Combining the Efforts of Humans and Computers to Reliably and Inexpensively Extract High Quality Spatio-Temporal Information About Cells in Videos

Danna Gurari
dgurari@bu.edu
Computer Science Department at Boston University
PhD Research Plan, Doctoral Consortium, HCOMP 2014, AAAI Conference on Human Computation & Crowdsourcing

Abstract

Microscopy imaging has become a common and important tool for the field of biology. In practice, e.g., in clinical or biotech workflows, images are often annotated by hand, because practitioners question the quality of automated results. However, more recently, practitioners are drawn to computer-produced annotations because they can be collected efficiently and inexpensively. I propose to build a system to extract spatial-temporal annotations of populations observed in videos that intelligently distributes the annotation work between crowdsourced workers and computers in order to simplify a challenge faced by biology researchers to extract high quality spatio-temporal information from their videos in an inexpensive, reliable, and scalable solution. I will evaluate the system on a freely-available test image set and also demonstrate its performance for two research-based studies. This work will highlight how human and computer resources can be leveraged together to consistently create high-quality annotations for a wide range of bioimage analysis tasks.

Introduction

Researchers are studying cell behavior with the goal of gaining an understanding of fundamental biological processes and using this knowledge in turn to diagnose diseases and engineer biomaterials. Many of these studies begin with the researchers collecting videos showing how populations of cells behave when exposed to a range of conditions. Then, the researchers extract statistics about single-cell behavior to discover the relationship between cell shape and function (Wada et al. 2011) and how the environment influences cell appearance (Yeung et al. 2005). Such analyses can depend on detecting subtle or rare appearance and behavior variations. The key challenge addressed in this paper is how to establish a generalized laboratory protocol for accurately demarcating the boundaries of cells (segmentation) and following cells over time (tracking). Amplifying this difficulty is that cells may undergo significant appearance variation and motion variation in short periods of time (**Fig. 1**).

Commonly, domain experts annotate their videos of cell populations using annotation software such as ImageJ (Rasband) to draw the segmentations and associate cell candidates in each image with cell tracks from previous images in the video. The key motivating assumption for this approach is that human annotators trained on how to interpret cells observed in images collected using different biomedical image acquisition systems can distinguish between true object boundaries and image noise/artifacts and so draw highly accurate boundaries and generate reliable tracks. However, leaving the annotation efforts in the hands of expert annotators is often prohibitive. For example, I estimated for a study of our collaborator on cancer cell progression that a single researcher annotating 15 videos with 200 images per video showing a population of 50 cells would take over 31 forty-hour work weeks, assuming the researcher took 30 seconds to draw the outline of each cell and 5 seconds to link associated cells between two frames. Moreover, the cost would be $31,575 assuming the researcher earned $25 per hour. This approach is not only time-consuming and expensive, but also error-prone, potentially biased, and not scalable.

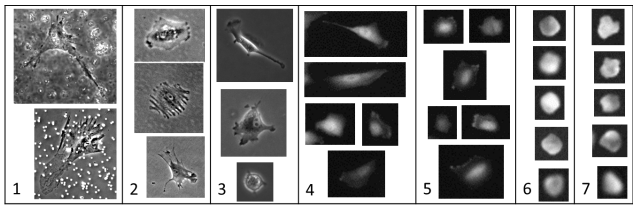

Figure 1: Examples of cells from seven phase contrast and fluorescent image sets. Methods that work well in general to detect, segment, and track cells are challenged to accurately handle the large variations in cell intensity in cell intensity, size, and shape, ill-defined boundaries separating cells from the background, image noise, and cluttered backgrounds. (To improve the visualization of cells in this figure, we manually enhanced the contrast of the images.) Reprinted from National Science Foundation Grant IIS-1421943 with permission from Margrit Betke.

