

NEW METHODS FOR QUANTIFYING AND VISUALIZING INFORMATION FROM IMAGES OF CELLS: AN OVERVIEW

Gustavo Kunde Rohde

Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA. 15213.

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA. 15213.

Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA. 15213.

ABSTRACT

New microscopy imaging techniques have enabled the acquisition of cellular and sub-cellular information with unprecedented accuracy and specificity. Fluorescence techniques have enabled labeling of numerous, previously inaccessible, molecules and organelles, while Raman spectrographic techniques, for example, have enabled label free acquisition. Together with the development of high throughput techniques, these technologies now allow for the acquisition of a significant amount of information about cellular processes and have enabled high throughput and high content screening. Beyond image formation and acquisition, computational techniques comprise an important part of the process of obtaining biological understanding from such experiments. Here we review the pros and cons of the main approaches that have been used to extract information from digital images of cells. In addition, we also offer an overview of modern computational techniques that beyond allowing for discrimination between two hypothesis, also allow for modeling, visualization, and understanding of biological phenomena.

Index Terms— digital microscopy, cytometry, image analysis

1. INTRODUCTION

For centuries, and dating back to the discovery of cells by Hooke in the 1600s [1], microscopy imaging techniques have served as an important experimental technique for biological discovery. In the past few decades, several technological advances have coalesced into elaborate imaging modalities that are able to access high resolution spatiotemporal information about cellular processes with unprecedented accuracy and specificity. Fluorescence molecules can be used to allow scientists to obtain information regarding the spatial arrangement and organization of numerous molecules, proteins, and organelles in live and fixed cells [2]. Combined with digital detectors such as photomultiplier tubes (PMTs), and charge coupled devices (CCDs), as well as intricate optical instrumentation, high resolution digital spatiotemporal imaging in 2 and 3 dimensions became possible [3]. Confocal fluorescence techniques, for example, are routinely used for live cell imaging experiments for a variety of applications, including dynamic aspects of sperm cells during fertilization [4], as well as the dynamic sub cellular localization of chloritin during endocytosis [5] for examples.

In addition to providing means for visualizing the location of different molecules and proteins inside cells, the quantitative nature of photon counting detectors (PMTs, CCDs) has also given rise to imaging techniques that enable the quantitative study of dynamic molecule behavior, as well as the colocalization of multiple molecules. Fluorescence (Froster) resonance energy transfer techniques, for example, have emerged as an important technique to

study changes in molecular proximity, and have been used, in conjunction with live cell imaging techniques, to study the role of different proteins (Rac1, Cdc42, and others) in glioblastoma cell invasion experiments [6], for example.

In addition, an important technological contribution has been the development of ultra fast, high throughput, imaging methods capable of imaging and analyzing millions of particles in a relatively short period of time [7]. These have the potential to revolutionize diagnoses of pathologies through rare event (single cell) detection. Finally, we mention that beyond precise quantification of cellular processes through fluorescence techniques, techniques for label free microscopic imaging using Raman spectroscopy techniques have been developed [8], with the major benefit that these technologies could pave the way for in vivo imaging and diagnosis.

Given the widespread capability for acquiring high resolution information from large quantities of images of cells, computational tools have gained importance and are often integrated into comprehensive imaging ‘pipelines’ to more fully characterize the biological processes being investigated. The purpose of this paper is to provide an overview of the main computational techniques that play crucial roles in the field of image-based cytometry. Rather than a comprehensive and exhaustive review, emphasis will be placed on highlighting the main ideas in current use as well as describe important emerging tools.

