lobes of fully expanded cells.

Following adjustment of the convex hull, the goal of the algorithm is to identify which of the points on the cell perimeter correspond to the positions of the protrusions. Additionally, not every point on the hull corresponds to a physical lobe on the pavement cell, and to some extent the identification of a lobe on the cell is subjective in nature with different individuals identifying different lobe positions and numbers. One design goal of the algorithm is to mimic the expert observer's approach to identify the geometric features, albeit by an objective computer algorithm. This goal informed the design of the geometric parameters for lobe geometry and spacing that were developed to optimize lobe identification. For each data point in the set of convex hull points (Figure 3.2B, Step 3), the distance between neighboring points is calculated. This distance between hull points determines if the algorithm should consider adjacent lobe points as part of the same lobe. To cull points on the hull and leave only those that are identified as the center of a lobe, two parameters (δ and λ) for the initial identification of lobes are used: the scaled spacing distance between lobe points (δ), and a ratio of the height (distance between hull and cell boundary) to the width (distance of hull segment) between prospective lobe points (λ). The distance between a lobe point and the convex hull is zero, however there must be a region between lobes where the distance is non-zero and above some threshold value. This module of the program calibrates LobeFinder to reduce the number of misleading or incorrect lobe points on the convex hull. This calibration is effective in most cases. However, because the parameters are tuned to be sensitive for small deviations in boundary shape, cells with relatively simple shapes with extended domains of the cell boundary that are close to the the threshold values for δ and λ are most likely to have false positives.

3.3.2 LobeFinder optimization and evaluation

The principal method for identification of lobes and lobing segments in pavement cells relies on observer-based inspection and identification of lobing events. An important research goal is to standardize pavement cell phenotyping and to create an objective computational method that can accurately quantify cellular geometry and be applied to time-lapse data and large ensembles of images to efficiently calculate population statistics. It is therefore necessary to ensure the algorithm produces consistent and accurate observations.

As an initial test of the sensitivity of the output to variability in an individual's choice of boundary points for cell segmentation, three pavement cells of varying sizes and shapes were manually segmented three times and analyzed using LobeFinder. For each of these cells the area, perimeter, and circularity values for the technical replicates were either identical or differed by a fraction of a percent. For the technical replicates the LobeFinder outputs for lobe number were more variable, with the coefficients of variation for lobe number varying between .06 and .10. This level of variabilty in the measurement of lobe number was much less than that observed when multiple individuals used subjective criteria to score an identical cell. For example, in our test population of pavement cells (Figure 3.3), the coefficient of variation for lobe number ranged from 0.03 to 0.21, with 10 of the 15 cells having a coefficient of variation greater than 0.1. In the LobeFinder program, variability in lobe number most often occurred along relatively straight cell perimeter segments with one tracing including a very small feature that was absent in another. The cause for this is discussed further below, but this result makes clear the importance of accurate sampling along the cell perimeter.

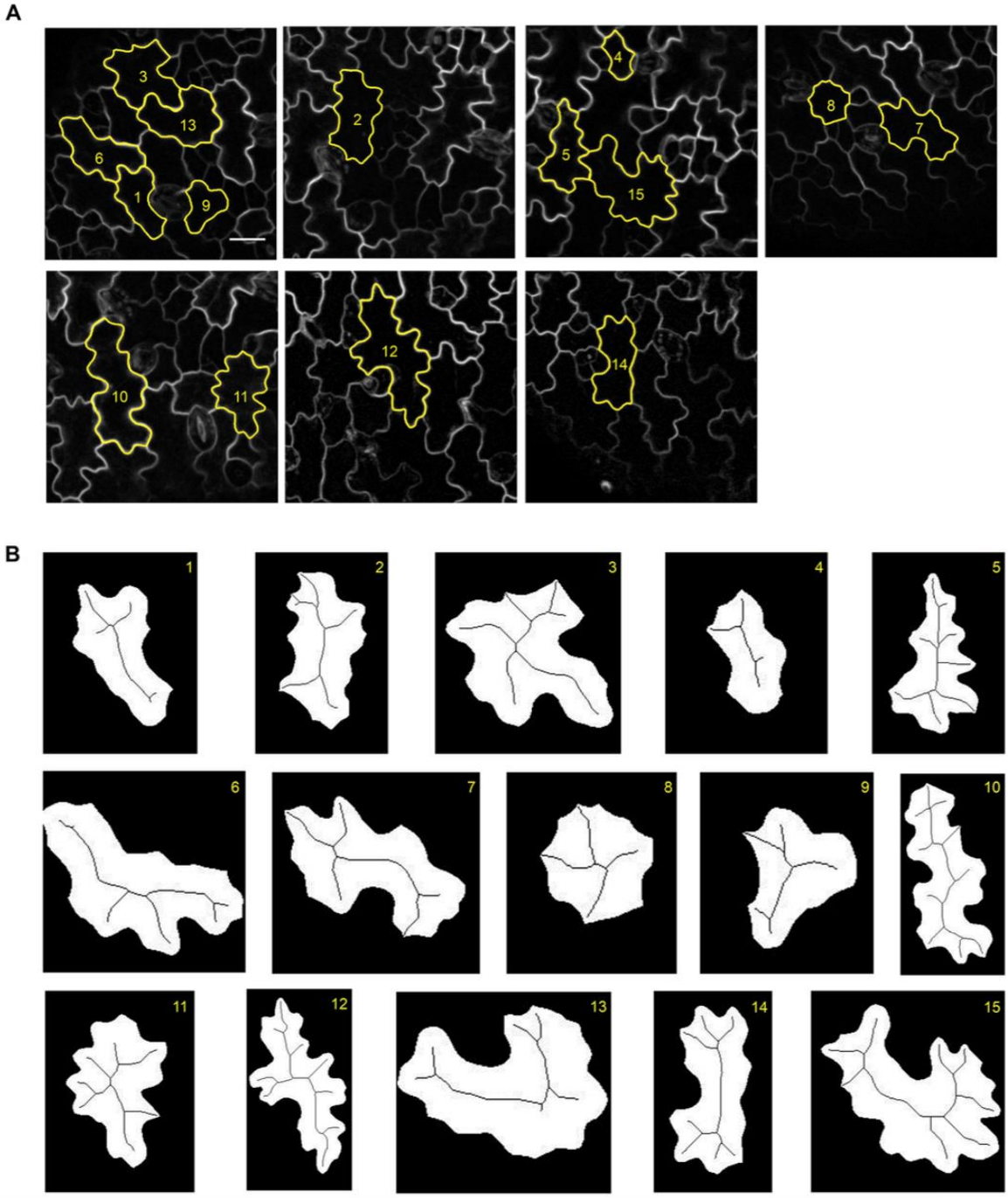


Figure 3.3. Examples of raw confocal images of pavement cells and skeletonization results. (A) Cell perimeter positions are manually segmented from raw images, scaled, and resampled. (B) A convex hull, defined as the as the minimal polygon which encloses the entire given cell perimeter is computed (step 1), then the perimeter is scanned for missed lobe points (the extrema between segments 4 and 5) using the PeakFinder algorithm within MatLab (step 2). The optimized values for thresholds (δTH and λTH) (TH: Threshold) for rule based lobe geometry and spacing (step 3) are used to identify putative lobe points (step 3), and last, groups of lobe points are merged and the final set of predicted lobe positions are extracted (step 4). Scale bar, 20μm.