In response to the deterrences from having experts use the brute force method to manually annotate videos, developers have been building computer vision systems to expedite or replace expert efforts (Rizk et al. 2014; Amat et al. 2014). However such systems are limited in design for fluorescent image sets. Furthermore, preliminary experiments reveal that these systems fail for fluorescent videos of our collaborators which show highly deformable migrating cells. Even when the are well separated so that there is no visual

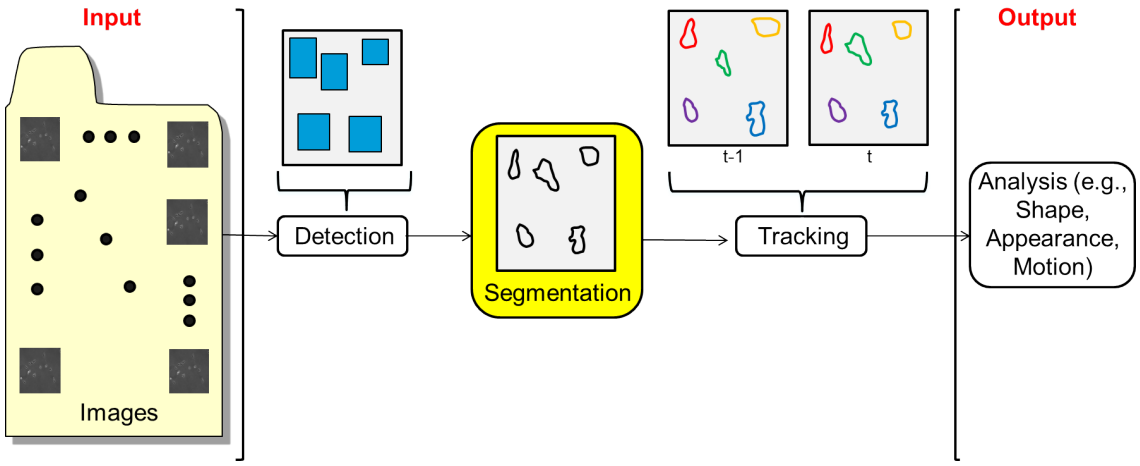

Figure 2: Overview of modules in the proposed system framework (Detection, Segmentation, and Tracking), within the context of producing biological analyses from an input video. The system will process all images sequentially. For each image, it will first be decomposed into regions containing objects, then segmentation methods will be applied to each region resulting in a binary mask identifying the silhouette of every object in the image, and finally a data association module will link objects detected in the current image to previously tracked objects. Toy examples are shown to exemplify the output of each module. Among the analysis tasks of detection, segmentation, and tracking of cells, segmentation is the most time-consuming task for human annotators and so will be the initial focus of the proposed work in order to provide the greatest opportunity for impact.

interaction or occlusion, it can be hard to track them all due to poor detection and delineation of cell boundaries.

I propose to eliminate expert involvement by abandoning the search for a single method that works well in general and, instead, developing a hybrid methodology to utilize crowdsourced humans and computers together to solve segmentation and tracking challenges. The idea to combine the efforts of crowdsourced humans and computers to expedite image-based biological research has recently been explored to reconstruct 3D neural circuits (Helmstaedter et al. 2013) by coupling weakly trained undergraduate students with segmentation algorithms. I instead will explore how to couple online paid crowdsourced workers with algorithms. The key contributions of this work will be:

- A methodology, web-based implementation, and crowdsourcing experiments that inform how to utilize the annotation efforts of crowdsourced workers and computer algorithms to create object boundaries that are of comparable quality to segmentations created by experts.
- A tracking system that integrates the proposed methodology to intelligently distribute the annotation work between crowdsourced workers and computers to yield higher-quality segmentation and tracking performance.
- Multiple experiments applying the proposed methodology to a variety of biomedical videos that demonstrates the usefulness of this system for domain experts to use as a laboratory tool.

This work simplifies a challenge faced by biology researchers to extract high quality spatio-temporal information from their videos in an inexpensive, reliable, and scalable solution. The proposed solution can have a significant and broad impact for biomedical research.

Proposed Research

I propose to develop an on-line video annotation system that can accurately find the boundaries of and track objects that exhibit significant variability in appearance (**Fig. 2**). I hypothesize that tracking and segmentation accuracy will improve if we intelligently adapt the segmentation method we apply based on image context in order to apply the method that will yield the highest quality segmentation among multiple options for each object. Furthermore, I hypothesize that we can eliminate expert involvement while extracting high quality spatio-temporal information consistently and inexpensively by coupling computers with crowdsourced workers in the system pipeline of object detection, segmentation, and tracking. Lastly, I will propose a performance evaluation method that connects the goals of the human computation, computer vision, and biology communities in order to provide "statistically significant" results with respect to a comprehensive set of performance metrics that address common application objectives.