2. COMPUTATIONAL TOOLS FOR QUANTITATIVE IMAGE-BASED CYTOMETRY

As described above, microscopic images of cells can be varied in signal source (e.g. fluorescence vs. bright light), resolution, dimension (2D vs 3D), etc. Ultimately, however modern devices produce digital images which are stored in computers as a collection of pixel measurements. For large images of high spatial resolution, numerous pixels are available. For most modern history, when conducting scientific experiments based on imaging assays, microscopic images were visually analyzed to determine the presence or absence of a certain morphological signature in a group of cells that were subjected to a specific controllable effect (e.g. drug, RNA interference, mechanical manipulation, etc.) in comparison to control cells. The morphological signature is normally reflected as a change (presence or absence) of a fluorescence signal, or changes in overall shape or appearance of the cells, in the acquired images. While the human visual system is capable of astounding tasks, it is well known that it has certain limits as far as precise quantification of phenomena, comparing large numbers of morphological exemplars (cells), and finding co variations amongst several factors [9], for example. As well known, for a scientifically oriented person to obtain confidence

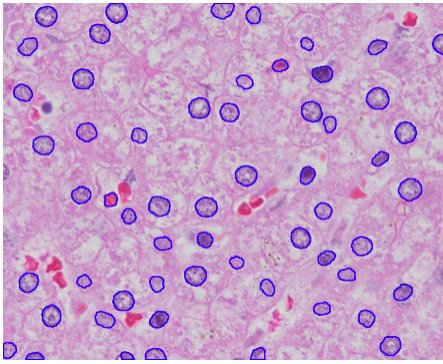


Fig. 2. Example segmentation of nuclei in an H and E stained image. The detected nuclei are shown with a blue (online) contour around them.

on a certain hypothesis (e.g. is drug X effective on cell type Y?), experiments using numerous, several thousands cells perhaps, are required. Computational image analysis methods can thus play a crucial role in helping scientists more categorically test and validate hypothesis, as well as analyze more subtle or complex processes.

As far as computational imaging pipelines for cytometry, before quantitative properties of individual cells can be analyzed, the individual cells must first be segmented from raw digital images when their field of view is large enough to contain many cells. Once properly segmented, certain of their properties are computed, based on which hypothesis can be statistically assessed using pattern recognition algorithms. Figure 1 contains a schematic of a common image cytometry pipeline.

2.1. Image segmentation

A wide variety of algorithms for segmenting cells are currently available for segmenting cells and sub cellular structures from diverse types of microscopy images [10]. This is not a trivial task, primarily due to the different appearance cells can have due to the variety of existing cell types, experimental imaging assays, as well as different microscopic imaging modalities (2D, 3D, fluorescence stains, etc.). Algorithms targeted towards segmenting certain cell types imaged using a particular modality tend not to perform as well for different cell types or different signal targets or imaging modalities. To deal with such large heterogeneity, algorithms that can automatically (or semi automatically) ‘tune’ themselves to a given application are poised to gain in significance. Figure 2 shows an example of such an algorithm [11] that utilizes hand annotated data to ‘calibrate’ itself so that it can accurately segment nuclei from cells in tissue stained with Hematoxylin and Eosin. Note that this method, as many modern algorithms, is able to accurately capture borders of individual nuclei as well as cells in clustered environments, and can be easily adapted (by providing annotated images) to various imaging modalities [11].

2.2. Feature-based methods for hypothesis testing

Once cells have been segmented, groups of cells can be analyzed for differences in control versus effect experiments by calculating relevant ‘features’ from each cell, and analyzing the measurements in multi dimensional ‘feature space.’ The idea is exemplified in Figure 1. A wide variety of features have been used over the past decades to quantify differences in shape and appearance of cells in digital

images. These can be broadly divided into features that aim to characterize shape or texture. Examples of shape related features include the perimeter and area of a cell (when measured in 2D images), as well as average curvature, etc. Example of texture features include Fourier and wavelet-like decompositions, Haralick texture features, as well as others. Once numerical features are computed from two sets (classes) of cells, the statistical significance of hypothesis (e.g. differences of means) can be computed. In addition, other pattern recognition tasks, such as automated classification, can also be computed. Altogether, these can give the experimental scientist the necessary evidence to conclude whether a significant effect has been observed in contrast to a control population of cells.

This approach has been in use for several decades [12]. More recently it has been used in drug discovery experiments [13], as well in automated digital pathology [14], for examples. However, a few limitations are apparent. First, the task of deciding which features are relevant for a particular problem is not trivial. Algorithms for sifting through a large number of features, and, given labeled training data, can attempt to select the most relevant features for a given problem. In addition, a researcher who wishes to find meaning, or biological understanding, from the feature space analysis is left with few direct options, though indirect approaches exist [15].