Fifteen randomly selected pavement cells (Figure 3.3) from a time-lapsed dataset were used to more thoroughly compare the accuracy of LobeFinder compared to exisiting methods. The cells had sizes that ranged from to 1588 μm2, and circularity values that ranged from 0.32 to 0.81. We evaluated how well LobeFinder outputs of lobe number and lobe position would agree with the scores generated by researchers with experience in the analysis of pavement cell shape. For each of the images (examples shown in Figure 3.4), six experienced pavement cell scientists visually inspected 8.5" X 11" printouts of each cell and identified lobe locations for each raw image. These data were used to determine the accuracy of lobe point position detection and to calculate the number of lobes present for each cell (Figure 3.4C and Figure 3.4D). A few of the cells used in the calibration of LobeFinder are shown in Figure 3.4A. The complete set of images (Figure 3.3), and a summary of the cell and convex hull properties (Table 3.1) are provided in the supporting online materials. A summary of the LobeFinder and voter results are shown in Figure 3.4E and Figure 3.4F. We next benchmarked LobeFinder and the existing skeletonize method against the images manually curated by members of the two labs. The subjective nature of the manual scoring of lobe number is evident in the plots of lobe number (Figure 3.4E and Figure 3.4F), with many cells having four or more features that were ambiguous. The median lobe number from the manually curated data was therefore used as a standard for comparison.

Table 3.1. Morphological properties of pavement cells measured using LobeFinder.

Morphological properties

Dataset	Time	Area (μm^2)	Convex Area (μm^2)	Perimeter (μm)	Convex perimeter (μm)	Mean radius (μm)
38-55 HAG	38 H	634 ± 288	747 ± 359	115 ± 34.3	106 ± 27.6	14.5 ± 3.52
	55 H	1.05e3 ± 405	1.35e3 ± 579	172 ± 51.9	143 ± 32.8	19.1 ± 3.98
48-120 HAG	48 H	576 ± 341	688 ± 430	106 ± 40.9	99.0 ± 35.8	13.7 ± 4.61
	120H	1.93e3 ± 961	2.48e3 ± 1.41e3	224 ± 81.5	191 ± 59.7	25.4 ± 7.31
72-120 HAG	72H	1.70e3 ± 1.45e3	2.32e3 ± 2.17e3	224 ± 135	176 ± 81.3	23.1 ± 9.63
	120H	2.70e3 ± 2.09e3	3.71e3 ± 3.20e3	292 ± 167	226 ± 94.8	29.7 ± 11.6

Ratio of morphological properties

Dataset		Area ratio	Convex Area ratio	Perimeter ratio	Convex perimeter ratio	Mean radius ratio
38-55 HAG	55H/38H	1.73 ± 0.22	1.87 ± 0.20	1.50 ± 0.05	1.37 ± 0.08	1.33 ± 0.08

