Tracing Branched Curvilinear Structures with a Novel Adaptive Local *PCA* Algorithm

Liya Wang[1, 2, *], Amir H. Assadi[2], and Edgar P. Spalding[1]
[1]Department of Botany, University of Wisconsin, Madison, WI, USA
[2]Departments of Mathematics, University of Wisconsin, Madison, WI, USA

Abstract – *A vast variety of branched structures differing in size, shape, and function are found in Nature and engineering. Needed are robust and sufficiently general methods for extracting geometric information about structure-function arrangements and relationships from images of branched structures. This paper describes a novel mathematical algorithm called Adaptive Local Principal Component Analysis (ALPCA) for tracing branched curvilinear structures. Utility of this algorithm is demonstrated by its application to quantitative analysis of shape in various plant organs undergoing development. Specifically, time series of digital images representing seedling stems (hypocotyls) responding to light or root systems undergoing branching were automatically analyzed by ALPCA. Automated quantification of such structures changing size and shape will enable algorithmic classification of phenotypes (consequences of altered gene functions) and therefore add a new technological element to the field of plant functional genomics. The general design of ALPCA enables the tracing of gray-scale digital images of other categories of branched structures.*

Keywords: Tracing curvilinear structures, principal component analysis, image processing, *Arabidopsis*.

1 Introduction

Tracing of curvilinear structures is a fundamental tool in the quantitative analysis of biological images, for extracting information about structures such as blood vessels, neurons, microtubules, and similarly-shaped entities. In general, a combinatorial representation known as the *medial axis* or *midline* is sufficient to provide a natural representation of the structure of interest. In practice, image analysis routines may or may not successfully extract the medial axis due to noise and clutter obscuring the object of interest in the images, which is often due to limitations of the data-acquisition techniques. For example, the present work involves monitoring the development of seedlings in darkness by forming images with infrared radiation. In plant biology (as well as many other domains) such images are typically analyzed manually by a laborious and approximate set of quantification instructions. In this work, we propose a robust curvilinear tracing algorithm that is insensitive to problems frequently encountered in biological imaging by combining computational intelligence and image processing techniques. The types of images studied below frequently present complications such as frequent intersection, occlusion and overlap, occasional loss of signal along the curve appearing as gaps, and high levels of noise due to water condensation within the experimental chamber. The proposed adaptive local principal component analysis algorithm traces the branched curvilinear structures in a fully automatic fashion, and it returns medial axis points in the desired order.

The paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the proposed algorithm for adaptive curvilinear structure tracing. Section 4 contains an application to developmental biology of the model plant *Arabidopsis thaliana*, including quantitative and qualitative tracing results. Finally, conclusions and work in progress are presented in section 5.

2 Related work

While the mathematical origins of the Singular Value Decomposition (*SVD*) trace back to the theory of linear systems of equations developed by the Italian mathematician Eugenio Beltrami, its statistical version, the popular Principal Components Analysis (*PCA*) has been independently discovered several times during the 20th century in different contexts such as regression, factor analysis and applications of dimensionality reduction. Limitations of *PCA* to data sets that are not normally distributed are well-known, and attempts to overcome the constraints on probability distribution of data have opened new directions in non-parametric regression and nonlinear generalizations of *PCA*, such as Local-to-Global *PCA* [1], Kernel *PCA* [2], principal curves [3], principal surfaces and manifolds [4] and a host of new and specialized algorithms motivated by specific applications. In [3], the authors construct a nonlinear generalization of the first principal axis of a data set in the Euclidean space, the so-called local principal curve (*LPC*), and suggest that it could serve as the best one-dimensional curvilinear approximation to the data (analogous to a nonlinear regression curve.) *LPC*s are based on the localization of principal component analysis, in connection with the mean shift technique [5]. The *LPC* algorithm constructs the principal curve, a series of local centers of mass, in a *bottom-up* manner. Every step of *LPC* considers only data in a local neighborhood of the currently

considered point. In contrast to *LPC*, the *ALPCA* algorithm introduces a new local-to-global integration method that adaptively adjusts the step size and the local neighborhood dimension (the kernel size), by incorporating prior information, such as, pixel intensity features.

