CrossMark

Extracting complex lesion phenotypes in *Zea mays*

Derek Kelly[1,2] · **Avimanyou Vatsa**[3] · **Wade Mayham**[3] · **Toni Kazic**[3]

Abstract Complex phenotypes are of growing importance in agriculture and medicine. In *Zea mays*, the most widely produced crop in the world (United States Department of Agriculture. World Agricultural Production. United States Department of Agriculture, Foreign Agricultural Service, Washington, 2015), the disease lesion mimic mutants produce regions of discolored or necrotic tissue in otherwise healthy plants. These mutants are of particular interest due to their apparent action on immune response pathways, providing insight into how plants protect against infectious agents. These phenotypes vary considerably as a function of genotype and environmental conditions, making them a rich, though challenging, phenotypic problem. To segment and quantitate these lesions, we present a novel cascade of adaptive algorithms able to accurately segment the diversity of *Z. mays* lesions. First, multiresolution analysis of the image allows for salient features to be detected at multiple scales. Next, gradient vector diffusion enhances relevant gradient vectors while suppressing noise. Finally, an active contour algorithm refines the lesion boundary, producing a final segmentation for each lesion. We compare the results from this cascade with manual segmentations from human observers, demonstrating that our algorithm is comparable to humans while having the potential to speed analysis by several orders of magnitude.

✉ Derek Kelly
derkelly@mail.med.upenn.edu

[1] Interdisciplinary Plant Group, Informatics Institute, Missouri Maize Center, University of Missouri, Columbia, USA

[2] Present Address: 421 Curie Boulevard, Philadelphia, PA 19104, USA

[3] Interdisciplinary Plant Group, Department of Computer Science, Informatics Institute, Missouri Maize Center, University of Missouri, Columbia, USA

1 Introduction

1.1 The impact of phenotypic complexity on image segmentation

From the susceptibility to chronic diseases to the yield of crops when environmentally stressed, understanding complex phenotypes lies at the heart of efforts to better the lives of humans and our planet's environment. Such phenotypes are difficult to characterize because of their large number of component traits; the wide variation in the components' values; and the need for large sample sizes to capture the phenotypic responses to the many influencing variables. The sample sizes needed for better characterization and resolution of the phenotypes multiply the logistical and methodological issues of the experiments, sharply limiting analyses.

High throughput imaging can increase sample size, improve quantitation, and better resolve visually related phenotypes, but introduces its own technical issues [2]. Chief among these is the ability to identify and segment the phenotype of interest from unaffected structures in the rest of the image. Image segmentation is an important task in many medical and biological applications [3–6].

Segmentation becomes harder as phenotypic complexity increases. One approach is to have people segment the images [7,8]. Manual segmentation may be feasible for small numbers of images with relatively few instances of the phenotypes, and for phenotypes that fall into very distinct classes. Usually, experts have time to segment only a few images, limiting the sample size and therefore the statistical power of
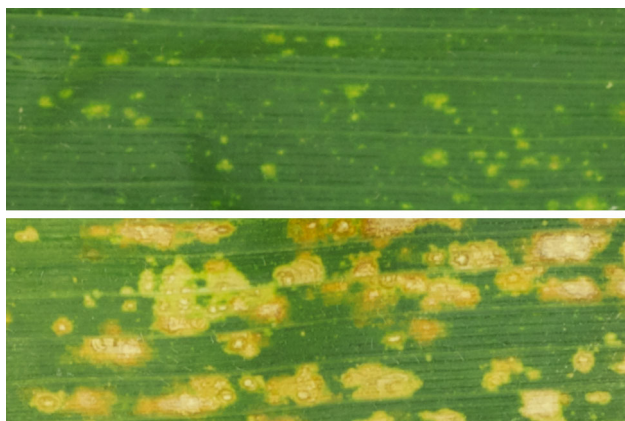
Fig. 1 Variations in lesion morphology as a function of mutant gene: *top Les2*, *bottom Les4*

an experiment; they disagree in their segmentations; and how well a segmenter performs can vary depending on the metric used to measure them [9–12]. Using manual segmentation in high throughput situations requires many more people, further complicating inter- and intra-observer consistency [8,13]. Thus, automating segmentation is highly desirable, despite the challenges of more numerous and more varied mathematical characteristics. Algorithms must be designed and tuned to fit the mathematical characteristics of the phenotype studied, while remaining general and flexible enough to include biologically salient variations.

The difficulties of segmenting complex phenotypes are amply illustrated by the disease lesion mimic mutants of *Z. mays* [14,15]. All the mutations produce macroscopic, irregularly shaped regions of chlorotic and necrotic tissue on otherwise healthy leaves, in the absence of exogenous pathogens. The size, shape, color, spatial distribution, sharpness and color of the lesion's boundary, internal morphology, and many other characteristics of the populations of lesions vary profoundly as a function of the particular lesion mimic mutation, the plant's genetic background, and the environmental conditions it experiences. Lesion plants with different underlying genetics may produce lesions with various numbers of distinct or overlapping features, but determining the relationships between lesion features requires accurate quantitation of the phenotypes. These phenotypes provide important clues about the mechanisms of lesion formation, a key part of the plant's defenses to disease, but their complexity makes segmentation challenging. Some variations are illustrated in Fig. 1.

1.2 Individual performance of standard algorithms on maize lesions

Segmentation is a common problem in medical and biological image analysis, and there are many approaches to

segmenting contiguous, visually distinct regions of an image. The problem of maize lesion mutants is so far untouched, but many similar problems have been tackled by the computer vision community. Macroscopic skin lesions, MRI scans of body regions, and cellular microscopy share several key features with maize lesions, though none of them exhibit the heterogeneity of disease lesion mimic mutants. Common approaches to these segmentation problems include color analysis and feature extraction [4], edge detection, watershed and thresholding, gradient vector flow and diffusion [3], active contour evolution, or some combination of these [5].

1.2.1 Color-based techniques: thresholding and feature extraction

If images are standardized sufficiently and phenotypes are homogeneous, it may be possible to apply a naïve approach such as thresholding. Barbedo recently developed an automatic lesion segmentation procedure for leaves of several species using color transformation and thresholding, which performed well for the images tested [16]. Notably, these were small leaves that are flat when laid in a scanner. Maize lesions, by comparison, show considerably more phenotypic variation; and our experimental throughput, several orders of magnitude larger than Barbedo's, results in higher technical variation between images. Maize leaves are much larger and their edges much longer than their central axes, producing significant rippling and highlights and shadows, necessitating a photographic technique. The combination of phenotypic and image complexity requires a more flexible and robust approach.