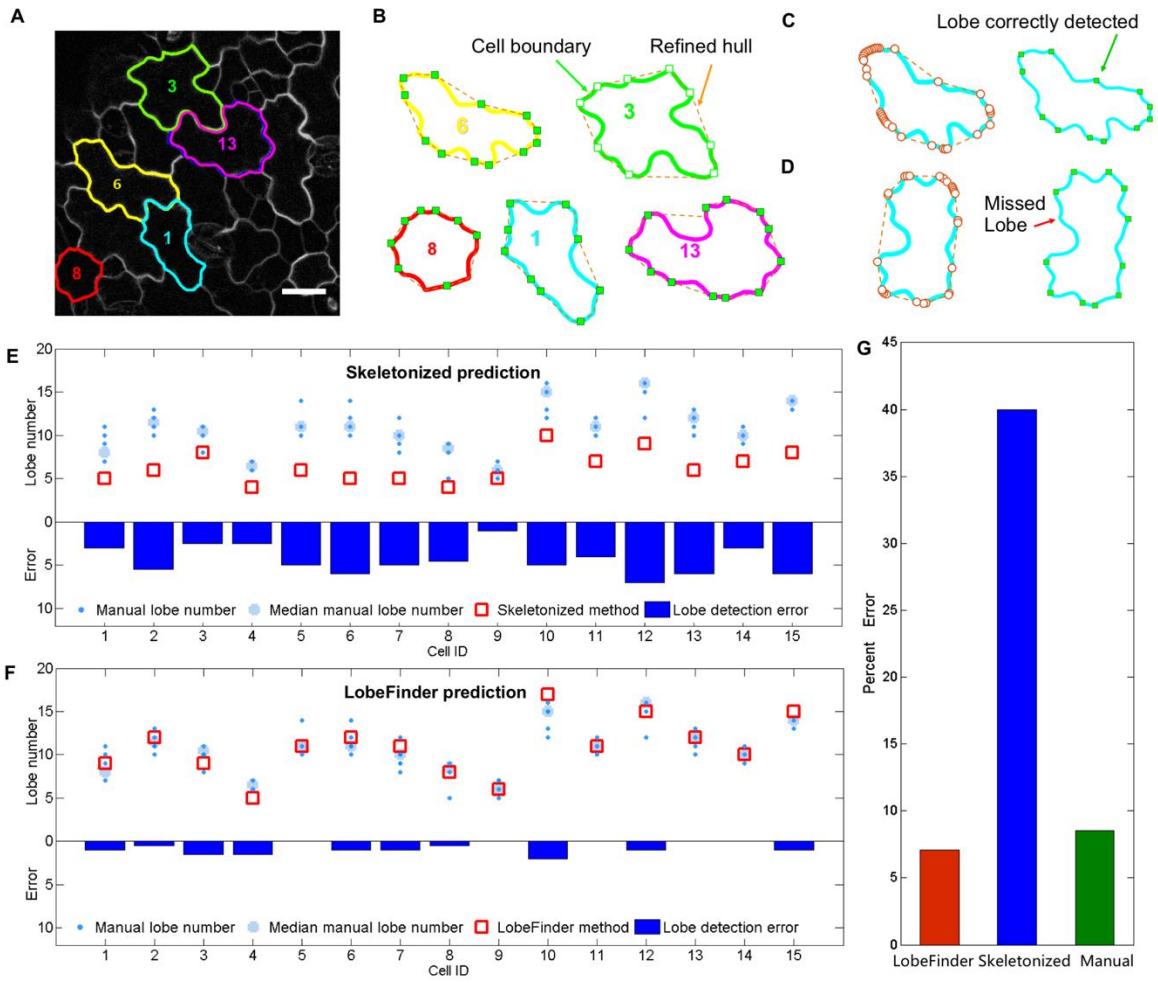


Figure 3.4. Evaluation of the LobeFinder accuracy using a calibration dataset and parameter optimization. (A) Example of a raw image containing five cotyledon pavement cells in the calibration dataset. (B) Outlines of extracted cells showing the cell boundary and the unrefined convex hull. (C, D) Example output of LobeFinder for two cells in which the correctly identified (green squares) and missed (red arrow) lobes are marked. (E) Comparison of the skeletonize method with manually curated results. The light blue circles are the median values from manual lobe identification results for each cell with individual independent values in small dark blue dots, and red boxes are lobe numbers predicted by skeletonize method. The dark blue bars plot are the absolute value of the difference between the lobe number count from the skeletonize method and the median value from the manual results. (F) Comparison of the LobeFinder method with manually curated results. The symbols and bars are as described in (E), but here the red boxes are the lobe numbers predicted by LobeFinder. The dark blue bars are the absolute value differences between the lobe number count from LobeFinder and the median value from manual results. (G) Comparison of the percent errors of the LobeFinder, skeletonize, and manual scoring methods that were calculated using the median lobe number as the correct value for each cell. Scale bar, 20μm.

Overall, the skeletonize method greatly underestimated lobe numbers (Figure 3.4E). Following an initial calibration to optimize the threshold values of δ and λ, the LobeFinder outputs for lobe number closely matched the median lobe numbers from the manually curated images (Figure 3.4F). The LobeFinder lobe number error was 5.7 times lower than that of the skeletonize method (Figure 3.4G). The accuracy of the manual lobe counts were similar to those of 308 LobeFinder when averaged across all individuals (Figure 3.4G); however for a given cell there was considerable spread in the lobe counts among the individuals (Figure 3.4F). For example, the error rate among the individuals differed by more than 20% for 8 of the 15 cells, even though each of the 6 individuals were similarly-trained to score the presence of lobes. This observation reinforces the strong need for objective methods for quantitative analysis of cells with highly variable shapes and sizes.

Two different types of features were typically identified as a lobe. First, there were instances of an undulation along a cell perimeter segment that were independent of a 3-way cell wall junction. This is the classic example of interdigitated growth among two adjacent cells, and we define these features as Type I lobes. A second class of cell protrusions, defined here as Type II lobes, were instances in which a protrusion was located at a 3-way cell wall junction. These tripartite junctions form during cytokinesis, and in some, but not all cases, the cell can grow asymmetrically at this location generating a protrusion with a shape that is often indistinguishable from Type I lobes. However, the growth mechanism that generates a Type II lobe may resemble "intrusive" growth [35] in which one cell expands asymmetrically at the interface of two adjacent cells. This form of asymmetric growth likely differs from that which generates lobes that are independent of tripartite junctions. There is certainly a need to distinguish between these different types when one analyzes phenotypes and gene function. Currently, this is a weakness of LobeFinder because the program operates on the coordinates of individual cells and information on the cell wall patterns of its neighboring cells is lost. At present, if a user wishes to distinguish Type I and Type II lobes, one can use the graphical output from LobeFinder to identify the subset of lobes that fall on 3-way cell wall junctions. In the future, we hope to use a semi-automated cell segmentation program like CellECT [33] to simultaneously extract cell coordinates from fields of cells and track the positions of 3-way cell wall boundaries.

To quantitatively evaluate the performance of the algorithm for lobe location, we compared the position of the predicted lobes against the manually determined lobes within a specified tolerance (0.025 radians). If LobeFinder identified the proper location within the tolerance, it is identified as a true positive (TP, Figure 3.5). If a predicted LobeFinder point was not within 0.025 radians of a manually identified point, it was considered a false positive (FP). Missed Lobe points were defined as false negatives (FN). We did not calculate true negatives since this would be an ambiguous number to determine and it would not inform the evaluation of the method. Related to these quantities, we also calculated the Sensitivity = TP/(TP+FN) and False Discovery Rate (FDR) = FP/(TP+FP). Both of these measures are used to determine the effectiveness of the algorithm.

A high sensitivity and a low FDR are the primary objectives for the application of LobeFinder as a tool for reliable and automated measurement of cell shape properties. Nine different combinations of the parameters δ and λ were tested that covered a wide range of parameter values. The highest parameter values yielded decreased senstivity and increased false positives; however, there was a fairly broad range of parameter combinations that yielded a sensitvity of ~0.8 and an FDR of ~0.25 (Figure 3.5). This indicates a relatively low 349 dependence of the algorithm on the specific parameters. The optimized parameter combinations yielded an average sensitivity of 0.95 or higher and an average FDR less than 0.2 (Figure 3.5).