To intelligently distribute the segmentation annotation work so that each method among a collection of popular segmentation approaches is applied only when it will perform the best, I will evaluate and compare trained experts, crowdsourced non-experts, and popular automated cell segmentation algorithms (Warfield, Zou, and Wells. 2004; He et al. 2008) on a generalized image library. I will compile an image library that includes datasets that capture the common as well as rare image artifacts and object appearances for various object types observed with multiple imaging modalities in order to identify the methods that are expected to perform well for the spectrum of possible unseen image and object appearances. I will analyze the strengths and weaknesses of these methods with respect to accuracy, consistency, and

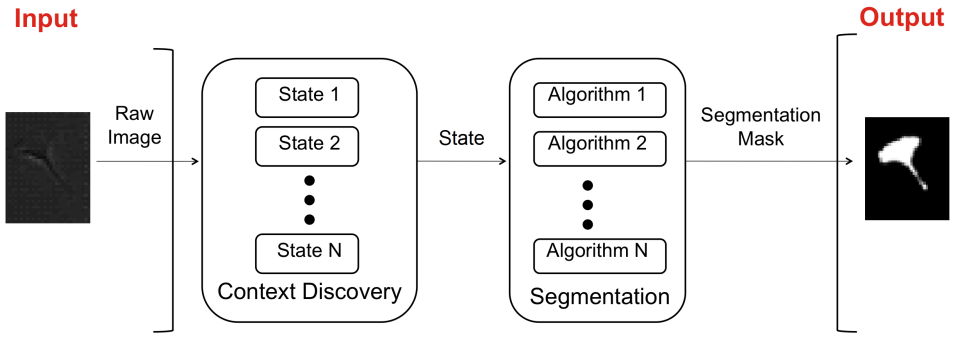

Figure 3: Example of a system that applies the online phase of the proposed machine learning based segmentation prediction framework to select the best segmentation algorithm among a set of options. For a cell detection region in an image, the best segmentation method is predicted for the image context and then applied. Reprinted from Gurari et al (Gurari, Theriault, and Betke. 2014).

cost and assess what factors make them succeed and fail. I will use these findings to train a machine learning system that determines automatically which types of cells to annotate with or without human involvement and, when not using human involvement, which algorithm. The proposed prediction system will first extract features that characterize the images of cells used in the benchmarking experiment that revealed whether human involvement was needed for analysis. Using the calculated features as a vector representation of each image and the associated class labels human needed or not needed, a binary SVM classifier will be trained in an offline phase. In the online phase, for a new image, the classification system will compute the feature vector representation of the image and apply the classifier to determine whether to recruit a human worker or use a segmentation algorithm to annotate the image. A similar approach will be applied to predict which algorithm will perform best among numerous algorithm options (**Fig. 3**).

I will then develop a tracking system embedded with the proposed segmentation system to achieve high tracking accuracy (**Fig. 2**). The system will process all images sequentially. I will benchmark crowdsourced non-experts using crowdsourcing and popular algorithms for the detection problems and examine how best to couple algorithms with crowdsourced workers to address these problems. Then, the system will apply the optimal segmentation method for each image subregion, resulting in a binary mask identifying the silhouette of every object in the image. Finally, the system will establish the correspondence between previously tracked objects and objects detected in the current image to perform tracking. I will examine whether I can build off of existing cell tracking algorithms, typically based on centroid or contour tracking, to automatically perform frame-by-frame data association or leverage freely-available online annotation systems (Vondrick, Patterson, and Ramanan 2013). I will initially evaluate both the Hungarian algorithm and level set-based algorithms (He et al. 2008). To further minimize human involvement, I will exploit the redundancy of information in videos by using humans in the loop for a subset of images and using interpolation to establish optimal segmentations for the remainder of images.