2.3. Explicit modeling methods for parameter extraction

While the approach above attempts only to describe the differences between two sets of cells, explicit parametric modeling approaches can be used to extract biologically relevant information by fitting the model being used to the image data available. This approach has been used in cytometry to model cellular and nuclear shapes, as well as other organelles [16]. Figure 3 shows how a modeling method [17] can be used to estimate the number and lengths of microtubule from 3D confocal fluorescence images of fixed cells, even though the size of such filaments is well below the optical resolution limit of these instruments, and therefore, in many regions, individual microtubules cannot be discerned by the human eye. Naturally, as all measurements from noisy data, these have errors associated with them. However, these errors have been well characterized and have found not to interfere with ones ability to obtain useful information from experiments involving real cells.

2.4. Implicit modeling for understanding and visualization

An emerging manner through which to mine information from sets of images of cells utilizes mathematical distances that measure the similarity between two or more images directly without the aid of pre conceived numerical features, nor physiological models. The advantage of such methods is that they can be more impartial to pre conceived notions regarding the known biology of the problem, but rather seek to utilize simple mathematical rules to automatically extract relevant information to both discriminate between two or more sets of cells (e.g. control vs. effect), and, as importantly, to visualize and understand what are the principal differences between them. The idea is demonstrated in Figure 4 where the approach developed in [18][19] was used to visualize the most significant (according to the p value for difference of means) difference between nuclear chromatin configuration in cells obtained from normal patients and patients diagnosed with fetal-type hepatoblastoma. It is clear that, aside from nuclear size differences, how much chromatin is placed in the interior of each nucleus as opposed to along its boundary is the major discriminant factor. It is clear from this figure that cancerous cells have a tendency to have its chromatin more evenly distributed

throughout the nuclear envelope. This is consistent with the fact that cancerous cells divide faster than their normal counterparts, and that the phenotype associated with nuclei just before cell division is one where nuclear chromatin is more evenly distributed. Finally, we mention that one can use this technique to visualize, in a completely automated fashion, the most significant differences between two or more sets of cell measurements, but also that the differences found through this method have been used to successfully classify cancerous versus normal patients [20].

3. SUMMARY AND DISCUSSION

Computational image analysis methods have become an important part of cytometry technology. The field has evolved tremendously since its early days [12]. It is now possible to obtain information from cellular images beyond abstract numerical features. Explicit parametric modeling techniques can and have been used to extract meaningful biological information such as detecting microtubules, which are well below the optical resolution used to acquire such images, for example. In addition, modern approaches for measuring distances between cellular shapes [21] and molecules and protein distributions [20, 18] have enabled direct visualization and biological understanding of the main variations within a signal dataset, or of the differences between two classes of cells. Looking forward, better integration of such modern techniques for analyzing cell images with experiments in biology and pathology should further facilitate more complete characterization of important biological processes.

4. ACKNOWLEDGEMENTS

The author acknowledges support by NIH grants GM088816, and GM090033. The author acknowledges important contributions from R.F. Murphy, K.N. Dahl, J.A. Ozolek, D. Slepcev, W. Wang, A. Sharif, C. Cheng, S. Basu, S. Kolouri, A.B. Tosun, S-R. Park, H. Huang, J. Guo, and J. Wang.

5. REFERENCES

- [1] R. Hooke, *Micrographia: or, Some physiological descriptions of minute bodies made by magnifying glasses*. London: J. Martyn and J. Allestry, 1665.
- [2] B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Tsien, "The fluorescent toolbox for assessing protein location and function," *Science*, vol. 312, pp. 217–223, April 2006.
- [3] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Processing Mag.*, vol. 23, no. 3, pp. 20–31, 2006.
- [4] Y. Hamamura, C. Saito, C. Awai, D. Kurihara, A. Miyawaki, T. Nakagawa, M. M. Kanaoka, N. Sasaki, A. Nakano, F. Berger, and T. Higashiyama, "Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in arabidopsis thaliana." *Curr Biol*, vol. 21, no. 6, pp. 497–502, 2011.
- [5] E. Ito, M. Fujimoto, K. Ebine, T. Uemura, T. Ueda, and A. Nakano, "Dynamic behavior of clathrin in arabidopsis thaliana unveiled by live imaging." *Plant J*, vol. 69, no. 2, pp. 204–216, 2012.
- [6] E. Hirata, H. Yukinaga, Y. Kamioka, Y. Arakawa, S. Miyamoto, T. Okada, E. Sahai, and M. Matsuda, "In vivo fluorescence resonance energy transfer imaging reveals