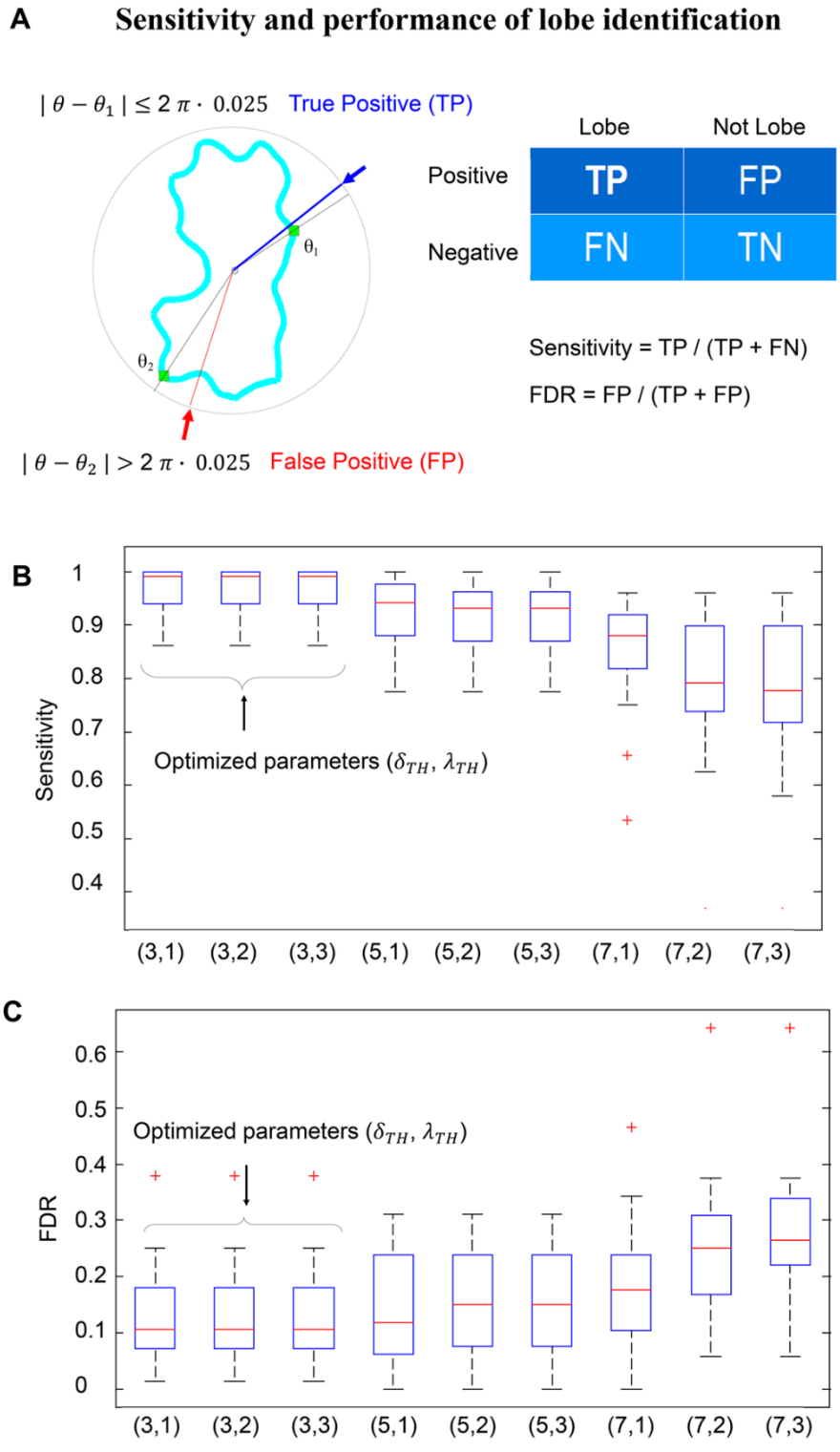


Figure 3.5. Sensitivity and accuracy analysis of LobeFinder performance. (A) Definitions of sensitivity and False Discovery Rate (FDR). (B) Sensitivity of LobeFinder outputs over a range of parameter values. (C) FDR of LobeFinder outputs over a range of parameter values.

3.3.3 Identification of new lobes in time-lapse images of pavement cells

The lobe number, shape, and size properties of pavement cells were analyzed on populations of cells at different intervals of cotyledon development. We applied the LobeFinder program to identify lobes in three time-lapse datasets of pavement cell growth. The first dataset represents early growth from 38 hours after germination (HAG) to 56 HAG. This slightly overlaps with the second dataset from 48 HAG to 120 HAG. The third dataset covers 72 HAG to 120 HAG. 359 Datasets 2 and 3 were part of a previous analysis of pavement cell growth [4], and the raw images were reanalyzed here using LobeFinder. Example cells from these three different datasets are shown in Figure 3.6A to Figure 3.6C. Each showed combinations of symmetrical lateral expansion, with datasets 1 and 2 including more cells that initiated new lobes during the time interval. For example, the cell that is representative of the 72-120 HAG dataset, we observed no change in lobe number over the course of time, but it increased in size (Figure 3.6C, left to right). In contrast, the image representing a cell in the 48-120 HAG dataset initiated five new lobes (Figure 3.6B, left to right) while the image in the 38-55 HAG added three new lobes in the timespan of 27 hours (Figure 3.6A, left to right). There is a great deal of variability in pavement cell size and lobe number as a function of cotyledon and leaf development [16,18]. As a result, in the relatively small windows of time that are analyzed here, there are examples in which lobe number and cell area are not strictly correlated with developmental time (Table 3.2). However, differences in lobe initiation rates of individual cells within the time intervals were apparent. Overall, the average number of new lobes per cell was about 2.5 for the 38-55 HAG and 48-120 HAG populations, and 0.5 for 72-120 HAG (Table 3.2). The percentage of cells in all datasets that grew new lobes, were 33 percent for 72-120 HAG, 93 percent for 48-120 HAG, and 80 percent for 38-55 HAG (Table 3.2). These LobeFinder outputs and the average number of lobes per cell at each time point (Table 3.2) indicate that lobing events are prevalent in early stages of growth, and that lobing events slow down at some point between 56 and 72 HAG. These results are consistent with the conclusions of a previous study (Zhang et al., 2011).

Table 3.2. Lobe number quantification for cotyledon pavement cells using LobeFinder.

	38 to 55 HAG*		48 to 120 HAG		72 to 120 HAG	
	38 H	55 H	48 H	120 H	72 H	120 H
Averaged lobe number	9.60 ± 2.68	12.10 ± 2.99	8.27 ± 2.89	10.87 ± 2.59	11.17 ± 2.89	11.67 ± 2.46
Percent with lobe initiation	80		93		33	
Average new lobes per cell	2.50 ± 2.46		2.60 ± 1.68		0.50±2.07	

* For 38 to 55 HAG, N=10 cells, 48 to 120 HAG N=12 cells, 72 to 120 HAG N=12 cells