Feature extraction techniques consider transformations of the image matrix rather than raw image values, and can include color transformations, edge detection, and spectral decomposition, etc., [17]. Color transformations and edge detection have been used to extract macroscopic skin lesions, which like maize lesions can be of many sizes and shapes with considerable internal structure and differences in coloring. Images of skin lesions usually show one in isolation, so the color gradients are very sharp and reasonably symmetric about the lesion [4]. But images of maize lesions usually have hundreds or even thousands of lesions, complicating edge detection of what should be considered independent lesions. Moreover, maize lesions often have very diffuse boundaries, similar to MRI images of breast or brain tissue. Diffusion of the boundaries further flattens the gradients, exacerbating the problems caused by crowding.

1.2.2 Microscopic cell nuclei: gradient vector diffusion

Identifying and separating individual cells in an image has been a central goal of biological image segmentation
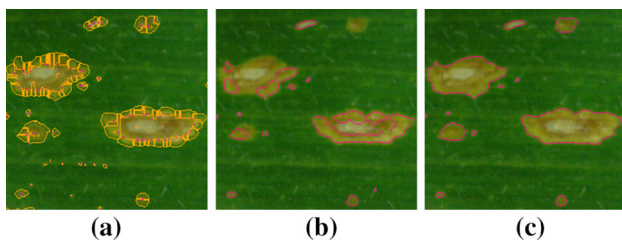
Fig. 2 Comparison of standalone algorithms. **a** Gradient vector diffusion followed by gradient tracking, **b** active contour refinement at a single resolution, **c** our approach. Segmentation boundaries shown in *pink* and *orange*

for over 50 years [18]. The task is difficult due to the size and shape differences between cells and the density of cells in an image. Cell nuclei and maize lesions often appear as localized regions of brightened pixels that can be sparsely to densely packed within the region of interest. The common approaches to segmenting cell nuclei include intensity thresholding, feature detection, morphological filtering, watershed transformation, deformable model fitting, and graph-cut approaches [3,19].

Gradient diffusion and flow tracking model the image as a deformable sheet and propagate strong gradients while suppressing weak ones. This approach was successful in segmenting the clustered cell nuclei of *C. elegans*, where the nuclei are very uniform in size, texture, and structure [3]. Not surprisingly, this approach worked well for detecting small and consistent lesions, even if they were densely packed. But it failed as the lesion phenotypes became more diverse (sizes varying over several orders of magnitude, variations in lesion texture and interior structure, and wider variation in lighting). For example, large lesions were severely oversegmented (Fig. 2a).

1.2.3 Active contours

Active contour algorithms, such as the Chan–Vese algorithm, evolve a segmentation boundary toward a local optimum, generally expressed as minimizing an energy function dependent on internal information (e.g., boundary curvature) and external information (e.g., image intensity or gradient) [20]. The final boundary is therefore a compromise between the locally optimal, "greedy" path around the object and smoothness constraints. These approaches provide smooth boundaries and accurate segmentations, but only if the algorithm is initialized sufficiently near to the object to be segmented. An example of mis-segmentation due to variable lesion size is shown in Fig. 2b. Smaller lesions are successfully identified, but larger lesions are oversegmented because the most salient features of smaller lesions struggle to properly characterize larger lesions.

The need to "seed" the active contour algorithm with an approximate boundary suggested the use of a wavelet-based multiresolution analysis (MRA) to identify putative lesions without prior knowledge of their size or shape [21]. MRA creates a series of image approximations, each representing the features of an image at different scales. The boundaries of these features are approximate, both because only a few scales are used in the MRA and because many objects may have irregularly crenelated polygonal boundaries.

1.3 Synopsis of our approach

Maize lesion segmentation requires an adaptive approach which is sensitive to small lesions and able to simultaneously detect and address large, diffuse, or internally structured lesions. Our method accomplishes this by combining several distinct techniques. First, wavelet decompositions at different scales identify putative lesions, independent of shape and size [21,22]. Then, the results of this multiresolution analysis are passed to a gradient vector diffusion algorithm for a preliminary estimate of the lesions' boundaries [3]. These two steps work well for detecting and approximating lesion boundaries, but for the final segmentation the Chan–Vese active contour segmentation algorithm is used [20]. The output is a mask that delimits each lesion in the image for subsequent quantitative characterization; example output of our algorithm is shown in Fig. 2c.

Our approach was validated using a collection of manually segmented images, which sample the range of maize lesion phenotypes and include some of the more difficult lesion types. We consider precision and recall measures to evaluate the correct identification of lesions, and the object-level consistency error (OCE) to measure overlap accuracy [23]. While our approach has been tuned to the problem of maize lesions, its strategy may be flexible enough for processing other, very heterogeneous images.

2 Materials

2.1 Plant materials

Fourteen different disease lesion mimic mutants were back-crossed into three different inbred maize lines. The number of back-crosses varied among the different line/mutant combinations. Plants were grown at the Missouri Maize Center, University of Missouri, Columbia, MO in the summer field seasons. Leaves were photographed shortly after the mutant plants had finished shedding pollen. At this stage, some combinations had lesions that had reached stasis and had well-defined boundaries, while other combinations had lesions that continued to expand and differentiate. In some cases, this latter class of phenotypes produce a diffuse chlorotic

zone surrounding one or more necrotic central lesions. As the chlorotic zones expand, intensify, and merge, it becomes difficult for humans to identify the original lesions.

2.2 Images for cascade development

A set of 18 images from the summer, 2012 field season was used for algorithm development. These images were chosen to maximize the number of distinct phenotypes, comparing among multiple lesion dimensions, while keeping the set small enough to permit detailed visual inspection of the algorithmic results.

Leaves were cut from plants, the cut ends immersed in ice water, rinsed, and air dried. All leaves were photographed within an hour after cutting, so they suffered no degradation. The leaf was attached to a field of blue cloth, creating a distinct leaf foreground and cloth background, allowing the leaf to be readily identified in the image. All images included an identifying tag with a unique plant ID, the identifier's barcode, and the number of the leaf. The images also included a Gretag–Macbeth (now XRite) color checker for possible color adjustment among images taken under differing lighting conditions.

Figure 3 demonstrates our setup for image capture. Images were taken with a Nikon D80 10.2 MP DSLR camera with an AF MICRO NIKKOR 60 mm lens and illuminated with fluorescent 5000 K lamps. Using a purpose-built jig, leaves were held parallel to the plane of the lens at a fixed distance and illuminated along their length and from above and below. Light was also reflected by aluminized bubble wrap enclosing the back, top, bottom, and part of the front of the apparatus, to further diffuse the angles of incident light. Most images included a third or more of the maize leaf; smaller leaves entirely filled the image.