A final important component of this work is establishing an evaluation methodology that addresses the goals of the multiple interested communities. The current standard performance evaluation methodology in the computer vision community is to compare algorithm generated annotations with a single set of manual annotations using a comprehensive set of performance metrics that penalize for the spectrum of possible segmentation mistakes and then to report algorithm performance using a single average score. In contrast, human computation and biology studies commonly adopt significance testing methods applied to performance metrics that address the experimental objectives in order to assess whether observed differences are negligible. I hypothesize that considering evaluation criteria pertinent to applications combined with significance testing may change the interpretation commonly found in the computer vision community such that algorithms may be reveal suitable replacements for manual annotation by demonstrating they have negligible differences from manual annotations when using less rigorous evaluation criteria.

Proposed Experiments

Experiments will focus on an application area in biotechnology, live-cell imaging, where the proposed solutions can have a significant and broad impact.

Image Sets. Initial testing will validate detection, segmentation, and tracking methods for a broad range of biomedical applications to examine generalized performance. I will employ image sets and associated ground-truth compiled at the Broad Institute (Ljosa, Sokolnicki, and Carpenter 2012) for testing, which includes 22 datasets of a combination of fluorescent, bright field, and differential interference contrast images capturing biological organisms. The number of objects to analyze per image range from one to hundreds.

Next, I will evaluate the methods and system on two collections of videos collected by our collaborators. The first collection contains 10 fluorescent videos containing 200 images each that show the migration behaviors of populations of approximately 200 cancer cells when exposed to various combinations of compounds. The video analysis goal is to use cell movement behavior to assess how various com-

pounds influence the spread of cancer. I will also evaluate the methods and system on 10 phase contrast videos which show migration behaviors of a population of fibroblasts on various substrates that mimic different physiological environments. The video analysis goal is to use cell boundary and movement behavior to aid biomaterial research.

Evaluation Metrics. I will initially leverage widely-accepted performance metrics commonly used in the computer vision community to analyze the methods, such as count to evaluate detection performance, Jaccard similarity index to evaluate segmentation performance, and the Multiple Object Tracking Accuracy (MOTA) to evaluate tracking performance. I will also use significance testing to compare the quantitative results of different methods and the ground truth and determine when differences are negligible. In particular, a one-way analysis of variance (ANOVA), followed by a multiple comparison test with Tukey's honestly significant difference criterion, will be conducted to perform pairwise comparisons of average annotation performance. Statistically significant results will be deemed those where the significance level p is less than 0.05.

Segmentation System Performance. To analyze the machine learning segmentation system, I will partition the segmentation data into three sets and then conduct three experiments. In each experiment, the classifier will be trained on one third of the data to learn the optimal segmentation algorithm for each cell state (offline phase) and then tested on the remaining two thirds of the data, with the optimal algorithm based on the state of each cell (online phase). I will then compute scores indicating the quality of the prediction segmentation system which determines whether to recruit a human worker or which segmentation algorithm to annotate the image. I will use significance testing to compare the segmentation results to the standalone segmentation methods.

Tracking Performance. I will compute scores indicating the performance of the proposed system. I will use significance testing to compare the results to two state of art fully-automated systems (Rizk et al. 2014; Amat et al. 2014).

Results

I expect the following results: (1) segmentation methods combined in a working system that together provide higher quality boundaries; (2) fewer tracking errors; and (3) improved models for describing cell shape and motion. Also, I expect the system to effectively handle differences in imaging conditions such as imaging modality, object type, microscope magnification, or environmental conditions.

Previous work shows that a single segmentation source may not be optimal for all image scenarios and that a hybrid approach of linking segmentation algorithms with domain-expert-provided classifications to find the boundaries of cells can yield higher quality segmentations than nine popular freely-available standalone algorithms (Gurari, Theriault, and Betke. 2014). Moreover, previous work highlights 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 segmentation algorithms effectively (Gurari et al. 2014). Preliminary experiments reveal how trained experts, crowdsourced non-experts, and algorithms compare when annotating 305 objects coming from six datasets that include phase contrast, fluorescence, and magnetic resonance images. A total of 6,148 segmentations created by 10 experts, 58 crowdsourced workers, and six algorithms were analyzed. We found that the quality of annotations of internet workers follows closely that of experts (**Figs. 4, 5**). We also found that combining the segmentations created by crowdsourced workers and algorithms yielded improved segmentation results over stand-alone non-experts and algorithms respectively. Finally, preliminary work highlights the limitations of existing segmentation performance evaluation methods and motivates a need for an approach that incorporates statistical significance analysis widely used in human computation (Gurari et al. 2013).

