

IMAGE ANALYSIS OF ARABIDOPSIS TRICHOME PATTERNING IN 4D CONFOCAL DATASETS

R. Bensch^{1,5}, O. Ronneberger^{1,5}, B. Greese^{2,3}, C. Fleck², K. Wester⁴, M. Hülskamp⁴, H. Burkhardt^{1,5}

¹Institute of Computer Science, University of Freiburg,
Chair of Pattern Recognition and Image Processing, Georges-Köhler-Allee Geb. 052,
79110 Freiburg, Germany, {bensch,ronneber,burkhardt}@informatik.uni-freiburg.de

²Center for Biological Systems Analysis (ZBSA), University of Freiburg

³Faculty of Biology, University of Freiburg, ⁴Botanical Institute III, University of Köln

⁵Centre for Biological Signalling Studies (bloss), University of Freiburg, Germany

ABSTRACT

In this article, we present an approach for the automated extraction of quantitative information about trichome patterning on leaves of *Arabidopsis thaliana*. Time series of growing rosette leaves (4D confocal datasets, 3D + time) are used for this work. At first, significant anatomical structures, i.e. leaf surface and midplane are extracted robustly. Using the extracted anatomical structures, a biological reference coordinate system is registered to the leaves. The performed registration allows to determine intra- as well as inter-series spatiotemporal correspondences. Trichomes are localized by first detecting candidates using Hough transform. Then, local 3D invariants are extracted and the candidates are validated using a Support Vector Machine (SVM).

Index Terms— Arabidopsis trichome patterning, surface extraction, registration, symmetry plane, trichome localization

1. INTRODUCTION

Motivation. Arabidopsis trichomes (leaf hairs) are an interesting example to study cell differentiation [1] and are well suited cell types for the functional analysis of *de novo* pattern formation. During cell growth, different processes provide for a regular spacing of the trichomes. A close interplay between theoretical modelling and experimental validations has been started to test the relevance of different models for trichome patterning. In this context, there is a need of quantitative data, e.g. trichome positions, to allow the analysis of these processes and their simulation with mathematical models.

State of the art. Lee et al. [2] and Kaminuma et al. [3] already performed a quantitative analysis of spatial distributions of Arabidopsis leaf trichomes, including surface extraction and extraction of trichome positions. Optical projection tomography (OPT) was used in [2] and micro X-ray computed tomography (μ CT) was used in [3] to acquire single channel 3D images. The marching-cubes algorithm, i.e. a threshold-based method, was used in both [2] and [3] to create a polygon model of the surface from the 3D image at a specified isovalue. The extraction of trichomes is based on the surface curvature, i.e. surface morphology, in both approaches. The analysis is based on 3D recordings of a single time step for each plant.

2. APPROACH

In contrast to [2] and [3], we intend to investigate the processes that guide the development of trichomes. For this purpose, it was necessary to record time series (4D data) of growing Arabidopsis leaves. Trichomes are tagged with a trichome specific green fluorescent protein (GFP) marker, such that they can be detected in an early developmental state. Confocal laser scanning microscopy (CLSM) was needed to obtain a higher resolution compared to OPT in [2] and to allow for recordings of GFP markers *in vivo*, which is not possible with μ CT in [3]. Firstly, a chlorophyll channel (representing the leaf volume) was recorded using its auto-fluorescence. Secondly, a GFP channel was recorded to detect trichomes. Variations in the chlorophyll concentration and effects like absorption due to CLSM lead to intensity gradients in the recorded images. Therefore, threshold-based methods for surface extraction can not be used here. We intend to investigate the development of trichomes starting from an early developmental state, when they are not yet morphologically visible, and thus our time series start with young leaves. In both [2] and [3] the detection of trichomes is based on the morphology and therefore is not applicable here because early trichomes are not visible morphologically.

Due to time-consuming reference labeling of trichomes, mainly two time series were selected for this work, series A (7 time steps) and series B (8 time steps). As an example, the dimensions of the youngest and oldest time step of series A are given: t_0 : $256 \times 256 \times 52$ voxels, t_6 : $521 \times 404 \times 158$ voxels of size $1.465 \times 1.465 \times 2.0 \mu\text{m}^3$.

