

How to use level set methods to accurately find boundaries of cells in biomedical images? Evaluation of six methods paired with automated and crowdsourced initial contours.

Danna Gurari, Diane Theriault, Mehrnoosh Sameki, and Margrit Betke

Boston University Computer Science Department

Abstract. Level set methods are popular tools for automatically collecting accurate outlines of biological objects in biomedical images and videos. The two main challenges with successfully applying these methods are identifying which among the many options will work well for a particular image set and choosing an initial contour that the method will successfully evolve to the desired final boundary. Little is known about the comparative performance resulting from different initial-contour method pairings. To examine the practical impact of this concern for biomedical applications, we compared six freely available level set methods with 12 different initializations on fluorescence and phase contrast images showing cells. The studies revealed that none of the initial-contour method pairings performed well for phase contrast images. These results motivated us to suggest using internet workers to draw estimates of cell boundaries. These crowdsourced boundaries can then serve as initial boundaries for a level set method to produce results closer to the true boundaries. We found that pairing segmentation algorithms with crowdsourced initial-contours yields over 50 percent points better performance than the other pairings for phase contrast images. Our results yield recommendations for initial-contour method pairings based on image modality and highlight the benefit of engaging non-expert internet workers to successfully leverage level set methods for biomedical images. We invite extensions of this work by sharing all code (<http://anonymous>).

1 Introduction

Level set methods are widely used to automate finding accurate boundaries of biological structures in biomedical images and videos. In general, level set methods deform an initial contour to a final contour that separates image foreground from background so that some method-specific image partition condition is enforced. While new energy functionals controlling how to partition an image continue to be proposed to address the spectrum of possible image conditions, the continued development and wide-spread sharing of new options is making it difficult for both non-experts and experts to know which method to use when. A further challenge for applying such methods is knowing which type of initial contour

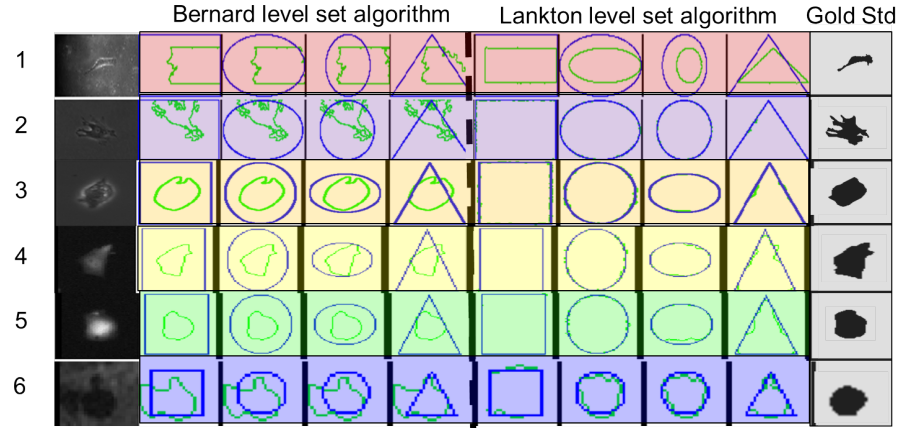


Fig. 1. Representative segmentation results exemplifying that a trial-and-error effort to find a contour initialization may or may not lead to successful use of a level set method. Raw images (column 1), shown for a biological object from each dataset in the image library, were processed with the “Bernard level set algorithm” [8] (columns 2–5), the “Lankton level set algorithm” [6] (columns 6–9), and manually (column 10). Blue lines show initial contours, green lines the final segmentation.

will be sufficiently close to the desired boundary since they often produce locally optimal segmentation results which may not match the desired globally optimal segmentations. As a result, a common question asked by individuals trying to apply level set methods is “which method with which initial contour will produce the desired boundary in my images?”

To address this question, we evaluate level set algorithms that currently have a potential widespread practical impact due to their inclusion in freely-shared bioimage analysis systems [1–3]. *Geodesic active contours* evolve the initial contour to end up in regions with strong edges (high contrast) [4]. *Active contours without edges* evolve the initial contour to try to separate the image into two homogeneous regions [5]. Both *Lankton region-based active contours* [6] and the *Li level set algorithm* [9] evolve the initial contour by using the local neighborhood statistics for each pixel in order to adjust how to separate the sub-region into two homogeneous regions. The *Shi approximation method* computationally speeds up the evolution process by replacing slow real-valued calculations with faster integer-based calculations [7]. The method by *Bernard et al.* uses a linear combination of B-spline basis functions for process speedup [8]. Currently, there is no work comparing these freely shared algorithms on biomedical image sets.