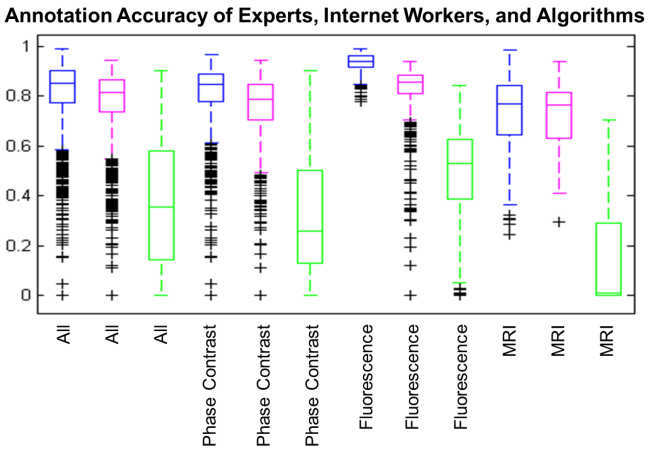

Figure 4: Jaccard similarity index scores for segmentations created by experts (red), non-experts (green), and algorithms (blue), averaged over all data, and data of each of the three image modalities. For each annotation source, the central mark of the box denotes the median score and the box edges the 25th and 75th percentiles scores. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually (black). Surprisingly, the quality of annotations of internet workers follows closely that of experts, and algorithms perform on average much worse. Automated segmentation techniques struggle particularly with interpreting the outlines of cells in phase contrast images and aortas in MRIs. The best annotations were collected for fluorescence images, followed by phase contrast images, and then m for all three annotation sources. Reprinted from Gurari et al (Gurari et al. 2015).

Conclusion

The novelty of the proposed work is designing a system that successfully combines the strengths of computers and crowdsourced humans to segment and track highly deformable objects in a computer vision system. The goal is to make concrete recommendations regarding how to leverage human involvement effectively for the system pipeline of object detection, segmentation, and tracking. I hypoth-

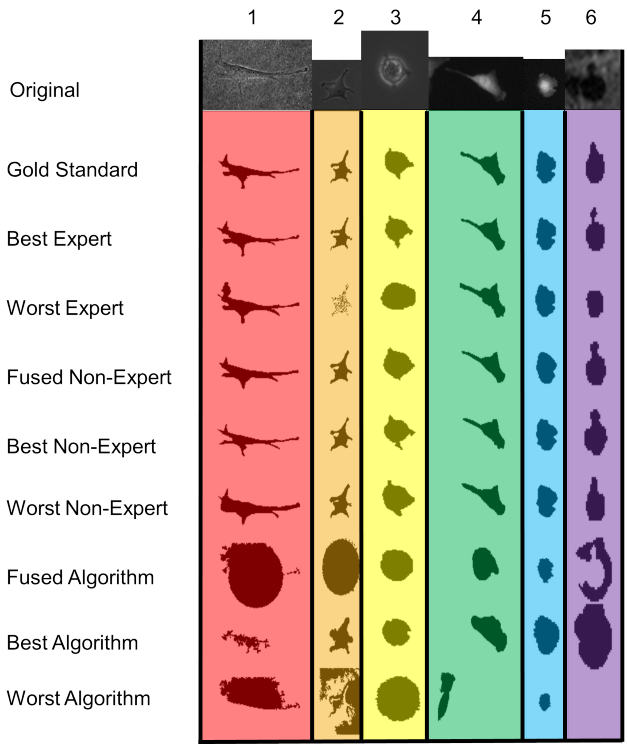

Figure 5: Representative segmentation results. Raw images (row 1), followed by fused, highest-scoring, and lowest-scoring segmentations for experts (rows 2–4), non-experts (rows 5–7), and algorithms are shown for a biological structure from each dataset in the image library (cols. 1–6). Reprinted from Gurari et al (Gurari et al. 2015).

esize both human computation and computer vision algorithms, which will be involved for demarcated subtasks in the system design, will benefit from each other. Success in this work will encourage future interdisciplinary collaborations by highlighting how such collaborations can improve computer vision systems and how computer vision systems can help human computation. Also, providing segmentation and tracking tools with capabilities comparable to or better than experts will significantly accelerate biological discoveries - it will accelerate progress for current researchers as well as encourage other researchers, previously deterred, to exploit research using image analysis.