The proposed approach mainly consists of three parts. The first part is the robust extraction of the leaf surface and midplane from the chlorophyll channel. Local filtering based on gradient directions and gradient magnitude is used to identify surface voxels. In addition voxels belonging to the midplane are identified similarly by using eigenvectors of the Hessian instead. Robustness is increased by applying a scale-space approach and removing unstable responses. The consideration of 4D data in our approach introduces the need for image registration. Symmetry analysis, based on the extracted surface, yields the symmetry plane of the leaf. It is used to robustly detect the pose of the leaf and thus to register a biologically motivated reference coordinate system. This allows for inter-subject registration of the leaves. Trichomes are localized in the GFP channel by first detecting candidates using Hough transform, extracting local 3D invariants, and then validating the candidates using a SVM.

3. EXTRACTION OF LEAF SURFACE

Leaf segmentation by thresholding is not appropriate here because the chlorophyll intensities increase along the leaf rib continuously towards the blade tip. An adequate approach to extract the leaf surface, that does not operate directly on the gray values, is presented in the following. Local non-linear filtering is applied to identify surface voxels. The image X is convolved with a Gaussian kernel of width σ . Then, gradient direction and gradient magnitude are calculated. For each voxel four neighbouring voxels are determined. They are located in mutual perpendicular directions on the plane that is perpendicular to the gradient direction and passes through the current voxel. A voxel is identified as a surface voxel if it and all of its four neighbours are local maxima of gradient magnitude along the gradient direction, see Fig. 1. This filter serves as an edge detector that is robust to noise, since also the neighbours are required to be local maxima. A similar technique was already successfully applied in [4]. Robustness to intra- as well as inter-subject variations is achieved using a scale-space approach by applying the filter to image X convolved with Gaussian kernels of widths in the range of $[6\mu m, \dots, 14\mu m]$. All binary filter responses are combined by adding them up, and only stable filter responses, occurring at least at three scales, are conserved and stored in the 3D binary surface mask M . The extracted surface itself is called S , consisting of all voxel positions where mask M has a value of 1. With a similar approach the midplane of the leaf is extracted. Instead of gradient direction the Hessian is computed and the most negative eigenvalue λ_0 and the corresponding eigenvector are used. In the midplane region of the leaf, this eigenvector points in a direction perpendicular to the midplane. Additionally, gray values are used instead of gradient magnitude for local maxima calculation, and the range for the width of the Gaussian kernel is set to $[10\mu m, \dots, 26\mu m]$. To achieve additional robustness, the condition $|\lambda_0| / (|\lambda_0| + |\lambda_1| + |\lambda_2|) > 0.7$ ensures that the eigenvalue λ_0 is dominant. From the extracted 3D surfaces, 2D depth maps are calculated, and the 2D leaf edge results from calculating the leaf contour in the 2D depth map.

4. REGISTRATION TO A BIOLOGICAL REFERENCE COORDINATE SYSTEM

This section describes how a meaningful, plant-based reference coordinate system is registered to the leaves, as depicted in Fig. 2. Using this coordinate system, trichome positions can be represented in a biologically meaningful manner. By registering a common coordinate system, inter-subject registration, within a single and also between different time series, can be performed. To completely define the coordinate system, the orientation and position of the leaf are detected automatically. Simply performing a principle component analysis (PCA) on the distribution of the extracted surface voxels failed, because the main direction is attracted towards additional plant material from the petiole, that is also included in the extracted surface. A robust approach proposed here is to exploit the symmetry property of the leaf to determine its orientation. For this purpose, the symmetry plane is calculated from the extracted leaf surface. In the literature, symmetry planes are used e.g. for brain registration in [5]. There, PCA is sufficient for the determination of the plane normal, whereas the author also refers to the known deficiencies of the PCA method. Here, the plane representation $\langle \mathbf{n}, \mathbf{x} \rangle = \alpha$ is chosen, where \mathbf{n} is the plane normal, α the plane offset from the origin, and \mathbf{x} a point on the plane. The list of plane parameters is $\mathbf{p} = (\varphi, \psi, \alpha)$, where φ and ψ define \mathbf{n} . In the following, the necessary steps for the calculation of the symmetry plane are described.

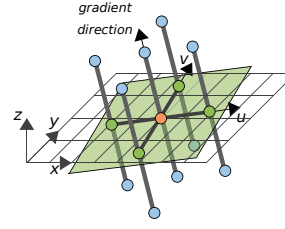


Fig. 1. Filter kernel used for surface extraction.