Domain experts planning to use level set methods on their biomedical images encounter an additional overhead of creating initial contours. With freely available image analysis software [2, 3], they create initial contours, clicking on images to create simple geometric shapes, points connected into polygons/splines, or free-hand tracings, and then typically wait for seconds or minutes per image for

Table 1. List of algorithms analyzed in comparison study, their inclusion in biomedical image analysis toolboxes, and initialization methods reported in the initial publications.

Tool	Software Options	Published Curve Initializations
Geodesic Active Contours [4]	Fiji[3], ITK[1], Creaseg [2]	Rectangle, Circle
Chan Vese level set method [5]	Creaseg [2]	Square, Circle
Lankton level set method [6]	Creaseg [2]	Rectangle, Square, Circle, Ellipse, Merged Rectangles
Li level set method [9]	Creaseg [2]	Square, Circle, Ellipse, Triangle
Shi level set method [7]	Creaseg [2]	Circle, Merged Circles
Bernard level set method [8]	Creaseg [2]	Square, Circle, Ellipse, Merged Circles

the input contour to finish evolving to a segmentation [2]. While recent as well as foundational publications reported that simple initial contour shapes such as bounding squares, rectangles, circles, ellipses, and triangles led to accurate segmentations [4–9], other recent publications suggested these initial contours can be insufficient. As an example, specialized contour initialization methods have been proposed for phase contrast image sets [10, 11] to avoid common curve evolution failures. It can be faster for domain experts to manually trace boundaries themselves than to run an algorithm and possibly risk running it repeatedly until finding an initial contour that returns an accurate segmentation (**Fig. 1**).

To provide practical guidance for obtaining accurate segmentations with level set methods, we conducted an extensive comparison study of six publicly-available level set methods paired with popular initial-contour shapes. We analyze when to use which method and how to use the methods effectively on fluorescence and phase contrast images. To further minimize the overhead for domain experts of creating initial contours for their biomedical images, we also propose to use crowdsourcing to create them. Finally, to facilitate extensions of this study, we publicly share all code (<http://www.cs.bu.edu/~betke/BiomedicalImageSegmentation>).

2 Methods

To find a contour initialization method that works well in general, we designed and implemented a system that supports trial-and-error analyses by applying all combinations of chosen initial-contour shapes and level-set algorithms to all image sets in an image library. A user runs the system with one command and can configure the system to apply a collection of level set algorithms initial-contour pairings to a collection of image sets.

2.1 Segmentation System

Images are processed sequentially. For each image, the system applies the segmentation method with the associated curve initialization method. Different segmentation and curve initialization configurations with different parameter settings can be applied for different image sets (described below). Next, the segmentation result is post-processed by filling holes and keeping only the largest object. Finally, the system saves the resulting binary segmentation as an image.

Segmentation Modules. Each of six publicly available implementations [2] of level set algorithms are wrapped into a single module that the user may use interchangeably in the system: geodesic active contours [4], Chan Vese level set method [5], Lankton level set method [6], Shi level set method [7], Bernard level sets [8], and Li level set method [9] (**Table 1**). Each segmentation module is decoupled from the initial contour by being linked to an *Initial Contour Module* option that, at run time, creates an initial contour.

Initial Contour Modules. Each initial contour module shares the same interface. Given an input image, it returns a binary mask of the same dimensions. The system supports four initial contour methods the user may use interchangeably: rectangle, ellipse, circle, and a triangle. To create the contour, the *rectangle* module uses the boundary of the rectangle drawn by removing n pixels from all sides of the image region, the *ellipse* module uses the boundary of an ellipse drawn to span the image region downsized by n pixels on all sides, the *circle* module uses the boundary of a circle drawn at the center of the image region with a radius of half of the smallest image region dimension minus n pixels, and the *triangle* module uses the boundary of a triangle drawn to span the image region downsized by n pixels on all sides using two corners of the bounding box on the bottom image side and the midpoint between both corners on the top image side. The user can configure parameter n to control the size of the shape.