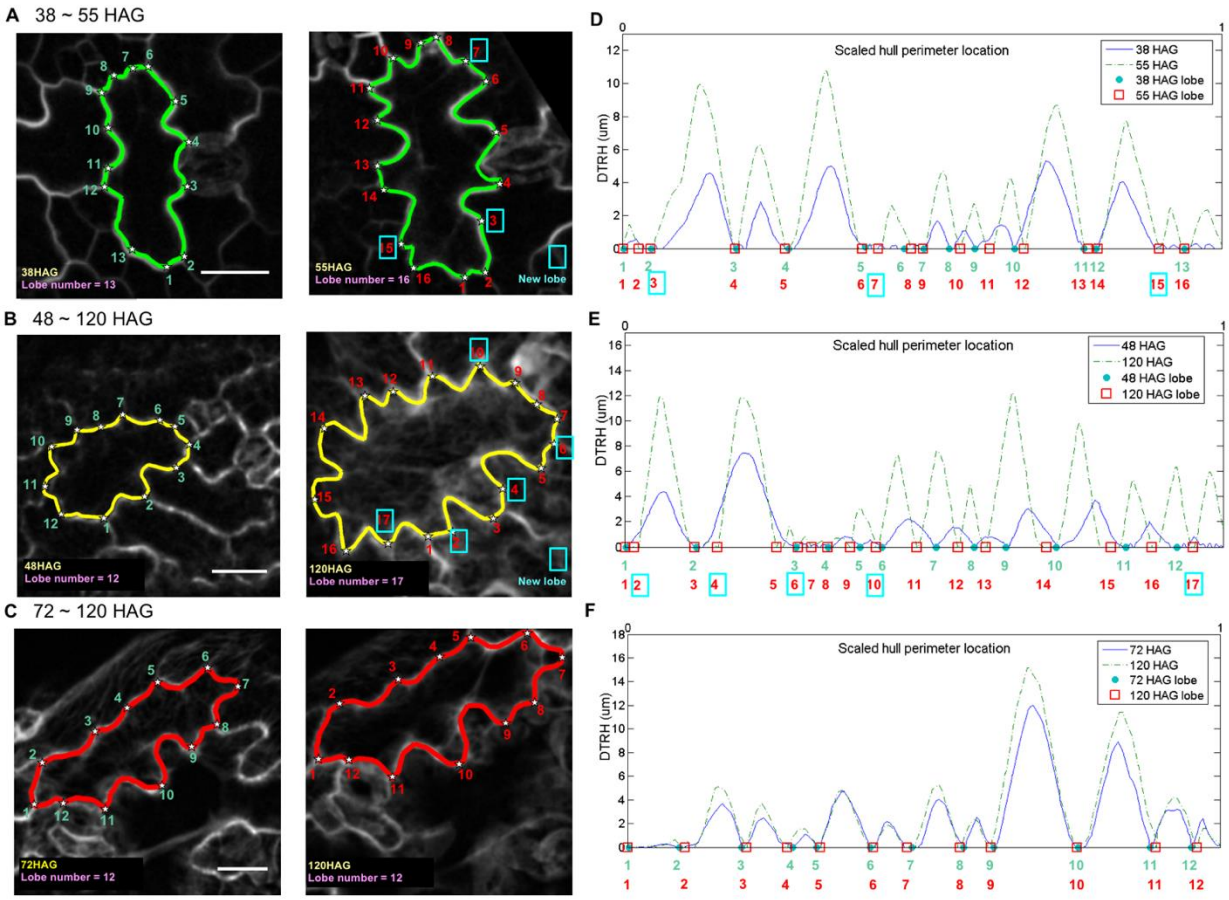


Figure 3.6. LobeFinder can be used to detect new lobes and quantify growth patterns in time-lapse images. (A-C), Examples of raw images of pavement cells with manually segmented cell shapes at three different intervals of cotyledon development. (A) Pavement cell at 38 (left) and 55 (right) HAG. (B) Pavement cell at 48 (left) and 120 (right) HAG. (C) Pavement cell at 72 (left) and 120 (right) HAG. The blue boxes indicate the detection of new lobes and their location in the images and on the DTRH plots. (D-F) The DTRH plots for pavement cells that were re-scaled to their original size. The x-axes of these plots are the scaled distance along the convex hull perimeter at the two different time points to enable visual comparisons of similar relative positions along the cell boundary at the two time points. The blue line is the DTRH at the initial time-point and dotted green line is DTRH at the final time-point. The time points in D to F, correspond to those of A to C, respectively, and are shown in the legend for each of the plots. The blue dots and red boxes on the x-axis identify lobe locations in the initial and final time points, respectively. Scale bar, 20μm

Additional scalar metric outputs from LobeFinder also correlate with different phases of pavement cell growth; however, they do not directly inform the generation of new lobes. Specifically, for example, the circularity of the individual cells decreases between the two time points (Table 3.3), likely due to the increased expansion of lobes that are initiated primarily in the first two and half

days after germination. This would also explain the observed decreases in the convexity (ratio of hull perimeter to cell perimeter) and solidity (ratio of hull area to cell area) of pavement cells. Overall, the identification of lobing events and the scalar metrics are consistent with the existence of a permissive developmental window for active lobe formation early in cotyledon development.

Table 3.3. Cell shape descriptors of cells analyzed with LobeFinder.

	38 to 55 HAG		48 to 120 HAG		72 to 120 HAG	
	38 H	55 H	48 H	120 H	72 H	120 H
Circularity	0.60 ± 0.09	0.46 ± 0.11	0.62 ± 0.10	0.49 ± 0.11	0.45 ± 0.13	0.42 ± 0.12
Roundness	0.69 ± 0.06	0.63 ± 0.06	0.70 ± 0.09	0.64 ± 0.09	0.63 ± 0.08	0.61 ± 0.07
Convexity	0.93 ± 0.04	0.85 ± 0.07	0.94 ± 0.03	0.87 ± 0.06	0.84 ± 0.10	0.82 ± 0.10
Solidity	0.86 ± 0.03	0.79 ± 0.05	0.85 ± 0.06	0.80 ± 0.06	0.77 ± 0.06	0.75 ± 0.05

Another outputs from LobeFinder that are useful for cell analyses are plots of the distance from the refined hull to the cell boundary, which provide a graphical representation of the magnitudes and directions of cell shape change near the cell periphery. This is due to the fact that as lobes expand, their height and width increase leading to corresponding changes in the DTRH plots. At the distal tips of cell protrusions, the DTRH is zero and corresponds to a lobe point of the cell of interest in the LobeFinder output. The shape of the cell boundary between lobes is captured by the contour of the DTRH, which is at a local maximum at the most concave position between lobes. Therefore, in a time lapse experiment, the DTRH plots reflect the local growth behavoirs of the adjacent protruding cell, and the shape change at the interface between the two cells. In Figure 3.6D to Figure 3.6F, the DTRH was plotted for each cell at the two different time points. The position along the hull is the x-axis, and this is scaled to the hull length of the initial time point to enable the DTRH values from different time points to be compared at similar relative positions along the hull perimeter. During the 72 to 120 HAG interval (Figure 3.6F), growth is highly symmetrical and lobe initiation is rare [4]. The corresponding DTRH plots were consistent with this result, because the contours of the plots at the two time points were highly symmetrical with well-aligned peaks. It is important to note that the peak widths for the later time points are

compressed because the x-axis is scaled. However, as previously shown [4], pavement cell growth during this phase is not perfectly symmetrical, and there were sub-regions of the DTRH plots that were not symmetrical (Figure 3.6F), indicating that some local warping of cell shape occurred during growth. The paired DTRH plots for cells that form new lobes (Figure 3.6E and Figure 3.6F), reflected a composite growth behavior. In some regions of the cell-cell interface growth appeared symmetrical with proportional increases in peak height and width at similar relative positions. The DTRH plots also revealed an obvious contribution of polarized growth to the shape change, because new peaks were detected. In addition, many of the peaks were shifted in position along the hull perimeter, reflecting symmetry breaking during lobe initiation and accumulation of local warping during the growth interval.