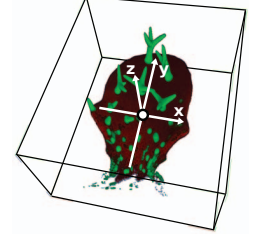


Fig. 2. Plant-based reference coordinate system registered to the leaf.

Symmetry metric. The symmetry metric assigns a value to a specific position \mathbf{x} that indicates its symmetry properties w.r.t. a specific plane with parameters \mathbf{p} , given the surface S . For this purpose, we calculate

$$\mathbf{m}_{\mathbf{p}}(\mathbf{x}) = \mathbf{m}_{\varphi, \psi, \alpha}(\mathbf{x}) = \mathbf{x} - 2(\langle \mathbf{x}, \mathbf{n}_{\varphi, \psi} \rangle - \alpha) \cdot \mathbf{n}_{\varphi, \psi} \quad (1)$$

which is the mirrored position of \mathbf{x} , and $\mathbf{m}'_{S, \mathbf{p}}(\mathbf{x})$, which is the nearest position on the surface S from $\mathbf{m}_{\mathbf{p}}(\mathbf{x})$. Several continuous and binary metrics based on distances to the surface and on surface normals were evaluated. The metric

$$\delta_{S, \mathbf{p}}(\mathbf{x}) = e^{-\frac{1}{2} \frac{\|\mathbf{m}_{\mathbf{p}}(\mathbf{x}) - \mathbf{m}'_{S, \mathbf{p}}(\mathbf{x})\|^2}{\sigma^2}} = e^{-\frac{1}{2} \frac{D_{S, \mathbf{p}}^2(\mathbf{x})}{\sigma^2}} \quad (2)$$

proved to have good properties in describing symmetry. It weights the distance to the surface $D_{S, \mathbf{p}}$ with a Gaussian to achieve values in the range of $[0, \dots, 1]$. Here, $\sigma = 9\mu m$ was chosen. A position that falls directly on the surface by mirroring yields the highest symmetry value of 1, while minor deviations still contribute to the score, which allows us to capture imperfect symmetries. This transformation is known as Gaussian Euclidean Distance Transform (GEDT) [6].

Symmetry score and symmetry plane. A symmetry score for a potential symmetry plane is calculated by adding up the symmetry values for all surface positions S (see sum in Eq. 3). The plane with parameters \mathbf{p}^* that maximizes the symmetry score according to

$$\mathbf{p}^* = (\varphi^*, \psi^*, \alpha^*) = \operatorname{argmax}_{\mathbf{p}} \sum_{\mathbf{x} \in S} \delta_{S, \mathbf{p}}(\mathbf{x}) \quad (3)$$

is the optimal plane w.r.t. the defined metric.

Implementation. The distances required in Eq. 2 are precalculated efficiently by applying a distance transform (DT) to the surface mask M . The DT stores for each volume position the distance to the nearest surface position. A 3D parameter space is defined for the plane parameters φ , ψ and α . Then, the symmetry score is calculated for a range of potential planes and stored in the parameter space. The position of the leaf tip \mathbf{x}_{tip} is estimated from the leaf tip region of the extracted surface. For the parameters φ and ψ a range of $[-40^\circ, \dots, +40^\circ]$ is chosen. These two parameters define \mathbf{n} . An offset for α is calculated such that the plane with normal \mathbf{n} passes through the leaf tip. The parameter α is varied in the range of $[-80\mu m, \dots, +80\mu m]$ around the offset. To define the optimal symmetry plane with parameters \mathbf{p}^* , the local maximum with the highest symmetry score in the parameter space is determined.

Remaining degrees of freedom. To determine the center and the tilt angle of the leaf, the surface mask M is aligned in a way that $\mathbf{n}_{\varphi^*, \psi^*}$ coincides with the x-direction.

The aligned mask M' is then projected in x-direction

$$P(y, z) = \int_{\mathbb{R}} M'(x, y, z) dx \quad (4)$$

Applying a PCA to this 2D grayscale projection reveals the center and the tilt angle using the eigenvector with the largest eigenvalue.