2.2 Crowdsourced Initial Contour Module

When basic geometric shapes are insufficient, as reported for phase contrast image sets [10, 11], a user can instead use a crowdsourced initial contour to create an estimate of the object boundary that is closer to the true boundary. We incorporate the publicly available on-line annotation software, LabelMe [13], into the system usage pipeline to collect the initial contours. LabelMe supports tracing the boundary of objects by taking as input points specified by the internet worker and connecting them sequentially with straight lines to produce a 2D segmentation. The internet worker finishes by clicking on the starting point. If a mistake is made, the internet worker can delete and redraw the object boundary.

Annotators are recruited from the Amazon Mechanical Turk online internet marketplace. Each drawing task for each image is posted to Amazon Mechanical Turk as a human intelligence task (HIT) paired with a price to be paid upon completion of the task. A worker can review the HIT before accepting the job. LabelMe comes with scripts to post HITs to Mechanical Turk and record the annotation results, making the annotation collection process simple.

To overcome concerns about trusting annotations from a single annotator, whether due to weaker skills or even malicious motivations, we incorporated into the pipeline the *Probability Maps (p-map)* algorithm so that the user can combine multiple crowdsourced segmentations for each image. This algorithm takes as input N segmentations and outputs a single segmentation where a pixel is labeled as foreground when at least M of the crowdsourced segmentations label it as foreground and background otherwise. Finally, the segmentation result is post-processed to fill holes and keep only the largest object.

3 Experiments

We conducted three studies using the proposed system on biomedical images to examine which among the six freely-available level set methods yield the most accurate segmentations for various image modalities, what is the impact of contour initialization on segmentation quality, and whether paid crowdsourced workers can be leveraged to expedite successful use of level set methods for biomedical images.

Image Library. We analyzed the algorithms on a total of 271 images from five datasets shared in a freely-available image library [12]. From this collection, 152 are phase contrast images that show 35 rat smooth muscles cells, 70 rabbit smooth muscle cells, and 47 fibroblasts. The remaining 119 are fluorescence images that show 65 Lu melanoma cells and 58 WM993 melanoma cells. The datasets represent cells that range in size from less than 1,000 pixels to over 32,000 pixels. There is exactly one dominant object in each image. Image resolutions are equal to the bounding box of the expert-drawn segmentation of the object in the image grown by 50 percentage of those dimensions, in order to maintain the original region proportions.

Segmentation Evaluation Methodology. We computed scores that indicate how closely algorithm-generated segmentations match gold standard segmentations provided with the image library. We quantitatively analyzed each algorithm for all images using the *Jaccard similarity index*, a standard evaluation metric. This metric computes the number of pixels common to both the algorithm-generated and gold-standard regions that is also in the combination of their regions (i.e., $\frac{|A \cap B|}{|A \cup B|}$ where A and B represent the set of pixels in the gold standard and algorithm-generated segmentations respectively).

Study 1: Impact of Initial Contour. We applied our system to all images in the library to collect segmentations using all six algorithms. We did this 12 times to analyze the impact of the shape and size of the initial contour by setting $n = 5, 15,$ and 25 pixels for the rectangle, ellipse, circle, and triangle.

Study 2: Comparison of Level Set Methods. We applied our system to all images in the library to collect segmentations using all six algorithms. We set the initial contour to the gold standard segmentation mask. We also compared algorithms using as the initial contour a circle with $n = 15$ since we found in Study 2 this pair performed well for both phase contrast and fluorescence images.

Study 3: Analysis of Using Crowdsourced Initial Contours. We applied our system to all images in the library to collect segmentations using all six algorithms paired with the initial contours created by crowdsourced workers. To create the initial contours, we collected five crowdsourced annotations per image and fused the segmentations into a single binary mask with the *p-map* algorithm setting $M = 3$ (i.e., a pixel is part of the object only if at least three annotators included it as part of the object). To minimize concerns about work quality, we only accepted workers that had previously completed at least 100 HITs and received at least a 92% approval rating. Workers receive a five step set of instructions detailing how to submit a HIT followed by pictures exemplifying