3.4 Discussion

LobeFinder is a novel convex hull-based tool to quantify the local boundary characteristics of a closed geometric shape and identify key features such as pavement cell lobes. The ability of LobeFinder to consistently and accurately identify and position lobes within a pavement cell is an important advance, because currently there is no reliable method to quantify the convoluted shape of pavement cells. Manual definition of lobe number [15, 36] or a feature such as the pavement cell neck-width (the shortest distance across the cell between two indentations) [37] is subjective and variable. Variation in human scoring is a major problem: we document here significant variability in lobe number scoring, even among well-trained individuals (Figure 3.4E and Figure 3.4G). The lack of standardized phenotyping methods can contribute to differing conclusions as to whether or not a particular mutant has a pavement cell phenotype [36,38]. Manual cell scoring is also time-consuming. It requires careful inspection of the cell boundary and the manual annotation of each feature in the image file that is scored as a lobe. Skeletonization of segmented, binary images of cells can identify pavement cell protrusions [18], and in some instances can be used to detect significant differences between mutant and wild type plants [19]. However, the skeletonize technique is very inaccurate and tends to miss approximately 40% of all lobes (Figure 3.4G). LobeFinder has a much greater accuracy compared to the skeletonize method, and performs with an accuracy that is only achieved by averaging the votes of several individuals with extensive experience in pavement cell analysis (Figure 3.4G).

The availability of an accurate method to directly identify pavement cell lobes is important because scalar shape descriptors such as circularity are sensitive to multiple features of a cell geometry, and do not contain information on the local cell features that are most useful for understanding cell growth behavior. For example, differences between cells in their scalar descriptors could reflect either differences in lobe number, reduced lobe expansion, or altered diffuse growth in the mid-region of the cells. This point is important because it is often assumed that any mutant with a reduced perimeter to area ratio has a lobe initiation defect. LobeFinder directly analyzes the local geometry of the cell and identifies lobes. In this regard, it is a powerful phenotyping tool that can be used to compare populations of cells and cell shape over time. However, the LobeFinder program is not perfect, and because of the local shape and spacing thresholds that are used for lobe detection, there are instances, most often along extended domains of low curvature, in which false positives are occasionally reported. Overall, LobeFinder has great potential for the community, and we anticipate that LobeFinder, provided as a user-friendly program in MatLab (Figure 3.7), will allow others to use this program to analyze mutants, and objectively test for direct effects on lobe initiation.

A major advantage of LobeFinder is that it creates a coordinate system to quantify local growth behaviors at the interface of two cells. Alternative approaches to lobe detection, such as the quantification of the local curvature of the cell perimeter using variation in the tangent to the cell boundary as a function of cell perimeter, could operate on splined images to identify regions of local curvature that accurately identify lobes. However, this strategy would not generate a coordinate system to analyze growth. Here we use LobeFinder and plots of the DTRH in time series data to illustrate a method to quantify local growth behaviors of an irregularly shaped cell (Figure 3.6). For example, the DTRH plots could be analyzed further in time-lapse experiments to generate spatial maps of how the magnitude and direction of growth at the interface of two cells change. These plots clearly indicate the timing, location, and pattern of polarized growth along the interface of interdigitating pavement cells. Importantly, these plots do not reveal the subcellualr patterns of growth that explain the shape change. To solve this problem, convex-hull based growth analysis coupled with the use of fiducial marks on the cell wall to track growth patterns [4,16,18] could provide improved methods to analyze the subcellular heterogeneity in polarized growth. While this paper was being written, there was a recent report in which externally applied particles

were used to track the growth patterns of the outer wall in fields of developing pavement cells [39]. The utility of externally applied particles to analyze the growth of the anticlinal wall is uncertain. However, the combined use of DTRH plots, high-density cell wall marking, and time lapse imaging has the potential to reveal how the polarized growth of individual cells and cell clusters can operate at broader spatial scales to dictate the growth patterns of leaf sectors and even whole organs [4,5,40].

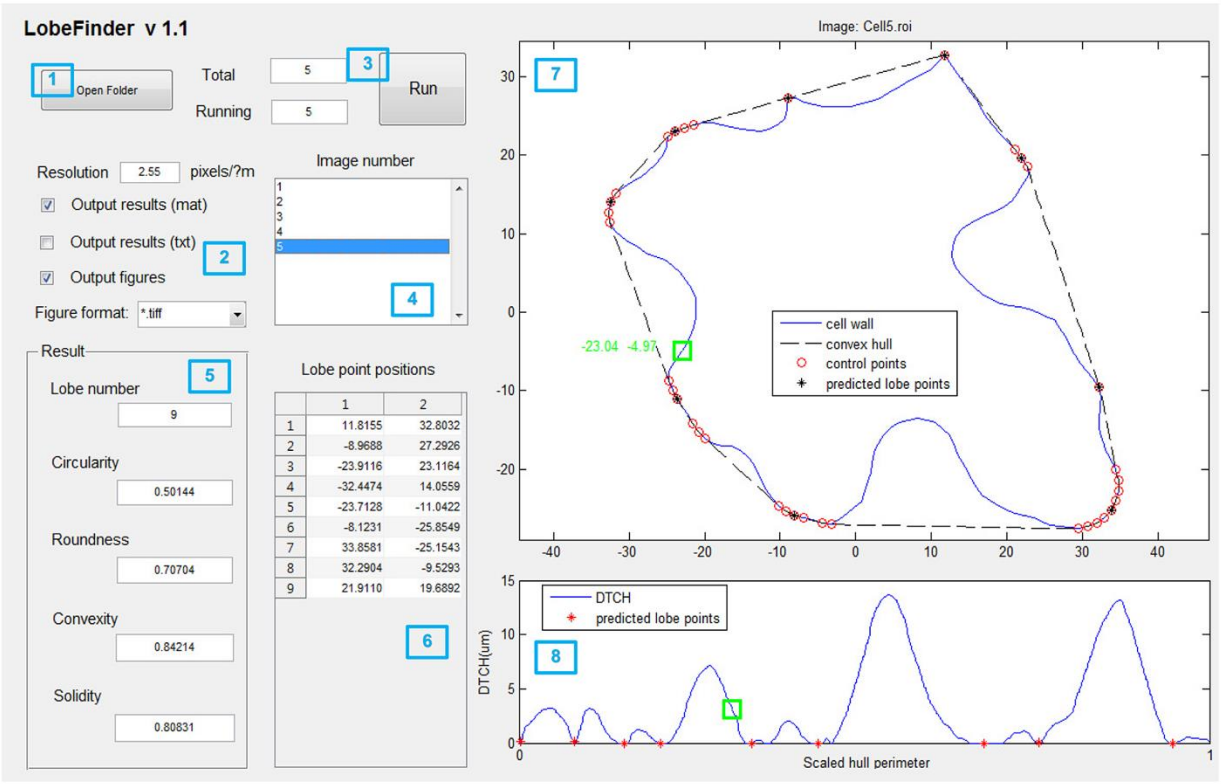
LobeFinder also has immediate applications in terms of more quantitatively dissecting the molecular control of lobe initiation. Hull based methods and the DTRH plots establish a perimeter coordinate system onto which the temporal and spatial patterns of lobe formation can be graphed (Figure 3.6). This is a boon for further analysis such as correlation of spatial geometric features with the localization of cell wall and intracellular signaling and structural factors that are believed to control symmetry breaking. Specifically, LobeFinder can provide the convex hull coordinate system to test for correlations between the local accumulation of proteins such as auxin efflux carriers [15,36] or microtubules [41,14,7,42,4] and lobe initiation. In this manner, an array of GFP-tagged proteins can be tested to determine those whose localization and activity at the cell cortex specify symmetry-breaking events.