2.3 Manual segmentation of test images

A set of 17 images, separate from the 18 used for algorithm development, was manually segmented and serves as a test set for the algorithm. These images came from the summer, 2011 field season, and were photographed differently from the 2012 images. In this case, the leaf was mounted horizontally, still parallel to the plane of the lens, and was illuminated by a combination of overhead fluorescent lights and two spotlights of 5000 K lamps, placed to the left and right of the leaf and angled onto it. Again, the identifying tag and color checker were included.

The images were chosen to maximize the variety of phenotypes as before. To prevent undue fatigue, the leaf area evaluated depended on the density of lesions, so that smaller areas of very crowded phenotypes were assessed. We aimed to have at least a hundred lesions in each segment, with

Fig. 3 Jig used for image capture. Views from front (*top*) and side (*bottom*)

approximately the same number of lesions for all segments. These parameters were evaluated by eye, however. With one exception, none of the observers had viewed lesions in planta before evaluating the image segments. Thus, they were at least as naïve as the cascade.

Each image segment was presented randomly to each observer, for a maximum of three presentations, each separated by at least one other image. The observer outlined the lesion boundary as accurately as possible using a stylus and touch screen monitor. Despite our best efforts, the pixels designated as the boundary were not always the one the observers intended. This boundary polygon was connected automatically between the first and last points, and its points recorded in a database. The observers varied widely in the number of image views they assessed: the range was 5–50 views, with a median of 17 views per person. The number of times each image was assessed also varied widely, ranging from 4 to 9 views by all observers.

3 Methods

3.1 Multiresolution analysis for lesion detection

Once the leaf is extracted from the raw image, the next step is to perform a multiresolution analysis, allowing for a segmentation approach that is agnostic to lesion size. Multiresolution analysis (MRA) approximates a function at different resolutions by means of wavelet decomposition. The wavelet function $\psi(x)$ is a function that can be dilated and translated to form a family of orthonormal basis functions spanning $f \in L^2(R)$ [21]. Intuitively, decomposing the function f can be thought of as finding how much each basis function contributes to f.

For analyzing signals that do not show strong periodicity, as is often the case for images, the wavelet transform is a considerable improvement over the Fourier transform, which has led to its adoption for the JPEG2000 image compression standard [22]. Images can be considered a discrete sampling of the continuous function f at even intervals, and so discrete approximations of the wavelet functions are scaled by a factor of 2^j and translated by k, where $(j, k) \in \mathbb{Z}$. Antonini et al. [24] define the discrete wavelet as finding the coefficients c which satisfy:

$$f = \sum c_{j,k}(f)\, \psi_{j,k} \tag{1}$$

$$c_{j,k}(f) = \langle f, \psi_{j,k} \rangle \tag{2}$$

$$\psi_{j,k}(x) = (2^{j/2}\psi(2^j x - k))_{(j,k)\in\mathbb{Z}} \tag{3}$$

where $\langle f, \psi_{j,k} \rangle$ represents the inner product (or "dot product") between f and $\psi_{j,k}$. In practice, this is equivalent to convolving f with the compact function ψ_j at all positions k.

The MRA approach developed by Mallat [21] uses two special wavelet functions, a difference function ψ and a scaling function ϕ. Transforming f with ϕ_j produces coefficients approximating f at resolution 2^j, while transforming f with ψ_j produces coefficients measuring the information difference between approximations at the scales 2^{j-1} and 2^j, effectively dividing f into a smoothed background component and a local detail component.

For images, a hierarchical algorithm has been developed which convolves f with low-pass and high-pass filters and downsamples by 2 at each step, decomposing an image into low-frequency and high-frequency components [24]. For our purposes, we are interested in the low-frequency image approximations generated via the Cohen–Daubechies–Feauveau 9/7 analysis coefficients [22, 25], which is a discretized version of the scaling function ϕ; these coefficients are effectively a low-pass filter (L) [24]. To ensure even sampling, the image must be expanded to be a multiple of 2^n prior to decomposition, where n is the number of times the image will be decomposed. The image is first convolved vertically with L and downsampled, followed by convolution horizontally and downsampling, creating a lower resolution approximation of the original image; this procedure can be repeated for as many scales as desired. For our purposes, we find four decompositions to be sufficient for detecting different-sized lesions. Figure 4b demonstrates the effect of the resolution reduction procedure for a small leaf section.

3.2 Gradient vector diffusion to find approximate position of lesions

Once the matrices of scaling coefficients have been computed, features across scales can be detected. While lesions can vary in size by several orders of magnitude, small lesions at a high resolution are visually very similar to large lesions at a low resolution, appearing as distinct foci of bright pixels. Measuring the gradient vectors of the images allows lesions to be detected as areas of gradient vector convergence, and by leveraging a multiresolution representation of the image, similar methods of detection can be used for lesions of all sizes. Due to the noise inherent in images, as well as irregularities in lesion shape and internal structure, we apply the gradient diffusion procedure of Li et al. [3], with minor alterations to adapt the method to two dimensions. This method was motivated by the high density of nuclei and diffuse boundary properties in microscopic embryo images of *C. elegans*.

The image is modeled as a deformable elastic sheet, where the gradient vector field $\mathbf{v}(x, y) = (u(x, y), v(x, y))$, with horizontal component $u(x, y)$ and vertical component $v(x, y)$, form solutions to the Navier–Stokes equation [26]:

$$\mu\nabla^2\mathbf{v} + (\lambda + \mu)\nabla\text{div}(\mathbf{v}) + (\nabla f - \mathbf{v}) = 0 \tag{4}$$

where ∇^2 is the Laplacian operator, ∇ is the gradient operator, div is the divergence operator, and ∇f is the original gradient. This equation is solved using a finite difference method, decoupling $\mathbf{v}$ and treating u and v as functions of time:

$$
\begin{aligned}
u_t(x, y, t) &= \mu\nabla^2 u(x, y, t) \\
&\quad + (\lambda + \mu)(\nabla\text{div}(\mathbf{v}(x, y, t)))_x \\
&\quad + ((\nabla f(x, y))_x - u(x, y, t)) \\
v_t(x, y, t) &= \mu\nabla^2 v(x, y, t) \\
&\quad + (\lambda + \mu)(\nabla\text{div}(\mathbf{v}(x, y, t)))_y \\
&\quad + ((\nabla f(x, y))_y - v(x, y, t))
\end{aligned} \tag{5}
$$