The combined use of geometry and intensity information allows *ALPCA* to overcome the possible shortcomings of other tracing algorithms for extracting curvilinear structures from images (e.g. the skeletal extraction and mathematical morphology in image processing.) Partial Differential Equation methods, such as diffusion and level set methods [6] use flow along curvature lines, boundary evolution, or other typically curvature-dependent flows such as the geodesic distance. The above-mentioned PDE-based methods do not use the direction information along the curve, so that, any neighboring structure could bias the trace. Curvature-guided techniques [7] could fail (and they often do) when a reliable and robust gradient cannot be computed around the intersections. In contrast to *ALPCA*, skeleton-extraction algorithms [8] that generate artifacts known as "barbs" due to small imperfections in the region boundary, and get heavily distorted for overlapping curvilinear structures, while in such situations, the *ALPCA* provides a smooth and precise medial axis.

3 Algorithm

In particular, our tracing algorithm provides precise medial axis points from a single local eigenvalue problem, making computation fast and easy. The key idea is to slide a circle that slightly overfills the object which is registered as a 2-dimensional "thickened smooth curve" of varying width, such as the hypocotyl image shown in Fig.1. The center of each circle is defined as a medial axis point. Successive medial axis points are determined by moving the circle in the appropriate direction guided by the morphology, while the radius is adjusted for a proper fit. This determines the coordinates of the new center, which is the new medial axis point. In short, *ALPCA* performs local *PCA* analysis to obtain the direction (using the 1st eigenvector), while it uses the 2nd eigenvalue to obtain the radius of the new circle. The local center of mass is determined to be the new medial axis point. Thus, the center of a rolling circle with a variable radius defines the medial axis point. While *ALPCA* appears closely related to the local principal curve method, the following details demonstrates its adaptive capability to capture the medial axis points of sharply curved structures with varying radii.

3.1 Initialization

The digital image that is processed by *ALPCA* is a two-dimensional data cloud represented by an *N-by-2* vector *P*. Since images are likely to be noisy, we preprocess the image by smoothing to improve the signal-to-noise ratio, or even remove the noise completely in favorable circumstances. The corresponding intensity values of a pixel $X = (x, y)$ is denoted by I_X, and are stored in the vector *I*. Next, construct a weight function,

$$W(I_X) \propto \frac{1}{|I_X - I_O| + 1}, \tag{1}$$

where I_O denotes the intensity value at the current medial axis point, and $w(I_X)$ weights every point *X* in its neighborhood by I_X.

Given any starting point $X = (x, y)$, its neighborhood is defined as a circle whose center and radius are estimated by a 2-dimensional Gaussian that fits the intensity distribution around *X*. The fitting with smaller variance gives μ_0 and σ_0. The position of the desired circle can be calculated from μ_0 and the coordinates of given starting point *X*. Thus, the radius of the first circle tightly fitting the curvilinear planar structure is,

$$r_0 = \frac{2\sigma_0}{p},$$

where *p* is a positive constant < 1 to ensure overfilling of the linear object. A small *p* results in a larger circle, that could track the regions of higher curvature with less precision. A larger *p* results in a smaller circle which could fail the algorithm in case of abrupt shape change as in sites of branching. Thus, the optimal choice of *p* depends on the trade-off between stability (i.e. robust tracking property) versus fidelity (more local accuracy of the approximation of the medial axis). As the traced medial axis extends, the radii of the rolling circles are automatically adjusted by *ALPCA*.

3.2 Tracing

Mathematically speaking, localization refers to the process of using an appropriate bandwidth kernel function to determine a desired neighborhood within the circles. For a symmetric positive definite 2-dimensional bandwidth matrix *H*, bandwidth kernel function is defined as in the following:

$$K_H(X) = |H|^{-1/2} K\left(H^{-1/2} X\right).$$

The medial axis is determined as an ordered set of local centers of mass driven by *ALPCA*. Given all components of X are measured on the same scale, the bandwidth matrix is defined by

$$H = \begin{bmatrix} h^2 & 0 \\ 0 & h^2 \end{bmatrix} = \eta^2 \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}, \tag{2}$$

Here, $h = \eta \cdot r$, and η is initially set to 0.5 and will be automatically adjusted with curvatures (see below). The Gaussian kernel $K(X) = 1/2\pi \exp\left(-1/2\, X^T X\right)$, allows us to rewrite $K_H(X)$ as follows:

$$K_H(X) = \frac{1}{2\pi|H|}\exp\left(-\frac{1}{2}X^T H^{-1} X\right).$$

This is also the density of the bivariate normal distribution with a mean vector $\mathbf{0}$ and a covariance matrix equal to $\boldsymbol{H}$ [6]. Given intensity I_X for any point X, a weighted bandwidth kernel function is defined as,

$$K_{HW}(X) = W(I_X)K_H(X).$$

The intensity-weighted local center of mass around x is calculated as below, resulting in pixels far from x or with different intensity values with the center of tracing circle receiving less weight.