- differential activation of Rho-family GTPases in glioblastoma cell invasion," *J. Cell Sci.*, p. in press, 2012.
- [7] K. Goda, A. Ayazi, D. R. Gossett, J. Sadasivam, C. K. Lonappan, E. Sollier, A. M. Fard, S. C. Hur, J. Adam, C. Murray, C. Wang, N. Brackbill, D. Di Carlo, and B. Jalali, "High-throughput single-microparticle imaging flow analyzer." *Proc Natl Acad Sci U S A*, vol. 109, no. 29, pp. 11 630–11 635, 2012.
- [8] L. Gau, H. Zhou, P. Luo, Y. Yang, A. A. Hammoudi, K. K. Wong, G. S. Palapattu, and S. T. Wong, "Label-free high resolution imaging of prostate glands and cavernous nerves using coherent anti-stokes raman scattering microscopy," *Biomed Opt Express*, vol. 18, no. 2, pp. 915–926, 2011.
- [9] G. A. Alvarez and P. Cavanagh, "The capacity of visual short-term memory is set both by visual information load and by number of objects." *Psychol Sci*, vol. 15, no. 2, pp. 106–111, Feb 2004.
- [10] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 140–145, 2012.
- [11] C. Chen, W. Wang, J. A. Ozolek, N. Lages, S. J. Altschuler, L. F. Wu, and G. K. Rohde, "A template matching approach for segmenting microscopy images," in *IEEE Int. Symp. Biomed. Imaging*, 2012, pp. 786–771.
- [12] J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images," *Ann N Y Acad. Sci.*, vol. 128, pp. 1035–1053, 1966.
- [13] L.-S. Loo, L. F. Wu, and S. J. Altschuler, "Image-based multivariate profiling of drug responses from single cells," *Nature Methods*, vol. 4, pp. 445–453, 2007.
- [14] W. Wang, J. Ozolek, and G. K. Rohde, "Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images," *Cytometry Part A*, vol. 77, pp. 485–494, 2010.
- [15] Z. Yin, X. Zhou, Y. Sun, and S. T. C. Wong, "Online phenotype discovery based on minimum classification error model." *Pattern Recognit*, vol. 42, no. 4, pp. 509–522, 2009.
- [16] T. Zhao and R. Murphy, "Automated learning of generative models for subcellular location: building blocks for systems biology," *Cytometry*, vol. A 71, pp. 978–990, 2007.
- [17] A. Shariff, R. F. Murphy, and G. K. Rohde, "A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images." *Cytometry A*, vol. 77, no. 5, pp. 457–466, 2010.
- [18] W. Wang, D. Slepcev, S. Basu, J. A. Ozolek, and G. K. Rohde, "A linear optimal transportation framework for quantifying and visualizing variations in sets of images," *Int. J. Computer Vision*, in Press, 2012.
- [19] W. Wang, Y. Mo, J. A. Ozolek, and G. K. Rohde, "Penalized fisher discriminant analysis and its application to image-based morphometry," *Pattern Recognition Letters*, vol. 32, no. 15, pp. 2128–2135, 2011.
- [20] W. Wang, J. A. Ozolek, D. Slepcev, A. B. Lee, C. Chen, and G. K. Rohde, "An optimal transportation approach for nuclear structure-based pathology," *IEEE Trans. Med. Imag.*, vol. 30, pp. 621–631, 2011.
- [21] G. K. Rohde, A. J. S. Ribeiro, K. N. Dahl, and R. F. Murphy, "Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells." *Cytometry*, vol. 73A, pp. 341–350, 2008.