Consistent with the approach of [3], Δx, Δy, and Δt are taken to be 1, and with indices i and j corresponding to x and y, respectively:

$$u_t = u_{i,j}^{n+1} - u_{i,j}^n, \; v_t = v_{i,j}^{n+1} - v_{i,j}^n$$

$$\nabla^2 u = u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}$$

$$\nabla^2 v = v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j}$$

$$(\nabla \mathrm{div}(\mathbf{v}))_x = u_{i+1,j} + u_{i-1,j} - 2u_{i,j} + v_{i+1,j+1}$$
$$- v_{i,j+1} - v_{i+1,j} + v_{i,j}$$

$$(\nabla \mathrm{div}(\mathbf{v}))_y = v_{i,j+1} + v_{i,j-1} - 2v_{i,j}$$
$$+ u_{i+1,j+1} - u_{i+1,j} - u_{i,j+1} + u_{i,j}$$

$$(\nabla f)_x = 2f_{i+1,j} + f_{i+1,j-1} + f_{i+1,j+1}$$
$$- 2f_{i-1,j} - f_{i-1,j-1} - f_{i-1,j+1}$$

$$(\nabla f)_y = 2f_{i,j+1} + f_{i-1,j+1} + f_{i+1,j+1}$$
$$- 2f_{i,j-1} - f_{i-1,j-1} - f_{i+1,j-1} \qquad (6)$$

The equations for the horizontal gradient components $(\nabla f)_x$ and the vertical gradient components $(\nabla f)_y$ of the original gradient field are analogous to the Sobel filter, a common technique for approximating image gradients.

Our current implementation uses λ and μ values of 0.015, repeating the finite difference method for 45 iterations. These values were decided upon by manual tuning to produce suitable segmentations of the eighteen test images, and to ensure speed and stability. This gradient diffusion technique is applied to each matrix of scaling coefficients $c_{j,k}(f)$ separately. After the gradients have been diffused, the divergence is calculated from the horizontal and vertical gradient components.

To create a consensus image, the divergence values at each resolution are expanded to the native resolution (e.g., all pixels that are at half scale are repeated twice vertically and horizontally, pixels that are one quarter scale are repeated four times, etc.,), and the minimum at each pixel coordinate is found across the calculated divergence matrices. Gradient vector diffusion at multiple resolutions automatically enhances the gradient features of a lesion at the resolution that it is most readily detectable, and reexpansion to the native resolution allows for a similar active contours refinement to be calculated on lesions of all sizes. Lesions are detected as pixels with a divergence of less than −45. This consensus image provides an "initial guess" for implementation of an active contour algorithm, which provides the final segmentation (see Fig. 4d, e).

3.3 Active contours to refine lesion boundaries

Once a consensus mask has been calculated through MRA and gradient vector diffusion, an active contour algorithm can be initiated at each potential lesion. Active contour algorithms are a popular segmentation approach for optimizing object boundaries or separating internally homogenous regions of an image. The basic premise is to guess the object's boundary and to evolve it toward an optimal position accord-

ing to image information and constraints on mean curvature [18,20].

We utilize the Chan–Vese level-set method, which separates regions of an image u_0 by minimizing the variance of their internal values and the curvature of the boundary C separating them [20]. Defining an energy function F:

$$F(c_1, c_2, C) = \gamma \cdot \mathrm{Curvature}(C)$$
$$+ \int_{\mathrm{inside}(C)} |u_0(x, y) - c_1|^2 \mathrm{d}x\mathrm{d}y$$
$$+ \int_{\mathrm{outside}(C)} |u_0(x, y) - c_2|^2 \mathrm{d}x\mathrm{d}y \qquad (7)$$

where c_1 and c_2 are the mean values of the interior and exterior regions, respectively, and γ is a weight parameter which constrains the curvature; the larger γ, the more rigid the boundary. This is therefore the minimization problem: $\inf_{c_1, c_2, C} F(c_1, c_2, C)$.

The evolution of C is achieved using a level-set approach. C is defined implicitly as the zero level (or 'level-set') of the higher dimensional Lipschitz function Ω, or:

$$C = \{(x, y) | \Omega(x, y) = 0\} \qquad (8)$$

Evolving the curve C in time is achieved by solving the partial differential equation:

$$\frac{\partial \Omega}{\partial t} = \delta_\epsilon(\Omega)$$
$$\times \left[\mathrm{div}\left(\frac{\nabla\Omega}{|\nabla\Omega|}\right) - (u_0 - c_1)^2 + (u_0 - c_2)^2 \right] = 0 \qquad (9)$$
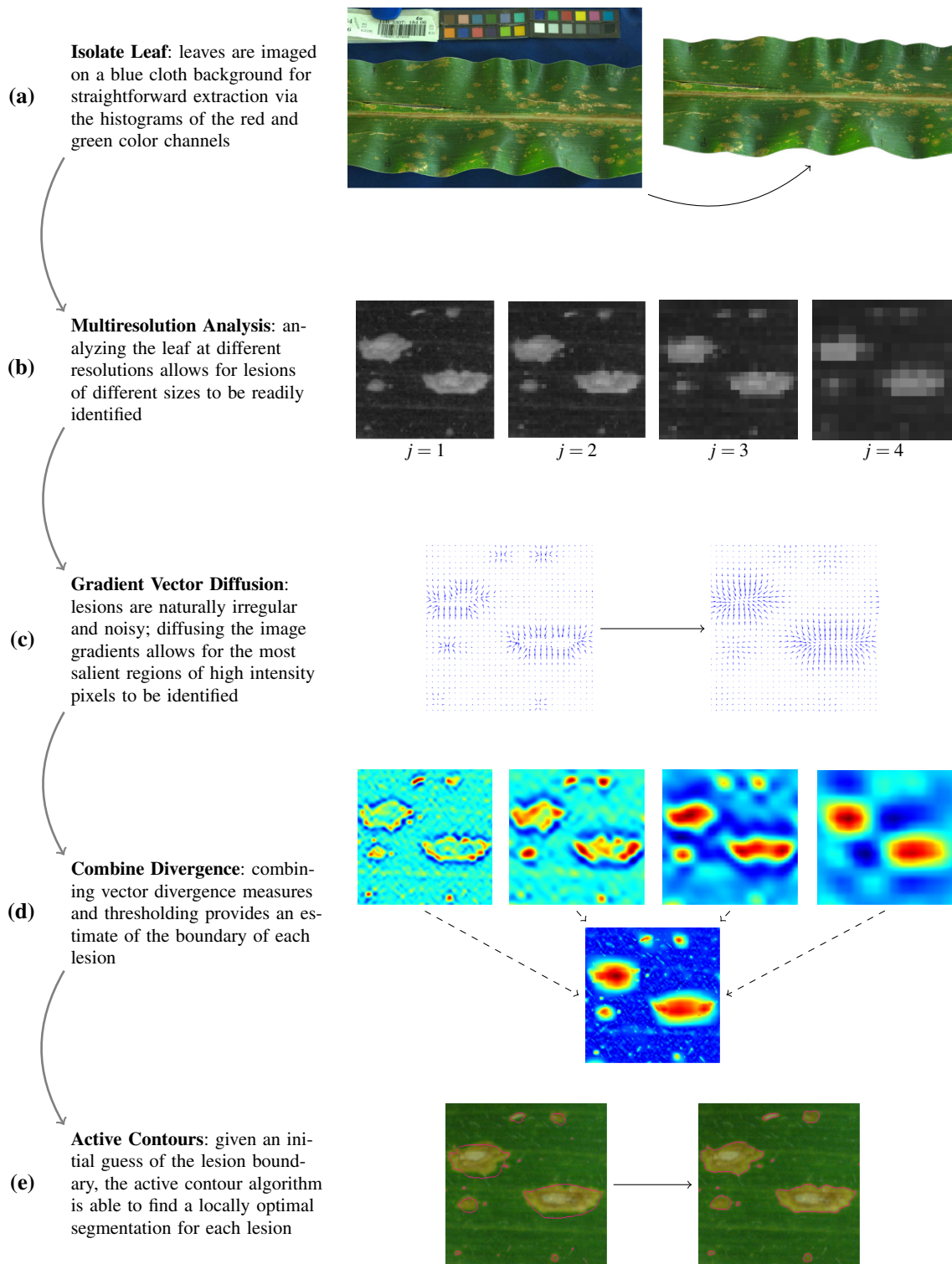
$$\Omega(0, x, y) = \Omega_0(x, y), \qquad (10)$$