$$\mu(x) = \frac{\sum_{i=1}^{n} K_{HW}(X_i - x)X_i}{\sum_{i=1}^{n} K_{HW}(X_i - x)},$$

The $(j,k)^{\text{th}}$ entry $(1 \le j,k \le d)$ of the local covariance matrix ($\sum^x$) of x is given by

$$\sigma_{jk}^x = \sum_{i=1}^{n} w_i\left(X_{ij} - \mu_j^x\right)\left(X_{ik} - \mu_k^x\right),$$

where $w_i = K_{HW}(X_i - x)\Big/\sum_{i=1}^{n} K_{HW}(X_i - x)$, and the abbreviation μ_j^x denotes the *j*-th component of $\mu(x)$. Let γ^x be the first eigenvector of $\sum^x$ and λ_k be the normalized 2nd eigenvalue at *k*-th step of tracing. An update value of x is acquired as

$$x' = \mu^x + t\gamma^x,$$

where the step size *t* is chosen to be correlated with the object width and curvature for better precision, as discussed below. Without loss of generality, we assume that the object width varies continuously along the direction that the object extends. Thus, the update rule for the radius is the given by the following estimate:

$$r_{k+1} = r_k \cdot \sqrt{\left(\frac{\lambda_{k+1}}{1-\lambda_{k+1}}\right)\Big/\left(\frac{\lambda_k}{1-\lambda_k}\right)}, \qquad (3)$$

where λ_{k+1} is computed from a rough estimation of $r'_{k+1} = r_k \cdot \sqrt{(\lambda_k/(1-\lambda_k))/(\lambda_{k-1}/(1-\lambda_{k-1}))}$. In this way, width change is initially estimated by the preceding step, then adjusted by (3), and as a result, as the local center of mass is translated forward, the radius of circle is automatically adjusted to tightly fit the object width.

The stopping rule for tracing is decided based on either arriving at image edge or registering the convergence of the local centers of mass.

3.3 Performance accuracy

The accuracy of the computed medial axis depends on the step size *t*. Although smaller *t* gives higher precision, the computation time increases linearly given its time complexity *O(n)*. More closely spaced points are required in tightly curved regions while straight regions require fewer points to accurately describe the medial axis. Thinner regions require smaller step sizes to capture the dynamic change over time. To compensate for less reliability in local center of mass and local *PCA* calculations, step size is first correlated with object width as shown below:

$$t_{k+1} \propto r_k .$$

As mentioned earlier, curvature of the object is also incorporated into the calculation of the step size. The difference in direction vectors among two consecutive steps is given by $\alpha_{(k)}^x = \left|\cos\left(\alpha_{(k)}^x\right)\right| = \left|\gamma_{(k-1)}^x \circ \gamma_{(k)}^x\right|$. The medial axis is first smoothed by adjusting the current direction,

$$\gamma_{(k)}^x = \alpha_{(k)}^x \cdot \gamma_{(k)}^x + \left(1 - \alpha_{(k)}^x\right) \cdot \gamma_{(k-1)}^x .$$

Then $\alpha_{(k)}^x$ is re-calculated and the step size is finally determined as,

$$t_{k+1} \propto r_k e^{\alpha_{(k)}^x - 1}. \qquad (4)$$

In this manner, both object width and curvature are incorporated into the algorithm in order to gain higher accuracy and at the same time, maintain a short computation time.

Furthermore, for a sharply curved object, the resulting medial axis is easily biased towards the inner boundary curve. That is, the width-fitting circle must capture the characteristic geometric shape due to presence of large curvature. To adjust tracing parameters to the geometry of such planar shapes, the algorithm automatically adapts the parameter η of equation (2) according to curvature, in order to provide an appropriately exaggerated weighting function for sharply curved regions. The formula:

$$\eta = 0.5 \cdot e^{\alpha_{(k)}^x - 1},$$

has the desired weighting property that is desirable for accuracy of tracing the medial axis when much less weight must be assigned to irregular boundary regions.