Acknowledgments

The authors gratefully acknowledge funding from the National Science Foundation (IIS-1421943, IIS-0910908) and thank Chentian Zhang, Matthew Walker, and Joyce Y. Wong for the images and annotations they provide.

References

Amat, F.; Lemon, W.; Mossing, D. P.; McDole, K.; Wan, Y.; Branson, K.; Myers, E. W.; and Keller, P. J. 2014. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. In *Nature Methods*.

Gurari, D.; Kim, S.; Yang, E.; Isenberg, B.; Pham, T.; Purwada, A.; Solski, P.; Walker, M.; Wong, J. Y.; and Betke, M. 2013. SAGE: An approach and implementation empowering quick and reliable quantitative analysis of segmentation quality. In *Proceedings of the IEEE Workshop on Applications in Computer Vision (WACV)*. 7 pp.

Gurari, D.; Theriault, D.; Sameki, M.; and Betke., M. 2014. 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. In *Interactive Medical Image Computation Workshop (IMIC)*.

Gurari, D.; Theriault, D.; Sameki, M.; Isenberg, B.; Pham, T. A.; Purwada, A.; Solski, P.; Walker, M.; Zhang, C.; Wong, J. Y.; and Betke, M. 2015. How to collect segmentations for biomedical images? a benchmark evaluating the performance of experts, crowdsourced non-experts, and algorithms. In *IEEE Winter Conference on Applications in Computer Vision (WACV)*. 8 pp.

Gurari, D.; Theriault, D.; and Betke., M. 2014. Informed segmentation: A framework for using context to select an algorithm and a case study using humans in the loop. In *Interactive Medical Image Computation Workshop (IMIC)*.

He, L.; Peng, Z.; Everding, B.; Wang, X.; Han, C. Y.; Weiss, K. L.; and Wee, W. G. 2008. A comparative study of deformable contour methods on medical image segmentation. *Image Vision Comput* 26(2):141–163.

Helmstaedter, M.; Briggman, K. L.; Turaga, S. C.; Jain, V.; Seung, H. S.; and Denk, W. 2013. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* 500:168–174.

Ljosa, V.; Sokolnicki, K. L.; and Carpenter, A. E. 2012. Annotated high-throughput microscopy image sets for validation. *Nature Methods* 9(7):637.

Rasband, W. ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij.

Rizk, A.; Paul, G.; Incardona, P.; Bugarski, M.; Mansouri, M.; Niemann, A.; Ziegler, U.; Berger, P.; and Sbalzarini, I. F. 2014. Segmentation and quantification of subcellular structures in fluorescence microscopy images using squassh. *Nature Methods* 9(3):586–596.

Vondrick, C.; Patterson, D.; and Ramanan, D. 2013. Efficiently scaling up crowdsourced video annotation. *International Journal of Computer Vision* 101,(1):184–204.

Wada, K.; Itoga, K.; Okano, T.; Yonemura, S.; and Sasaki, H. 2011. Hippo pathway regulation by cell morphology and stress fibers. *Development* 138:3907–3914.

Warfield, S. K.; Zou, K. H.; and Wells., W. M. 2004. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Med Imaging* 23(7):903–921.

Yeung, T.; Georges, P. C.; Flanagan, L. A.; Marg, B.; Ortiz, M.; Funaki, M.; Zahir, N.; Ming, W.; Weaver, V.; and Janmey, P. A. 2005. Effects of substrate stiffness on cell morphology, cytoskeletal structure, and adhesion. *Cell Motility and the Cytoskeleton* 60(1):24–34.