$$\frac{\delta_\epsilon(\Omega)}{|\nabla\Omega|} \frac{\partial \Omega}{\partial \overrightarrow{n}} = 0 \qquad (11)$$

where $\mathrm{div}(\frac{\nabla\Omega}{|\nabla\Omega|})$ is the curvature of Ω, $\Omega_0(x, y)$ is the initial boundary, and $\delta(\Omega)$ is the Dirac delta function. A discretized Matlab implementation of the algorithm was used.[1]

3.4 Validation

A segmented image is considered as a binary image where lesion pixels are labeled 1 and nonlesion pixels are labeled 0. An individual lesion is a set of contiguous lesion-labeled pixels. Methods exist to create a ground truth image from the

---

[1] The code can be found at: http://www.mathworks.com/-matlabcentral/-fileexchange/-19567-active-contour-segmentation.

(a) Isolate Leaf: leaves are imaged on a blue cloth background for straightforward extraction via the histograms of the red and green color channels

(b) Multiresolution Analysis: analyzing the leaf at different resolutions allows for lesions of different sizes to be readily identified

$j = 1$ $\quad$ $j = 2$ $\quad$ $j = 3$ $\quad$ $j = 4$

(c) Gradient Vector Diffusion: lesions are naturally irregular and noisy; diffusing the image gradients allows for the most salient regions of high intensity pixels to be identified

(d) Combine Divergence: combining vector divergence measures and thresholding provides an estimate of the boundary of each lesion

(e) Active Contours: given an initial guess of the lesion boundary, the active contour algorithm is able to find a locally optimal segmentation for each lesion

Fig. 4 An overview of our approach. The output of each algorithm is passed to the next in the cascade, producing a final segmentation of each putative lesion in an image. The boundaries for initial and final segmentation are shown in *pink (e)*

input of multiple users by voting schemes [27] or Expectation Maximization [28], but these approaches are tailored to situations where one or a handful of objects in an image are segmented, not the hundreds to thousands present in *Z. mays* lesion images. This, along with the low number of manual segmentations available, resulted in highly irregular and

user 2   user 5   algorithm
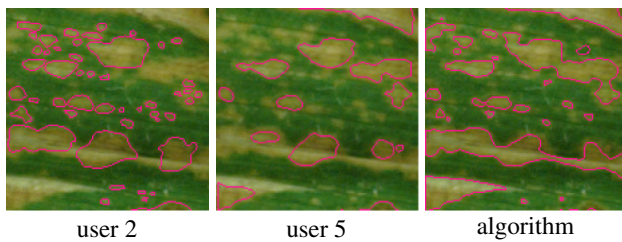
Fig. 5 Disagreement between two human segmenters and the algorithm

unnatural ground truth images. We instead opted to perform a "round robin" validation, wherein each segmenter (including the algorithm) was compared against all other users. Rather than evaluating the segmenter in absolute terms, this procedure determines whether a segmenter is in agreement with the other segmenters; our measure of success is therefore whether the algorithm statistics lie within the range of inter-human variation.

When comparing two segmentations of a lesion in separate images of the same leaf, there is a distinct set of relationships the putative lesions may share:

1. *Exact match* (*True Positive*): for every pixel belonging to a lesion in one image, there is an equivalent pixel in the second image, and vice versa.
2. *Near match* (*True Positive*): the centroid of one and only one lesion in the first image falls within the convex hull of a lesion in the second image, or vice versa.
3. *Oversegmentation* (*True Positive*): the centroids of more than one lesion in the first image fall within the convex hull of a lesion in the second image.
4. *Undersegmentation* (*True Positive*): the centroids of more than one lesion in the second image fall within the convex hull of a lesion in the first image.
5. *False positive* (*False Positive*): the centroid of a lesion in the first image does not fall within the convex hull of a lesion in the second image.
6. *False negative* (*False Negative*): no lesion in the first image has a centroid which falls within the convex hull of a lesion in the second image.

Relationships 1–4 are all examples of *true positives* (*TP*), wherein a lesion has been correctly identified, with or without some error.

Figure 5 demonstrates some of the variability that exists between users, and illustrates several of the mis-segmentation cases described above. In this case, it is apparent that user 5 is considerably more conservative than user 2 or the algorithm, while the algorithm tends to undersegment (relative to the users) and user 2 tends to oversegment (relative to user 5 and the algorithm). These examples also highlight the difficulty

in deciding whether a segmented region constitutes a "true" lesion.

To evaluate a segmentation, we consider the precision, recall, and object-level consistency error (OCE) of the segmented image against a ground truth [23]. When evaluating a given segmentation, the ground truth image is taken to be all pixels labeled as "lesion" pixels by the other participants, which includes the automated algorithm when evaluating manual segmentations. We considered excluding the results of the algorithm for purposes of validation, but this unfairly decreases the human metrics while increasing the algorithm's metrics. In this way, a lesion is considered valid if it was detected by more than one observer.