Substitution of $h = \eta \cdot r$ (cf (2)) in assertion (4) provides the following estimate that establishes the high correlation between the automatically adjusted kernel size and the computed medial axis:

$$t_{k+1} \propto h_k . \qquad (5)$$

4 Application

As an example application of our proposed curvilinear structure tracing algorithm, we tested it on the *Arabidopsis* development video. The *Arabidopsis thaliana* plant is a

valuable model system for studying plant gene function. An important source of information in such gene function studies is a mutant phenotype, some observable difference in a plant missing the gene under study relative to the wild type. The dynamic growth information of both the stem (hypocotyls) elongation and root development is well related to particular genes. The model plant is quite small and usually studied in a small Petri plate. Consecutively imaging with an infrared camera has the advantages of keeping plant under development intact and easily capturing the dynamic information. Some particular problems of this technique include the condensation that severely degrades the image contrast and the uneven illumination resulted from the non-uniform distributed infrared light. Specifically, the vaporization of agar solution due to the infrared light heating effect condensates on the relative cold Petri plate cover, and results in blurred images. Note that uneven infrared light background causes considerable signal degradation in addition to the condensation. Furthermore, frequent intersection and overlaps cause intensity variations along the curvilinear objects. Thus, typically tremendous manual pre-processing was needed even for analyzing images captured at a single time point.

Use of *ALPCA* algorithm has reduced the manual preprocessing work to zero for the quantitative analysis of both hypocotyls and lateral roots development. The rich information extracted during robust tracing process has been easily combined and carried into the feature point detection for hypocotyls analysis, which is crucial for separating hypocotyls from cotyledons and thus provide precise morphological measurements. Furthermore, the easy addition of auxiliary circles along with the main tracing circle allows *ALPCA* to trace lateral roots that cross over and overlap heavily in a totally automatic fashion. Overall, the application here points out that *ALPCA* can be easily adapted for robustly tracing various curvilinear structures in image analysis.

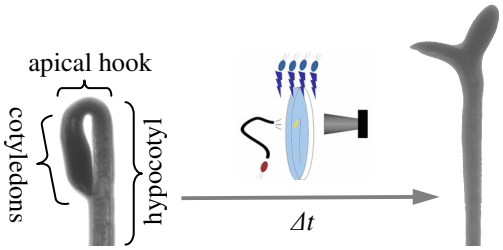

Fig. 1. Examples of recorded images show the morphological changes of apical hook, cotyledons and hypocotyl during seedling growth. Also shown are a CCD camera, infrared light behind the Petri plate and blue light on top.

4.1 Hypocotyls

In the case of Arabidopsis seedlings growing initially in darkness, extreme curvature of the apical hook results in the cotyledons being pressed against the hypocotyl as shown in Fig. 1. After turning on the blue light, the apical hook opens and the cotyledons bifurcate in less than 10 hours. The elongation rate of the hypocotyls and the opening of apical hook with time is regulated by many environmental and endogenous (e.g. hormonal) factors. Much progress toward the understanding of gene function and the effects of light, temperature, stress, etc. has been made by studying hypocotyl growth in wild-type and mutant plants.

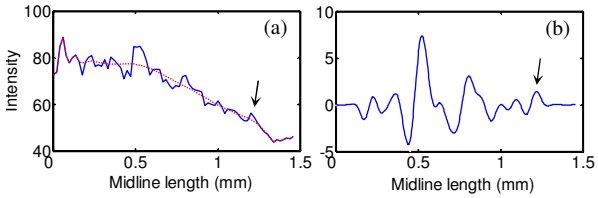

Fig. 2. (a) Intensity values at medial axis points are plotted as a function of the medial axis length (solid line); Fitted baseline is plotted as dashed line. Arrow points to the peak corresponding to the bifurcation point to be tracked. (b) Subtraction of the baseline from the intensity is smoothed and plotted as a function of medial axis length.

Because cotyledons appear darker in images than hypocotyls, due to their smaller, more opaque cells, intensity weighting in (1) allows *ALPCA* to trace the hypocotyl upward through the cotyledons intersection. Fig. 2(a) shows how image intensity along the traced medial axis (solid line) decreases from hypocotyl to cotyledons. The arrow points to a lighter feature, a peak that corresponds to the end of hypocotyls and the "bifurcation point" of cotyledons. Subtracting the baseline (dashed line in Fig. 2(a)) highlights this useful feature peak in Fig. 2(b). To take advantage of this unique but weak feature in an automated fashion, baseline removal and locating of a "feature window" were necessary.