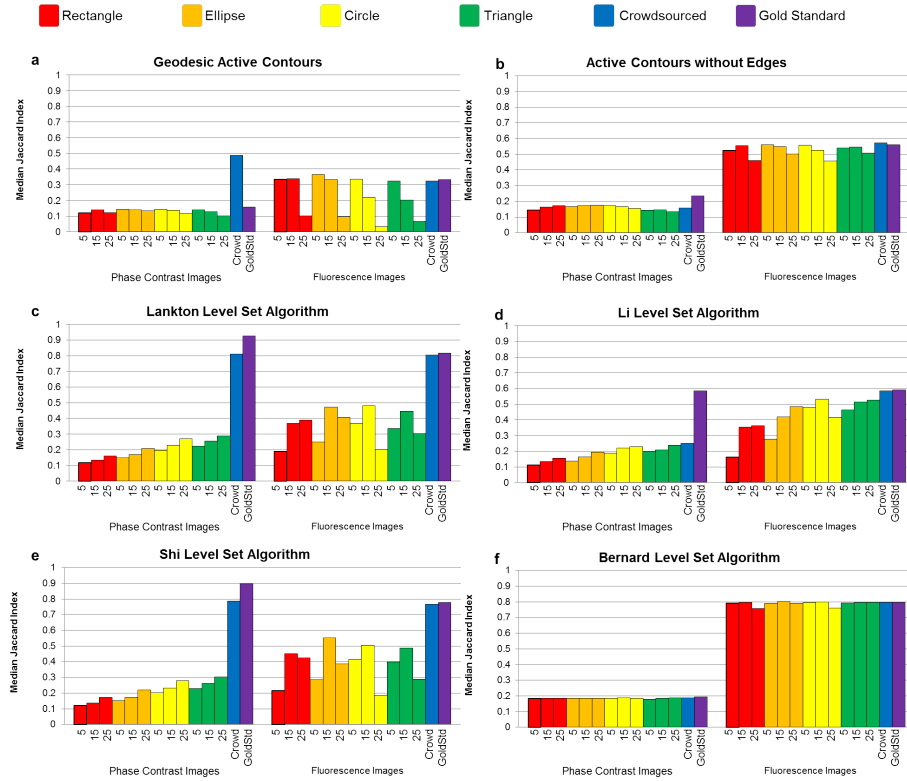


Fig. 2. Results showing the performance of six level set methods paired with 14 unique contour initializations. Each plot shows the median *Jaccard similarity index* score for all phase contrast images and fluorescence images independently when using as the initial contour four geometric shapes at three different sizes, the crowdsourced segmentation boundary, and the gold standard boundary.

good and bad annotations. All submitted HITs were accepted and workers were paid \$0.02 for completing each drawing task.

4 Results

Study 1: Impact of Initial Contour. We found that the shape and size of the initial contour can impact algorithm performance for both phase contrast and fluorescence images (**Figs. 1, 2**). For *fluorescence* images, we found a noticeable difference in algorithm performance based on initial contour shape and size for all but the Bernard level set algorithm. For initial contour shape, the ellipse and circle led to the best performance for most of the algorithms. For initial contour size, for most algorithms, the medium-sized bounding region led to the best performance ($n = 15$). For *phase contrast* images, we found a slight difference in

algorithm performance based on initial contour shape and size for the Lankton, Li, and Shi level set based algorithms. For initial contour shape, the ellipse, circle, and triangle each led to better performance for different algorithms and only the rectangle consistently led to inferior or equal performance. For initial contour size, for most algorithms, a smaller bounding region led to the best performance regardless of initial contour shape (larger n value).

Study 2: Comparison of Level Set Methods. We found, when the initial contours were set to the boundaries of the gold standard segmentations, that only the Lankton, Shi, and Bernard level set algorithms performed well (**Fig. 2**). For the *phase contrast* images, the Lankton and Shi level set algorithms yielded the best performance. For the *fluorescence* images, the Bernard and Shi level set algorithms yielded the best performance. We found that the top-performing algorithms resulted in scores over 10% higher for *phase contrast* images than for *fluorescence* images.

We found, when comparing algorithms using the circle as an initial contour, that different algorithms performed well for different image modalities (**Fig. 2**). For *phase contrast* images, we found that the Lankton and Shi level set algorithms led to the best performance. For *fluorescence* images, we found that the Bernard level set algorithms led to the best performance. We found that the top-performing algorithms resulted in scores over 50 percent points higher for *fluorescence* images than for *phase contrast* images.

Study 3: Analysis of Using Crowdsourced Initial Contours. We found that pairing segmentation algorithms with our proposed initial contour method yielded over 50 percent points performance improvement for *phase contrast* images and comparable performance for *fluorescence* images in comparison to the top-performing algorithm initial-contour pairings found in study 2.