Object-level consistency error (OCE) is a metric that quantifies the overlap of a segmented and ground truth image at the object level [23]. This makes it preferable for evaluating segmentations of multiple objects. The commonly used Dice and Jaccard coefficients are appropriate for evaluating the overlap of single objects, but topological disagreements (oversegmentation and undersegmentation) are not properly penalized as all labeled pixels are considered to be the same object. OCE overcomes this by measuring the mismatch between pairs of objects and weighting the mismatch by the magnitude of the mismatch relative to the overall image. OCE values range from 0 for perfect agreement to 1 for no agreement between segmentations.

Consider a ground truth image $I_g = \{A_1, A_2, \ldots, A_M\}$, where A_j is the jth fragment in I_g, and a segmented image $I_s = \{B_1, B_2, \ldots, B_N\}$, where B_i is the ith fragment in I_s. A partial error measure $E_{g,s}(I_g, I_s)$ is defined as:

$$E_{g,s}(I_g, I_s) = \sum_{j=1}^{M} \left[1 - \sum_{i=1}^{N} \frac{|A_j \bigcap B_i|}{|A_j \bigcup B_i|} \times W_{ji} \right] W_j, \tag{12}$$

$$W_{ji} = \frac{\bar{\delta}(|A_j \bigcap B_i|)|B_i|}{\sum_{k=1}^{N} \bar{\delta}(|A_j \bigcap B_k|)|B_k|}, \quad W_j = \frac{|A_j|}{\sum_{l=1}^{M} |A_l|}, \tag{13}$$

where $\delta(x)$ is the Dirac delta function and $\bar{\delta}(x) = 1 - \delta(x)$. The OCE is thereby defined as:

$$\text{OCE}(I_g, I_s) = \min(E_{g,s}, E_{s,g}). \tag{14}$$

Precision (P) and recall (R) are measured as:

$$P = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FP}},$$
$$R = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FN}} \tag{15}$$

For a pair comprised of a ground truth and test image, centroids for all possible lesions are calculated and true positive, false positive, and false negative values are calculated as described above. Using these values, the precision and recall

is easily calculated. These measures effectively evaluate the presence or absence of a lesion in a test image, while the OCE score evaluates how well a particular segmentation captures the total lesioned tissue.

4 Results
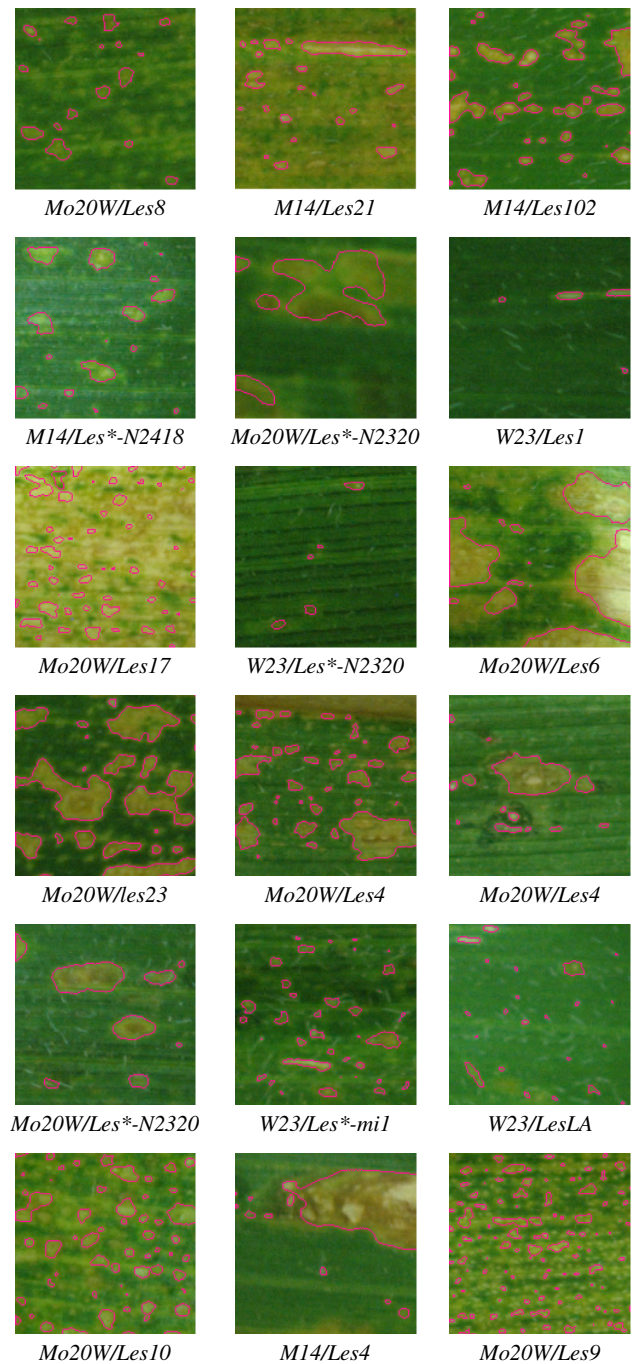
4.1 Variation in lesion phenotypes

Figure 6 demonstrates the results of our algorithmic cascade on the 18 development images. These images sample from the broad phenotypic variation present among maize lesion mimic mutants, and highlight variation in phenotypic characteristics and the difficulties in measuring them. Lesions can vary in size from only a handful of pixels, as seen in the *W23/Les*-N2320* mutants, to several thousand, shown by the *Mo20W/Les6* mutants. Lesion boundaries also vary considerably in sharpness and regularity, as demonstrated by the defined but amorphous lesions of *Mo20W/les23* and the diffuse, ovoid lesions of *Mo20W/Les8*. Proximity of lesions is yet another key though confounding phenotypic trait; a handful of lesions may be scattered throughout the leaf (*W23/Les*-N2320*) or they may be densely packed or touching (*Mo20W/Les9*). Finally, the background healthy tissue itself presents challenges to lesion detection due to variable lighting conditions, leaf damage, and non-lesion biological structures such as leaf hairs and the leaf midrib.

4.2 Illustration of cascade

The principal stages of our approach are demonstrated by Fig. 4. Multiresolution analysis decomposes the maize lesion image into representations of decreasing resolution, while the distinct lesion foci are detected by gradient vector convergence. It is apparent that small lesions are readily identified at a higher resolution and larger lesions at lower image resolutions. After combining the gradient convergence measures across the multiresolution image representation, the active contours algorithm can be initiated at the native image resolution for final boundary refinement. Parameters and the respective values used are shown in Table 1.