4.1.1 Baseline substraction

Using iteratively re-weighted least square algorithm, robust analysis [10] provides stable regression results even in the presence of outliers. Piecewise robust analysis has been utilized in NMR data validation [11] by applying it on non-overlapped fractions of data separately. Here, by sliding a window over the data (overlapped fractions), robust local regression was implemented to subtract the baseline and enhance the bifurcation signal ("outlier"). The algorithm can be formalized as (6). Assuming local linearity ($\hat{g}(x_i) = \theta_0 + \theta_1(x_i - x_0)$), a weighted least squares technique can be applied to a fraction of the data around any point x_0,

$$\hat{\theta}(x_0) = \arg\min_{\theta} \sum_{i=1}^{n} w_r(x_i) K\left(\frac{x_i - x_0}{h_i'}\right)\left[y_i - \hat{g}(x_i)\right]^2 , \quad (6)$$

where $K(u) = \left[\max\left(1 - |u|^3, 0\right)\right]^3$ is a weight function for the point *i*, which descends smoothly to zero and is zero outside the neighborhood defined by h_i'. $w_r(x_i)$ is called

"robustness weight" assigned to each point,

$$w_r(x_i) = \begin{cases} 1 & \text{if} \quad r_i < 0, \\ \left\{\max\left[1-(r_i/b)^2, 0\right]\right\}^2 & \text{otherwise}, \end{cases}$$

where $r_i = \left[y_i - \hat{g}(x_i)\right]/\sigma$ is the residuals of the initial/previous least squares fitting divided by the scale parameter $\sigma = median\left[y_i - \hat{g}(x_i)\right]/0.6745$. The parameter b determines the "robustness" of the procedure. Assuming the baseline is asymmetrically "contaminated" by the peak, the asymmetric $w_r(x_i)$ is applied for repeating weighted least squares technique until convergence.

Because the bifurcation signal can be very weak, the above baseline subtraction routine was repeated multiple times. To avoid over-fitting the boundary points with the repeated procedure and to take into account the variable step size taken in the medial axis tracking process, the local neighborhood is taken as:

$$h_i' = \left[1 + \cos(\pi + i \cdot 2\pi/N)\right] \cdot h_i + 1,$$

where N is the total number of medial axis points within the locked window (discussed below). With this definition, boundary points get a smaller neighborhood which decreases their chance of being over-fitted.

4.1.2 Locating feature window

As shown in Fig. 2(b), intensity variability along the medial axis is quite large compared with the bifurcation signal. This variability is in part due to dark and light patterns formed by the end walls of cells in the hypocotyl. Intensity varies less in the region around the bifurcation point, probably due to the smaller cell size in this area. Thus a "feature window" containing the bifurcation signal was identified to filter out the "noisy signal" and correctly locate the "bifurcation point". Moreover, considering that the iterative robust baseline subtraction algorithm is time consuming, faster computation results from applying it to a fraction instead of all of the medial axis points.

Fig. 3 summarizes the entire process of medial axis tracing and bifurcation point. As shown in the *bifurcation point tracking* section, the start of the feature window is determined by locating the medial axis point associated with the minimal stem width, which is captured by the 2nd eigenvalue at the point. Then the feature window size L is taken as twice the hypocotyl width W_h to cover the bifurcation point,

$$L = 2W_h \approx 4\bar{h},$$

where $\bar{h} = \langle h_i \rangle$ is the average value of the kernel widths applied within such a window. Given a stack of images each separated by a short time interval, the feature windows can be easily relocated by centering it in each successive image at previous bifurcation point without determining the minimal width point each time, since the later one is no longer a robust feature if condensations are accumulated.