5. LOCALIZATION OF TRICHOMES

In order to quantify experimental results, it is beneficial to automatically detect trichome positions. This section describes our corresponding approach, that mainly consists of two steps.

5.1. Candidate Detection

Since the young trichome candidates are nearly spherical and the base of old trichomes resembles a sphere, the idea is to detect spheres in the trichome channel using a Hough transform [7]. This candidate detection consists of four parts.

1. Selecting voting points. Firstly, noise is removed by convolution with a Gaussian kernel. Then, an edge detection is performed by searching for local maxima of gradient magnitude along gradient direction. For the gradient magnitude, a threshold is defined heuristically, and the gradient direction is restricted to enclose an angle in the range of $90^\circ \pm 45^\circ$ with the leaf surface direction. This accounts for the fact that trichomes are expected to stand approximately orthogonal to the leaf surface.

2. Voting (for a given radius r). For each voter, the accumulator is incremented by one in the distance of r for both the positive and negative gradient direction.

3. Post-processing. The raw accumulator is convolved with a Gaussian kernel of width σ_r that is relative to the radius r to compensate for imperfect spheres and to obtain smooth local maxima. In addition to that, the accumulator is weighted according to the proximity to the leaf surface. In this way, the fact that trichome bases are located in the proximity of the leaf surface is incorporated.

4. Searching spheres. To detect spheres, local maxima in the Hough space are identified.

The steps 2 to 4 are performed for radii in the range of $[3\mu\text{m}, \dots, 18\mu\text{m}]$. The resulting list of candidates consists of all of the found local maxima.

5.2. Candidate Validation

Feature extraction. To validate the candidates, local 3D invariants are extracted at each candidate position $\mathbf{q} = (q_x, q_y, q_z)^T$. To gain invariance, the Haar integration framework is used. Invariance can be achieved by integration over the transformation group using a kernel function, that is an arbitrary, non-linear combination of the gray values in X . Here, two kernel functions have been chosen, firstly,

$$h_{1,i}(X) = \int_0^\infty w_i(r) \cdot X(r, 0, 0) dr \quad (5)$$

and secondly, $h_{2,i}$, for which X is substituted by the gradient magnitude $\|\nabla X\|$. The triangular weighting function in Eq. 6

$$w_i(r) = \begin{cases} \left(1 - \frac{|r-r_i|}{\Delta r}\right) & , \frac{|r-r_i|}{\Delta r} \leq 1 \\ 0 & , \frac{|r-r_i|}{\Delta r} > 1 \end{cases} \quad (6)$$

weights the values with a value of 1 at position r_i , which is the radius of an i -th spherical shell. The weighting factor decreases linearly to

a value of 0 for the width Δr in both positive and negative direction. To gain invariance, we integrate over the group of 3D rotations around \mathbf{q} in the sense of Haar integration and obtain the description of two parts of our final feature vector, firstly,

$$f_{1,\mathbf{q},i}(X) = \frac{1}{4\pi r_i^2} \int_{\mathbf{x} \in \mathbb{R}^3} w_i(\|\mathbf{x}\|) \cdot X(\mathbf{q} + \mathbf{x}) d\mathbf{x} \quad (7)$$

and secondly, $f_{2,\mathbf{q},i}$. For $f_{2,\mathbf{q},i}$ we use the kernel function $h_{2,i}$ instead, and thus X is substituted by $\|\nabla X\|$. The feature vector $\mathbf{f}_{1,\mathbf{q}}$ ($\mathbf{f}_{2,\mathbf{q}}$) describes gray value (gradient magnitude) averages on spherical shells of radius r_i around position \mathbf{q} , where i is the index of the feature. To gain invariance to linear gray value transformations, the feature vectors are normalized to unit mean and are called $\mathbf{f}'_{1,\mathbf{q}}$ and $\mathbf{f}'_{2,\mathbf{q}}$ in the following.