4.3 Validation

Four individuals hand segmented the 17 test images, which included some of the most diverse lesion phenotypes. Only one individual segmented the full set of extracts, with the others completing some subset of the extracts. Partially segmented extracts were not considered valid. Hand segmentations vary considerably between individuals, and to a lesser degree within individuals, so inter- and intra-researcher variation are both important to consider.


Fig. 6 Examples of maize lesion phenotypes. The approximately horizontal striations are leaf veins and the short, slender, mostly vertical features are leaf hairs. The *pink* edges around the lesions denote the segmentation results

Figure 7 and Table 2 summarize the results for hand-segmented and machine-segmented images. For both humans and the cascade, the segmentation results vary, a result of the complexity of *Z. mays* lesion images as well as the subjectivity in defining what constitutes a lesion. Because in each case the ground truth was considered to be those lesions identified

Table 1 Parameters used across the various stages of the algorithm and their values

Parameter	Value	Description
n_{MRA}	4	Number of decomposition stages used for MRA
μ and λ	0.015	Control speed of gradient vector diffusion
n_{GVD}	45	Number of iterations to run the gradient vector diffusion
γ	0.8	Rigidity of active contour boundary; higher values result in more ovoid segmentations
n_{AC}	25	Number of iterations to run the active contour boundary refinement
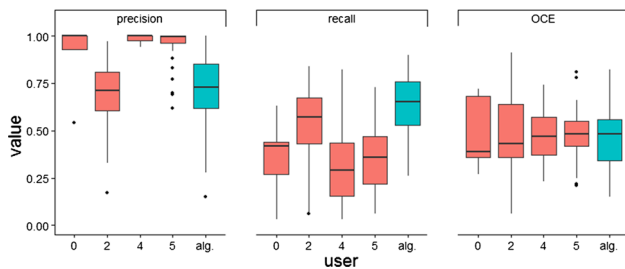
Fig. 7 *Box and whisker plots* for precision (*left*), recall (*center*), and OCE (*right*). Human segmenters shown in *red* and the algorithm in *blue*. A precision and recall of 1 is ideal, while an OCE of 0 is ideal

by the other participants, low precision and recall can result from disagreement. It is worth noting that in cases where the algorithm performed poorly, there is also disagreement amongst users, implying that these are particularly difficult and perhaps subjective cases.

Multivariate analysis of variance (MANOVA) was conducted for each pair of users (0, 2, 4, 5, and the cascade) with the dependent variables of precision, recall, and OCE. The cascade demonstrated lower precision than users 0, 4, and 5, but only comparisons with users 4 and 5 were statistically significant ($p < 0.01$). The cascade's precision was comparable to that of user 2. The cascade demonstrated higher recall values than all users, with comparisons against users 0, 4, and 5 being significant ($p < 0.01$). Finally, when comparing OCE values for all users, there was no detectable difference among

users. This demonstrates a somewhat expected loss of precision due to automation, but an increase in overall recall with no significant change in OCE.

5 Discussion

Zea mays lesion mimic mutants are an excellent example of a distinct, measureable complex phenotype. By comparing the phenotypic similarities and differences between mutants in a quantitative manner, the underlying processes producing these phenotypes can be better understood. Quantitating complex phenotypes, however, is itself a challenge, and the phenotypic and technical variation of maize lesion mimic mutants makes segmentation a particularly difficult problem. Our initial attempts at background subtraction had modest success, but were unable to overcome the range of local intensities and biological structures present in our images. Standard edge detection techniques such as the Sobel filter were successful at segmenting highly defined lesions, but diffuse boundaries and densely packed lesions made edge detection infeasible. Use of active contour algorithms was promising, but as has been previously noted [4,5,18,20,29], the results are highly dependent on initialization. The gradient vector diffusion technique of Li et al. [3] was very successful at identifying small lesions, which are visually similar to *C. elegans* embryos, but larger lesions with variable internal structure led to high degrees of oversegmentation. Multiresolution analysis overcame this barrier, allowing for gradient vector diffusion at multiple scales to detect lesions of many shapes and sizes. This permitted a final refinement of their boundaries by the active contours algorithm.

By combining common image processing techniques, including thresholding, multiresolution analysis, gradient vector diffusion, and active contours, we have created a novel approach which is able to segment lesioned *Z. mays* leaves on par with human segmenters. Due to the complexity of *Z. mays* leaf images and the difficulty of *Z. mays* lesion segmentation, results between and within segmenters can vary widely, with precision as low as 15 % and recall as low as 2 % for some of the more difficult examples. These results are a product of both the technical difficulties in segment-

Table 2 Comparison of human and algorithm segmentation performance

User	Total	Uniq	$\overline{P}$	P SD	$\overline{R}$	R SD	$\overline{OCE}$	OCE SD
0	5	5	0.89	0.20	0.36	0.22	0.48	0.20
2	31	15	0.69	0.17	0.52	0.23	0.48	0.20
4	11	9	0.98	0.02	0.31	0.23	0.47	0.16
5	50	17	0.96	0.09	0.35	0.18	0.49	0.12
Alg.	17	15	0.69	0.22	0.64	0.18	0.46	0.19

Included are the number of total images segmented, number of unique images segmented, and the mean and standard deviation of each user's precision, recall, and OCE measures

ing natural images as well as the subjective interpretation of what constitutes a lesion. In cases where the cascade struggled, there is higher disagreement among users, such as the *W23/Les1*, *Mo20W/Les17*, and *W23/Les*-2320* mutants. Nonetheless, the cascade demonstrates a high recall rate and was comparable to humans with respect to OCE; that is, our algorithm captured most of the lesions identified by human segmenters. In the case of precision, the cascade did under-perform some of the human segmenters, though it was comparable to user 2, who segmented the most images. For our phenotyping purposes, OCE and recall are the most pertinent measures: we will at least capture the lesions humans do, for most of their area. Our approach demonstrates results comparable to human segmenters and the potential to vastly improve throughput, an essential requirement for our images, which are comprised of hundreds of images with thousands of lesions each.

To improve results, the simplest solution may be in the field. By photographing leaves earlier, lesions will be more distinct and better separated, while the leaf will show less damage, simplifying analyses. Future improvements on the cascade may include a more structured training of parameters, inclusion of additional features such as color transformations, or a more principled method for combining gradient information across levels in the multiresolution analysis. Other performance measures would likely produce different values. While this method has been tailored to the *Z. mays* lesion problem, we believe its general approach can be generalized to other problems in computer vision.