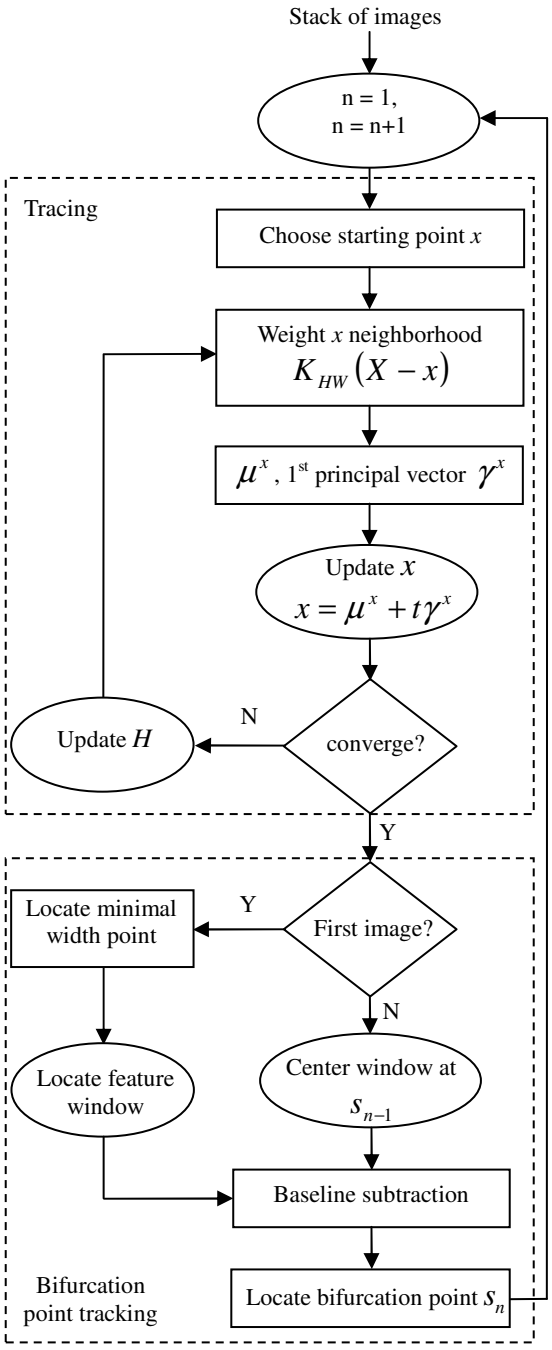

Fig. 3. Flowchart of the automatic hypocotyl growth analysis algorithm given a stack of images.

4.1.3 Results

The average growth rate of hypocotyl with standard error is shown in Fig. 4(a). Blue light from above inhibits hypocotyl elongation [12] in a fluence-rate (intensity) dependent manner. In the experiment shown, relatively weak blue light (1 μmol photons m^{-2} s^{-1}) was applied after 2 hours of growth in darkness. The well-known transient slowing of growth rate followed by a sustained inhibitory phase was automatically detected by the *ALPCA* algorithm.

Fig. 4(b) shows the hook angle, also automatically extracted from the images series by *ALPCA*, as a function of time. The inserted images show how the hook angles are calculated. The direction at the bifurcation point (round dot) is given by the first eigenvector *ALPCA*. The other direction is obtained by traveling back L along the medial axis from the bifurcation point and taking the first eigenvector from there. The hook angle is defined as the angle between these two. The directions at all medial axis points are acquired through the tracing process, thus no extra calculations are needed here.

As demonstrated, the information acquired through *ALPCA* is highly correlated with the featured bifurcation point detection as well as the hook angle calculation. The robustness of all steps in the quantitative analysis is guaranteed by such a one-time calculation of all required pieces of information.

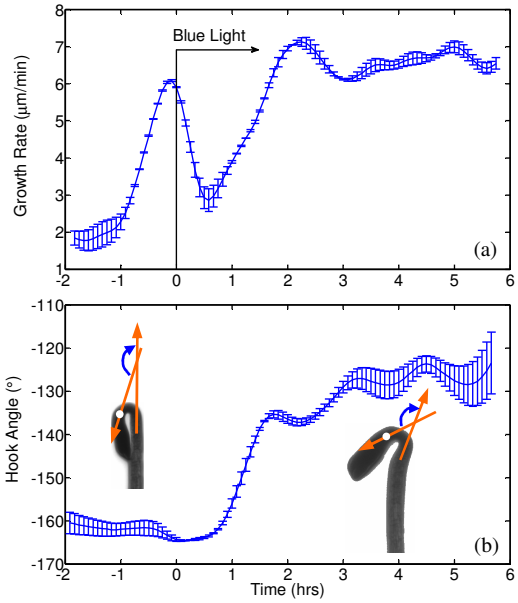

Fig. 4. (a) Hypocotyl growth rate is plotted as a function of time with light on time pointed. (b) Hook angle measurement is plotted as a function of time. Error bars represent the standard error of the mean from 20 individual seedlings.