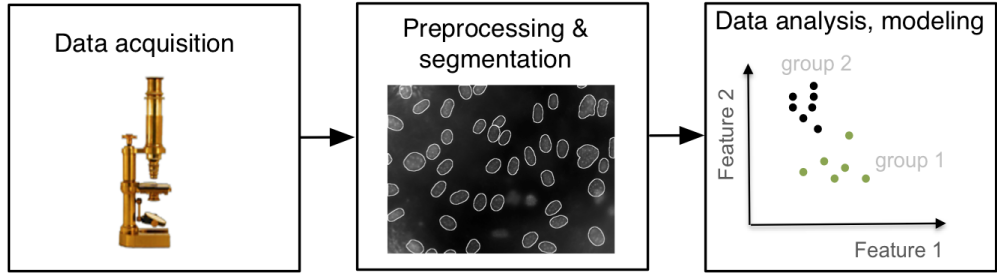


Fig. 1. An overview of a typical quantitative image cytometry pipeline.

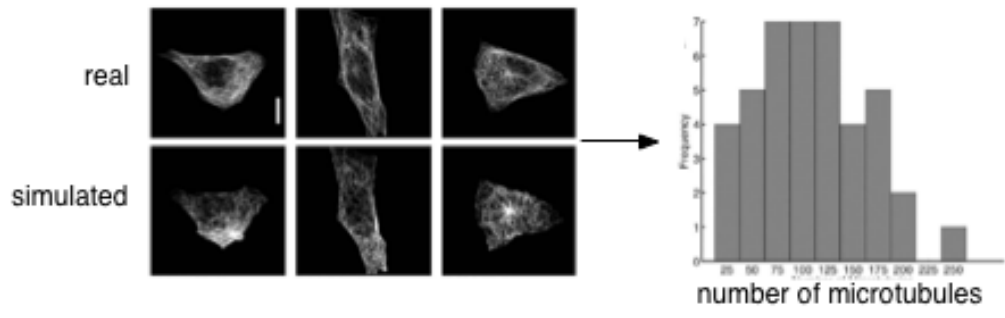


Fig. 3. Microtubule organization modeling and parameter extraction. The left pane shows three confocal images of real cells and their counterpart simulated images (computed through the modeling approach described in [17]). The histogram on the right shows the number of microtubules extracted using this modeling approach for each cell is a set of 50 fixed cells.

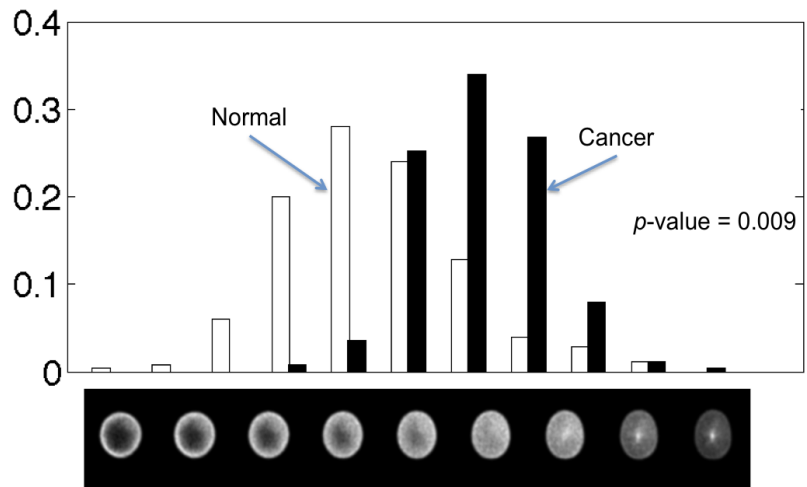


Fig. 4. Differences in chromatin organization between normal and cancerous (fetal-type hepatoblastoma) cells computed using the implicit approach described in [18, 20]. Each bar in the histogram shows the relative number of cells that had their phenotype most closely associated with the nuclear chromatin image displayed directly below it. The p value for differences of means shows the trend is highly significant.