We found the mean elapsed time from the five annotation collection batches in Mechanical Turk was 32 hours and 35 minutes from posting the drawing tasks until submission of all the 271 completed drawing tasks. 40 unique workers completed all 1,355 drawing tasks and, on average, each worker took a mean time of 46 seconds to complete a single drawing task.

5 Discussion

We analyzed freely-available level set algorithms to report about algorithms with immediate wide-spread relevance. We were surprised that most of the algorithms yielded low-quality segmentations when evolving the gold standard boundary to a final boundary. We infer from these results that the algorithm energy functionals most closely matching the inherent properties of the studied image modalities and biological structures are Lankton and Shi level set algorithms for the phase contrast images and Bernard and Shi level set algorithms for the fluorescence images. We infer from our results that, when applying these algorithms in practice, all the studied initial contour shapes and sizes yield high quality segmentations when paired with the Bernard level set algorithm, while the other three level set algorithms should be paired with an initial contour that closely hugs the true

object boundary. Lastly, we infer from our results that non-expert paid crowdsourced workers can replace domain expert involvement to create initial contours for biomedical images.

6 Conclusions

Greater wide-spread use of algorithms to successfully collect high quality segmentation annotations relies on knowing which algorithm to choose and then how best to use it. We found that only a few of the studied freely-available level set algorithms are designed with assumptions that are well-suited for the studied phase contrast and fluorescence images of cells. For the well-suited algorithms, we found that one simple detail, the initial contour, can trigger over a 50 percent point improvement for phase contrast images. Finally, our results show the potential of using paid crowdsourced workers without domain-specific training to reliably and inexpensively replace domain experts in creating initial contours that are needed to use these algorithms effectively. Our study may be a start point towards a larger community effort to empower those applying level set methods to make an informed choice about which algorithm to use, how to use it effectively, and how to replace their efforts with non-experts. We encourage the reader to leverage our system so that the number of comparison studies of this sort can grow to address a wider range of biomedical problems important to the research community. Future work will explore how to more efficiently utilize crowdsourcing by analyzing the reliability of crowdsourced workers and what number of annotations are necessary. Possible future research directions also include running the study on a larger image set and quantitatively analyzing the causes in images that influence the successes and failures of the different algorithms and initial contours.

Acknowledgments

The authors gratefully acknowledge funding from the National Science Foundation (IIS-1421943, IIS-0910908).

References

1. L. Ibanez et al. The ITK software guide. 2003.
2. T. Dietenbeck et al. Creaseg: A free software for the evaluation of image segmentation algorithms based on level-set. *ICIP*, pages 665–668, 2010.
3. J. Schindelin et al. Fiji: an open-source platform for biological-image analysis *Nature Methods*, 9(7):676–682, 2012.
4. V. Caselles, R. Kimmel, and G. Sapiro. Variational b-spline level-set: A linear filtering approach for fast deformable model evolution. *IEEE Trans Imag Process*, 22(1):62–79, 1997.
5. T. Chan and L. Vese. Active contours without edges. *IEEE Trans Imag Process*, 10(2):266–277, 2001.

6. S. Lankton and A. Tannenbaum. Localizing region-based active contours. *IEEE Trans Imag Process*, 17(11):2029–2039, 2008.
7. Y. Shi and W. C. Karl. A real-time algorithm for the approximation of level-set based curve evolution. *IEEE Trans Imag Process*, 17(5):645–656, 2008.
8. O. Bernard et al. Variational b-spline level-set: A linear filtering approach for fast deformable model evolution. *IEEE Trans Image Process*, 18(6):1179–1191, 2010.
9. C. Li et al. Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans Imag Process*, 17(10):1940–1949, 2008.
10. K. Li et al. Cell population tracking and lineage construction with spatiotemporal context. *Medical image analysis*, 12(5):546–566, 2008.
11. I. Ersoy et al. Cell spreading analysis with directed edge profile-guided level set active contours. *MICCAI*, pages 376–383, 2008.
12. D. Gurari et al. How to Collect High Quality Segmentations: Use Human or Computer Drawn Object Boundaries? *Department of Computer Science Technical Report BUCS-2013-20*, Boston University, December 2013.
13. B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1–3): 157–173, 2008.