4.2 Lateral roots

About 3-4 days after the bifurcation of cotyledons and under proper conditions, *Arabidopsis* plants start to develop a significant number of lateral roots. In addition to the primary root, development of lateral roots is important for absorbing adequate nutrition from the plants' environment. As a result, the genes that regulate this process are of particular interest to plant biologists. Quantification of growth of lateral roots by image analysis would ideally require 3-D image acquisition of the plant growth in space. While the 3-dimensional growth would be more realistic, 2-D images acquired in a Petri plate could still provide adequate amounts of information about root development, and in particular, it is more suitable for high throughput genotype or phenotype screening.

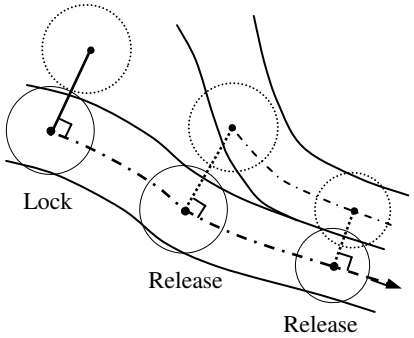

Fig. 5. "Lock and Release" schema on two overlapping roots.

One challenge encountered in image analysis of plants growing on agar surface in a Petri plate is crossover and overlapping of lateral roots that are far more frequent in comparison with the natural growth in three dimensions. Tracking crossovers can be easily solved by *ALPCA* with the 1st eigenvector, but overlaps cannot be tracked by the same method. To solve this problem and to increase the robustness of *ALPCA* tracing algorithm, we introduce an improvement of *ALPCA* by adding *an auxiliary circle* along with the main tracing circle. Together, their function is to perform simultaneous tracing of lateral roots via a "lock and release" schema as shown in Fig. 5.

In Fig. 5, the bottom branch is the lateral root under tracing and it is touched by another lateral root from the top. Two auxiliary circles (dashed, only one is shown) are attached to the main tracing circle (solid). The auxiliary circle always moves in same direction and same step size as the main tracing circle; but it can be in either lock or release state, as follows:

1. Lock – *tangent to the main circle and with the same radius as the main circle.*
2. Release – *tangency to the main circle is not required and radius varies as (3).*

Before being released, the auxiliary circles do nothing instead of counting $w(I_x)$ inside them. At a certain threshold, the auxiliary circle is released (tangency not required) so that it can vary its radius based on (3). Then the auxiliary circle can be either outside of, or overlapping with its adjacent main circle. For the overlapping region between two circles, *ALPCA* is further weighted by the distances to the centers of the main and auxiliary circles.

Fig. 6. shows the tracing results for three images collected at 7 days, 7½ days and 8 days after germination. The tracing starts with the tip of the primary root with a corner detection procedure [13], and then the lateral roots are traced by the released auxiliary circles, which are then becoming the new "main tracing circles" and use their own auxiliary circle to trace through the overlapping regions. Fig. 6. shows that severed overlapping of lateral roots can be properly traced.

Furthermore, to acquire the dynamic information, such as growth rate, images can be collected at certain time intervals, such as every 15 minutes for lateral root

development. To automate the process, first locate the primary root tip for the first image, and then scan its corresponding neighborhood for the adjacent image. The result of tracing in image *n* can also be used in tracing image *n+1* to speed up the tracing. This frame-amalgamation of image processing could also serve to catch undetected lateral roots, for instance due to the blurring caused by condensation.

As in the quantitative analysis of hypocotyls growth, the information collected here gives us the length, growth rate, curvature, distribution of lateral roots, and other measurements. Analysis of raw images could be performed without any manual operation, and the process captures precise quantities for root development.