Currently, the greatest limitation for LobeFinder is that it operates on cell coordinates from manually segmented cells. Manual segmentation is a reliable, but time consuming process, presenting a major bottleneck for high-throughput phenotyping. As discussed previously, the use of individual, segmented cells also makes it impossible to distinguish between Type I and Type II lobes, which complicates one's ability to test for alterative genetic control mechanisms and differing contributions of the lobe types to cell expansion. The obvious solution is an automated cell segmentation program that accurately extracts cell boundary coordinates and marks 3-way cell wall junctions in the dataset. Currently, there is no existing segmentation method to accurately extract pavement cell coordinates from fields of cells and track 3-way junctions. However, the development of watershed-based cell segmentation coupled with user-guided validation in a program like CellECT [33] has the potential to be integrated with LobeFinder to create a more robust and efficient cell analysis pipeline.

We show here that LobeFinder is an effective new tool for pavement cell phenotyping and growth analysis. We believe this algorithm has a broader utility for the quantification and many lobed cell types [7] and the analysis of objects with closed and highly irregular geometric shape at any spatial scale. For example, there is great interest in the quantitative analysis of leaf shape, and the complex boundary shapes of many types of leaves could be analyzed with LobeFinder. In this context, LobeFinder, could complement other leaf shape analysis programs like LeafProcessor [44]. Similar hull-based methods could operate on projected images of shoots and roots to analyze whole plant architecture over time. We also believe LobeFinder could have broad applications in other fields such as human medicine and environmental science. For example, the progression of irregularly-shaped tumors could be quantified over time, and correlated with other features such as tumor location or drug treatment regimes. The local spatial dynamics of spreading plumes of contamination, floods, and the retraction of glaciers could be similarly analyzed, and tested for cross-correlations with any variable of interest. Our efforts will focus on the use and integration of LobeFinder within a completely automated image analysis platform, with the goal of accelerating discovery in the field of leaf morphogenesis.

Starting the LobeFinder_GUI

1. Click the ‘’Open Folder’’ button to select folder where the ROI files are located.
2. Select checkbox(es) for output data type and/or figure format options.
3. Click ‘’Run’’ button to start LobeFinder.



Results of LobeFinder_GUI

4. Select image ROI number in the box to display the LobeFinder result.
5. Predicted lobe numbers and Cell-shape descriptors.
6. Predicted lobe x-y coordinates.
7. Cell perimeter, convex hull, and predicted lobe point positions.
8. Corresponding distance to the convex hull with predicted lobe point locations on it.

Supplemental Figure S3. Graphical user interface of LobeFinder.

Operation steps and navigation of the results are outlined.

Figure 3.7. Snapshot of the graphical user interface of LobeFinder. Operation steps 638 and navigation of the results are outlined.

3.5 References

1. Zhu, X. G., Long, S. P., & Ort, D. R. Improving photosynthetic efficiency for greater yield. Annual review of plant biology, 61, 235-261 (2010).

2. Walter, A., Silk, W. K., & Schurr, U. Environmental effects on spatial and temporal patterns of leaf and root growth. Annual review of plant biology, 60, 279-304 (2009).

3. Savaldi-Goldstein, S., Peto, C., & Chory, J. The epidermis both drives and restricts plant shoot growth. Nature, 446(7132), 199 (2007).

4. Zhang, C., Halsey, L. E., & Szymanski, D. B. The development and geometry of shape change in Arabidopsis thaliana cotyledon pavement cells. BMC plant biology, 11(1), 27 (2011).

5. Kuchen, E. E., Fox, S., De Reuille, P. B., Kennaway, R., Bensmihen, S., Avondo, J., ... & Coen, E. (2012). Generation of leaf shape through early patterns of growth and tissue polarity. Science, 335(6072), 1092-1096 (2012).

6. Asl, L. K., Dhondt, S., Boudolf, V., Beemster, G. T., Beeckman, T., Inzé, D., ... & De Veylder, L. Model-based analysis of Arabidopsis leaf epidermal cells reveals distinct division and expansion patterns for pavement and guard cells. Plant physiology, 156(4), 2172-2183 (2011).

7. Panteris, E., & Galatis, B. The morphogenesis of lobed plant cells in the mesophyll and epidermis: organization and distinct roles of cortical microtubules and actin filaments. New phytologist, 167(3), 721-732 (2005).

8. Szymanski, D. B., & Cosgrove, D. J. Dynamic coordination of cytoskeletal and cell wall systems during plant cell morphogenesis. Current Biology, 19(17), R800-R811 (2009).

9. Geitmann, A., & Ortega, J. K. Mechanics and modeling of plant cell growth. Trends in plant science, 14(9), 467-478 (2009).

10. Szymanski, D. B. The kinematics and mechanics of leaf expansion: new pieces to the Arabidopsis puzzle. Current opinion in plant biology, 22, 141-148 (2014).

11. Sampathkumar, A., Krupinski, P., Wightman, R., Milani, P., Berquand, A., Boudaoud, A., ... & Meyerowitz, E. M. Subcellular and supracellular mechanical stress prescribes cytoskeleton behavior in Arabidopsis cotyledon pavement cells. Elife, 3, e01967 (2014).

12. Tsuge, T., Tsukaya, H., & Uchimiya, H. Two independent and polarized processes of cell elongation regulate leaf blade expansion in Arabidopsis thaliana (L.) Heynh. Development, 122(5), 1589-1600 (1996).

13. Fu, Y., Li, H., & Yang, Z. The ROP2 GTPase controls the formation of cortical fine F-actin and the early phase of directional cell expansion during Arabidopsis organogenesis. The Plant Cell, 14(4), 777-794 (2002).

14. Qiu, J. L., Jilk, R., Marks, M. D., & Szymanski, D. B. The Arabidopsis SPIKE1 gene is required for normal cell shape control and tissue development. The Plant Cell, 14(1), 101-118 (2002).

15. Fu, Y., Gu, Y., Zheng, Z., Wasteneys, G., & Yang, Z. Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. Cell, 120(5), 687-700 (2005).