References

1. United States Department of Agriculture, F. A. S: World Agricultural Production. United States Department of Agriculture, Foreign Agricultural Service, Washington (2015)

2. Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S.: Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. **126**, 867–887 (2013)

3. Li, G., Liu, T., Tarokh, A., Nie, J., Guo, L., Mara, A., Holley, S., Wong, S. T.: 3D cell nuclei segmentation based on gradient flow tracking. BMC Cell Biol. **8** (2007)

4. Cavalcanti, P.G., Scharcanski, J., Persia, L.E.D., Milone, D.H.: An ICA-based method for the segmentation of pigmented skin lesions in macroscopic lesions. In: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Boston, pp. 5993–5996 (2011)

5. Yuan, Y., Giger, M.L., Li, H., Suzuki, K., Sennett, C.: A dual-stage method for lesions segmentation on digital mammograms. Med. Phys. **34**, 4180–4193 (2007)

6. Zhou, H., Schaefer, G., Celebi, M.E., Lin, F., Liu, T.: Gradient vector flow with mean shift for skin lesion segmentation. Comput. Med. Imaging Graph. **35**, 121–127 (2011)

7. Lingutla, N.T., Preece, J., Todorovic, S., Cooper, L., Moore, L., Jaiswal, P.: AISO: annotation of image segments with ontologies. J. Biomed. Sem. **5**, 50 (2014)

8. Plaza, S.M., Scheffer, L.K., Saunders, M.: minimizing manual image segmentation turn-around time for neuronal reconstruction by embracing uncertainty. PLoS One **7**, e44448 (2012)

9. Iii, S.A., et al.: Assessment of radiologist performance in the detection of lung nodules. Dependence on the definition of "truth". Acad. Radiol. **16**, 28–39 (2009)

10. Siena, S., Zinoveva, O., Raicu, D., Furst, J., III, S.A.: A shape-dependent variability metric for evaluating panel segmentations with a case study on LIDC data. In: Karssemeijer, N., Summers, R.M. (eds.) Medical Imaging 2010: Computer-Aided Diagnosis. Proceedings of SPIE, Bellingham. Society For Optical Engineering, vol. 7624 (2010)

11. Restif, C.: Revisiting the evaluation of segmentation results: introducing confidence maps. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007. Proceedings of the 10th International Conference, Brisbane, 29 October–2 November 2007, Part II, Berlin. Lecture Notes in Computer Science, vol. 4792, pp. 588–595. Springer, New York (2007)

12. Hripcsak, G., Wilcox, A.: Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. J. Am. Med. Inform. Assoc. **9**, 1–15 (2002)

13. Pearl, R.: The personal equation in breeding experiments involving certain characters of maize. Biol. Bull. **21**, 339–366 (1911)

14. Neuffer, M.G., Calvert, O.H.: Dominant disease lesion mimics in maize. J. Hered. **66**, 265–270 (1975)

15. Neuffer, M.G., Coe Jr. E.H., Wessler, S.R.: Mutants of Maize. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1997)

16. Barbedo, J.G.A.: An automatic method to detect and measure leaf disease symptoms using digital image processing. Plant Dis. **98**, 1709–1716 (2014)

17. Nixon, M.S., Aguado, A.S.: Feature Extraction and Image Processing, 2nd edn. Elsevier, New York (2008)

18. Meijering, E.: Cell segmentation: 50 years down the road. IEEE Sig. Proc. Mag. **29**, 140–145 (2012)

19. Dimopoulos, S., Mayer, C.E., Rudolf, F., Stelling, J.: Accurate cell segmentation in microscopy images using membrane patterns. Bioinformatics **30**, 2644–2651 (2014)

20. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Proc. **10**, 266–277 (2001)

21. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)

22. Skodras, A., Christopoulos, C., Ebrahimi, T.: The jpeg 2000 still image compression standard. IEEE Sig. Proc. Mag. **26**, 36–58 (2001)

23. Polak, M., Zhang, H., Pi, M.: An evaluation metric for image segmentation of multiple objects. Imaging Vis. Comput. **27**, 1223–1227 (2009)

24. Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I.: Image coding using wavelet transform. IEEE Trans. Image Proc. **1**, 205–220 (1992)

25. Cohen, A., Daubechies, I., Feauveau, J.-C.: Biorthogonal bases of compactly supported wavelets. Commun. Pure Appl. Math. **65**, 485–560 (1990)

26. Davatzikos, C., Prince, J.L., Bryan, R.N.: Image registration based on boundary mapping. IEEE Trans. Med. Imaging **15**, 112–115 (1996)

27. Artaechevarria, X., Muñoz-Barrutia, A., de Solórzano, C.O.: Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imaging **28**, 1266–1277 (2009)

28. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging **23**, 903–921 (2004)

29. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. IEEE Trans. Image Proc. **17**, 2029–2039 (2008)

Derek Kelly received his B.S. in Anthropology and Biological Sciences from the University of Missouri, Columbia in 2013. He is currently a PhD student in the Genomics and Computational Biology Graduate Group at the University of Pennsylvania under Dr. Sarah Tishkoff. His interests include human evolutionary genomics, regulation of gene expression, and statistical genetics.

Avimanyou Vatsa is pursuing Ph.D and received MS (Computer Science) and minor in Statistics from University Of Missouri, Columbia in 2015. He obtained his M.Tech (Computer Engineering) with Hons. from Shobhit University, Meerut and B.Tech (I.T.) from V.B.S. Purvanchal University, Jaunpur, (U.P), India, in 2009 and 2001, respectively. He is working as teaching and research assistant at University of Missouri – Columbia and worked as an assistant professor for ten years in several engineering colleges and universities in India. He had been member of academic and administrative bodies. During his teaching he had been coordinating many technical fests and national conference at college and university level. He has worked as software engineer in software industry. He is a member and reviewer on the editorial board of several international and national journals in networks and security field. His area of research includes Low- and High-Dimensional Data (Big Data), Image Processing, MANET (Mobile Ad Hoc Network), Computer Network and Network Security.

Wade Mayham received his B.S. in Computer Science from the University of Missouri, Columbia in 2015. Currently he is working as a freelance software developer and will pursue Masters of business administration in 2016. His interests include artificial intelligence, computer networks, and virus research and defense.

Toni Kazic is a computational biologist and an associate professor in the Department of Computer Science, University of Missouri. She is also a member of that University's Interdisciplinary Plant Group, Informatics Institute, Missouri Maize Center, and Genetics Area Program. Her research is focussed on understanding complex phenotypes and using these to infer biological networks. A geneticist by training, she is finishing her ninth year working in her corn field and thinking about the computational and mathematical challenges maize poses.