Implementation. Here, 20 indices have been chosen for both $\mathbf{f}_{1,\mathbf{q}}$ and $\mathbf{f}_{2,\mathbf{q}}$, where the last index corresponds to the radius $r = 1.8 \cdot r_{\mathbf{q}}$, and $r_{\mathbf{q}}$ is the radius of a candidate at position \mathbf{q} . In this way, the indices of the features are normalized to the scale of the candidates. The final feature vector is composed as follows: $\mathbf{f}_{\mathbf{q}} = (r_{\mathbf{q}}, \mathbf{f}'_{1,\mathbf{q}}, \mathbf{f}'_{2,\mathbf{q}})$. When averaging the gray values over too few elements, noise is introduced. The number of elements we use to average the gray values decreases when the radius of the shell, i.e. the index of the feature, decreases. To limit the noise, we define a minimum number of elements to average over. Hence, we calculate a first valid index for which the minimum number of elements is achieved and that depends on $r_{\mathbf{q}}$. The values for invalid indices are set to 0, and normalization to unit mean is only performed for the valid indices.

Classification. A Support Vector Machine (SVM) is trained and used to classify the candidates as “trichome” or “no trichome”. For SVM we use the implementation LIBSVM¹, perform grid-search to optimize the parameters, whereas linear and RBF kernels are investigated. Candidates for the training set are extracted from time series A, whereas time series B is used as a test set. After classification, overlapping candidates are removed based on the SVM decision-value.

6. RESULTS

Extraction of Leaf Surface and Registration. Figure 3 shows results for the extraction of the surface and midplane for series A, t_3 . Figures 4d and 4e compare the parameter space for the applied metric GEDT and for mutual information (MI) applied to the raw gray values. MI was chosen as comparison, because it performed best among standard similarity measures like sum of absolute differences (SAD), cross correlation and MI. Figures 4a and 4b show the resulting symmetry plane for the best local symmetry value in each case (crosses in Fig. 4d and 4e). As the example shows, the GEDT metric performs better than standard metrics like MI. The GEDT parameter space is smoother, has only few local maxima (here MI even yields 246 maxima in the 3D parameter space compared to 2 maxima for GEDT), and the best local symmetry value yields the expected symmetry plane, which is not the case for MI here. GEDT does not directly depend on the gray values and thus is robust to gray value transformations, as they e.g. appear for the reflected leaf part when the leaf is tilted to the recording direction. Furthermore, the influence of additional plant material is decreased if it is rejected by surface extraction. GEDT is robust to imperfect symmetries and thus also to

¹LIBSVM: <http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvm>

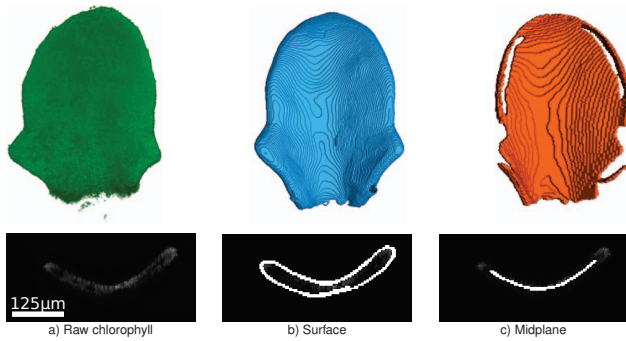


Fig. 3. Extraction of leaf surface and midplane.

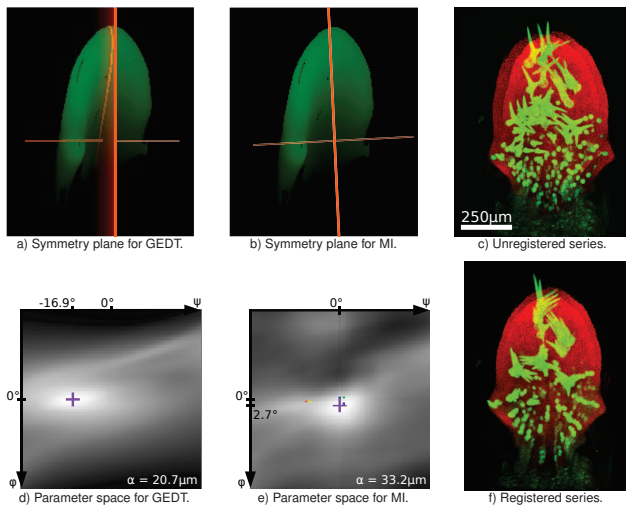


Fig. 4. Parameter space (d, e) and symmetry plane (a, b) for GEDT and MI (Series A, t_0). Unregistered and registered series A (c, f).

missing leaf parts. Besides robustness, our approach has a better runtime since symmetry evaluation only has to iterate the surface voxels and not all of the voxels as for standard metrics like MI. Due to the smooth parameter space, our approach can be accelerated by applying gradient ascent methods. The parameter space was evaluated completely here to investigate different symmetry metrics. Figures 4c and 4f show the result of registering a complete time series.