16. Elsner, J., Michalski, M., & Kwiatkowska, D. (2012). Spatiotemporal variation of leaf epidermal cell growth: a quantitative analysis of Arabidopsis thaliana wild-type and triple cyclinD3 mutant plants. Annals of botany, 109(5), 897-910 (2012).

17. Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., ... & Inzé, D. Exit from proliferation during leaf development in Arabidopsis thaliana: a not-so-gradual process. Developmental cell, 22(1), 64-78 (2012).

18. Staff, L., Hurd, P., Reale, L., Seoighe, C., Rockwood, A., & Gehring, C. (2012). The hidden geometries of the Arabidopsis thaliana epidermis. PLoS One, 7(9), e43546 (2012).

19. Le, J., Mallery, E. L., Zhang, C., Brankle, S., & Szymanski, D. B. Arabidopsis BRICK1/HSPC300 is an essential WAVE-complex subunit that selectively stabilizes the Arp2/3 activator SCAR2. Current Biology, 16(9), 895-901 (2006).

20. Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., & Ecker, J. R. CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases. Cell, 72(3), 427-441 (1993).

21. Le, J., El-Assal, S. E. D., Basu, D., Saad, M. E., & Szymanski, D. B. Requirements for Arabidopsis ATARP2 and ATARP3 during epidermal development. Current Biology, 13(15), 1341-1347 (2003).

22. Djakovic, S., Dyachok, J., Burke, M., Frank, M. J., & Smith, L. G. BRICK1/HSPC300 functions with SCAR and the ARP2/3 complex to regulate epidermal cell shape in Arabidopsis. Development, 133(6), 1091-1100 (2006).

23. Zhang, C., Mallery, E. L., Schlueter, J., Huang, S., Fan, Y., Brankle, S., ... & Szymanski, D. B. Arabidopsis SCARs function interchangeably to meet actin-related protein 2/3 activation thresholds during morphogenesis. The Plant Cell, 20(4), 995-1011 (2008).

24. Blilou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Friml, J., ... & Scheres, B. The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature, 433(7021), 39 (2005).

25. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Tinevez, J. Y. Fiji: an open-source platform for biological-image analysis. Nature methods, 9(7), 676 (2012).

26. Russ, J. C. The image processing handbook. CRC press (2016).

27. Robert, C., Noriega, A., Tocino, Á., & Cervantes, E. Morphological analysis of seed shape in Arabidopsis thaliana reveals altered polarity in mutants of the ethylene signaling pathway. Journal of plant physiology, 165(9), 911-919 (2008).

28. Marcuzzo, M., Quelhas, P., Campilho, A., Mendonça, A. M., & Campilho, A. Automatic cell segmentation from confocal microscopy images of the Arabidopsis root. In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 712-715). IEEE (2008).

29. Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A., & Ortiz-De-Solorzano, C. 3D reconstruction of histological sections: Application to mammary gland tissue. Microscopy research and technique, 73(11), 1019-1029 (2010).

30. Graham, R. L. An efficient algorithm for determining the convex hull of a finite planar set. Info. Pro. Lett., 1, 132-133 (1972).

31. Ma, W. Y., & Manjunath, B. S. EdgeFlow: a technique for boundary detection and image segmentation. IEEE transactions on image processing, 9(8), 1375-1388 (2000).

32. Roeder, A. H., Chickarmane, V., Cunha, A., Obara, B., Manjunath, B. S., & Meyerowitz, E. M. Variability in the control of cell division underlies sepal epidermal patterning in Arabidopsis thaliana. PLoS biology, 8(5), e1000367 (2010).

33. Delibaltov, D. L., Gaur, U., Kim, J., Kourakis, M., Newman-Smith, E., Smith, W., ... & Manjunath, B. S. CellECT: cell evolution capturing tool. BMC bioinformatics, 17(1), 88 (2016).

34.Yoder N PeakFinder http://www.mathworks.com-/matlabcentral/ fileexchange/25500 (2011)

35. Jura, J., Kojs, P., Iqbal, M., Szymanowska-Pułka, J., & Włoch, W. Apical intrusive growth of cambial fusiform initials along the tangential walls of adjacent fusiform initials: evidence for a new concept. Australian journal of botany, 54(5), 493-504 (2006).

36. Xu, T., Wen, M., Nagawa, S., Fu, Y., Chen, J. G., Wu, M. J., ... & Yang, Z. Cell surface-and rho GTPase-based auxin signaling controls cellular interdigitation in Arabidopsis. Cell, 143(1), 99-110 (2010).

37. Lin, D., Cao, L., Zhou, Z., Zhu, L., Ehrhardt, D., Yang, Z., & Fu, Y. Rho GTPase signaling activates microtubule severing to promote microtubule ordering in Arabidopsis. Current Biology, 23(4), 290-297 (2013).

38. Gao, Y., Zhang, Y., Zhang, D., Dai, X., Estelle, M., & Zhao, Y. Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. Proceedings of the National Academy of Sciences, 112(7), 2275-2280 (2015).

39. Armour, W. J., Barton, D. A., Law, A. M., & Overall, R. L. Differential growth in periclinal and anticlinal walls during lobe formation in Arabidopsis cotyledon pavement cells. The Plant Cell, 27(9), 2484-2500 (2015).

40. Remmler, L., & Rolland-Lagan, A. G. Computational method for quantifying growth patterns at the adaxial leaf surface in three dimensions. Plant Physiology, 159(1), 27-39 (2012).

41. Panteris, E., Apostolakos, P., & Galatis, B. Microtubules and morphogenesis in ordinary epidermal cells ofVigna sinensis leaves. Protoplasma, 174(3-4), 91-100 (1993).

42. Ambrose, J. C., Shoji, T., Kotzer, A. M., Pighin, J. A., & Wasteneys, G. O. The Arabidopsis CLASP gene encodes a microtubule-associated protein involved in cell expansion and division. The Plant Cell, 19(9), 2763-2775 (2007).

43. Kirik, V., Herrmann, U., Parupalli, C., Sedbrook, J. C., Ehrhardt, D. W., & Hülskamp, M. CLASP localizes in two discrete patterns on cortical microtubules and is required for cell morphogenesis and cell division in Arabidopsis. Journal of Cell Science, 120(24), 4416-4425 (2007).

44. Backhaus, A., Kuwabara, A., Bauch, M., Monk, N., Sanguinetti, G., & Fleming, A. LEAFPROCESSOR: a new leaf phenotyping tool using contour bending energy and shape cluster analysis. New phytologist, 187(1), 251-261 (2010).

45. Wu, T. C., Belteton, S. A., Pack, J., Szymanski, D. B., & Umulis, D. M. LobeFinder: a convex hull-based method for quantitative boundary analyses of lobed plant cells. Plant physiology, 171(4), 2331-2342 (2016).