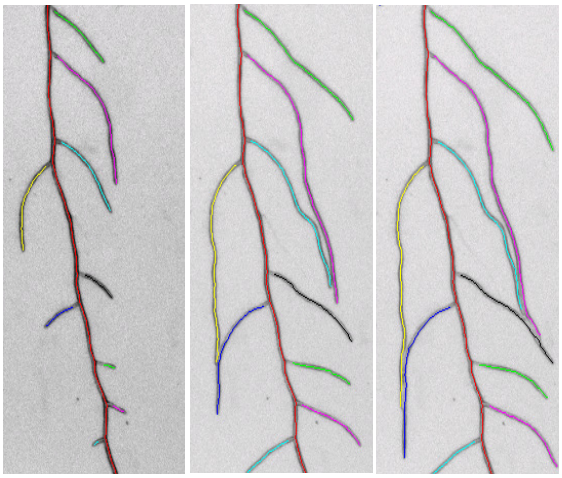

Fig. 6. Examples of tracing overlapping lateral roots. The medial axis of primary root is shown in red and all traced lateral roots are shown in various colors.

5 Conclusion

We have presented in this paper a tracing algorithm for curvilinear structures and applied it to plant photo-morphogenetic analysis. The most noteworthy contribution of the proposed algorithm is the easy adaptation for analyzing different objects and the robustness independent of any preprocessing operations. Note that the kernel size or tracing step size is adaptive to the object width and curvature and is automatically adjusted in each tracing step. It is expected that the resulted medial axis not only captures the geometric information of the curvilinear structures but also performs a compression operation for removing possible redundancies. Such compact information will be particular useful for future analysis, such as, high throughput phenotype based gene screening.

In analysis of raw images, the region of interest (ROI) such as roots occupies at most 10% of the total pixels. *ALPCA* uses a "bottom-up" approach to follow along the branched structure, so that the tracing is computationally efficient. Furthermore, the local *PCA* portion of the algorithm requires arithmetic operations on matrices of relatively small size, so that *SVD* is typically completed in minute fractions of a second. Even more, the one time *ALPCA* tracing operation also provides useful angular information along with the medial axis, which usually requires an additional post-processing step.

Finally, computational complexity of *ALPCA* is linear, so results scale favorably in applications to high throughput screening that typically is computationally very demanding. Furthermore, the most demanding part of the computation, such as tracing lateral roots, can be parallelized because *ALPCA* performs only local operations.

6 References

[1] A. H. Assadi, H. R. Eghbalnia, and J. Carew, "Local-to-global topological methods in data analysis and applications to fMRI". *IJCNN '99*. pp. 1010-1015, 1999.

[2] B. Schölkopf, A. Smola, and K. R. Muller, "Kernel Principal Component Analysis", *Advances in Kernel Methods-Support Vector Learning*, MIT Press Cambridge, MA. pp. 327-352, 1999.

[3] J. Einbeck, G. Tutz, and L. Evers, "Local principal curves", *Statistics and Computing*, vol. 15, no. 4, pp. 301-313, 2005.

[4] Z. Zhang, and H. Zha, "Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment", *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313-338, 2005.

[5] D. Comanicu, and P. Meer, "Mean shift: A robust approach toward feature space analysis", *IEEE Trans. Pattern Anal. Machine Intell.* vol. 24, pp. 603-619, 2002.

[6] J. Sethian, "Level Set Methods and Fast Marching Methods", *Cambridge Univ. Press*, 1999.

[7] J. August, and S.W. Zucker, "Skeches with curvature: the curve indicator random field and markov processes", *PAMI*, vol. 25, no. 4, pp. 387-400, 2003.

[8] T. Y. Zhang, and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns", *Comm. ACM*, vol. 27, no. 3, pp. 236-239, 1984.

[9] M. P. Wand, and M.C. Jones, "Comparison of Smoothing Parameterizations in Bivariate Kernel Density-Estimation," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 520-528, 1993.

[10] P. W. Holland, and R. E. Welsch, "Robust Regression Using Iteratively Re-Weighted Least-Squares," *Communications in Statistics Part A-Theory and Methods*, vol. 6, no. 9, pp. 813-827, 1977.

[11] L. Y. Wang, et. al., "Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications," *Journal of Biomolecular NMR*, vol. 32, no. 1, pp. 13-22, 2005.

[12] B. M. Parks, K. M. Folta, and E. P. Spalding, "Photocontrol of stem growth", *Current Opinion in Plant Biology*, vol. 4, no. 5, pp. 436-440, 2001.

[13] G. Harris and M. J. Stephens. "A combined corner and edge detector", *Proceedings Fourth Alvey Vision Conference*, Manchester. pp. 147-151, 1988.

* LIYAWANG@WISC.EDU

FUNDED BY NSF DBI-0621702

RELATED MOVIE – HTTP://PHYTOMORPH.WISC.EDU/MOVIE/