Localization of Trichomes. Figure 5 shows results for the localization of trichomes. The recognition rates using a RBF kernel, which performed best, are listed in table 1 for the complete series B. A manual labeling was used as “ground truth”. The results obtained so far are significantly better than earlier results achieved with a threshold and morphology based approach.

7. CONCLUSION AND OUTLOOK

The proposed methods allow for a robust extraction of the leaf surface and midplane. Symmetry analysis based on GEDT is used to robustly detect the pose of the leaf. This method for pose estimation may be applied to various biological structures. It further allows for inter-subject registration. Localization of trichomes uses a detection step, extraction of local 3D invariants and a classification step. For the localization, first results are presented that are supposed to be

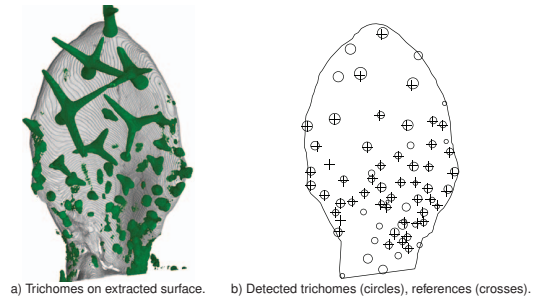


Fig. 5. Localization of trichomes. Detected trichome positions and radii (circles), reference positions (crosses). (Series B, t_7)

Step	Recall / precision	Step	Recall / precision
t_7	90% / 68.2% (45/5/21)	t_3	70.3% / 86.7% (26/11/4)
t_6	93.9% / 83.6% (46/3/9)	t_2	65.5% / 86.4% (19/10/3)
t_5	84.3% / 89.6% (43/8/5)	t_1	59.3% / 88.9% (16/11/2)
t_4	78.3% / 94.7% (36/10/2)	t_0	61.9% / 50% (13/8/13)

Table 1. Localization results (RBF kernel) for series B. Recall/precision (hits/misses/false alarms). The tolerance for matching detections against the reference is set to $\tau = 10\mu m$.

improved by e.g. improving the local features, incorporating time consistency, and training with examples from several series.

8. ACKNOWLEDGMENTS

This study was supported by the Excellence Initiative of the German Federal and State Governments (EXC 294) and BMBF Project FRISYS No. 0313921.

9. REFERENCES

- [1] M. Hülkamp, “Plant trichomes: a model for cell differentiation,” *Nat Rev Mol Cell Biol*, vol. 5, no. 6, pp. 471–480, 2004.
- [2] K. Lee, J. Avondo, H. Morrison, L. Blot, M. Stark, J. Sharpe, A. Bangham, and E. Coen, “Visualizing plant development and gene expression in three dimensions using optical projection tomography,” *Plant Cell*, vol. 18, no. 9, pp. 2145–2156, 2006.
- [3] E. Kaminuma, T. Yoshizumi, T. Wada, M. Matsui, and T. Toyoda, “Quantitative analysis of heterogeneous spatial distribution of arabidopsis leaf trichomes using micro x-ray computer tomography,” *Plant J*, vol. 56, pp. 470–482(13), 2008.
- [4] O. Ronneberger, Q. Wang, and H. Burkhardt, “Fast and robust segmentation of spherical particles in volumetric data sets from brightfield microscopy,” in *Proc. of the ISBI*, 2008, pp. 372–375.
- [5] Xin Liu, C. Imielinska, A.F. Laine, and A. D’Ambrosio, “Symmetry based multi-modality registration of the brain imagery,” in *Proc. of the ISSPIT*, Cairo, 2007, pp. 807–812.
- [6] M. Kazhdan, B. Chazelle, D. Dobkin, T. Funkhouser, and S. Rusinkiewicz, “A reflective symmetry descriptor for 3d models,” *Algorithmica*, vol. 38, pp. 201–225, 2003.
- [7] D. H. Ballard, “Generalizing the hough transform to detect arbitrary shapes,” *Pattern Recogn*, vol. 13, no. 2, pp. 111–122, 1981.