Computational image analysis of cancerous epithelial tissue using MATLAB

Dao Sen Wang under supervision by Dr. Jason Levy
Faculty of Science, University of Ottawa

Introduction

Traditional analysis of cellular images is performed manually and usually on smaller data sets. With advances in technology and computing power, high-throughput automated image acquisition systems have become more prevalent and can produce large image datasets. Quantitative measurements of these datasets can produce big data for achieving big results.

In collaboration with the Kim Lab at the University of Toronto, the overarching goal of the research project is to develop new methods of detecting early stages of cancer. The project focuses on identifying patterns of peroxisome distribution in epithelial tissue of the prostate gland which are strongly linked with cancer progression.

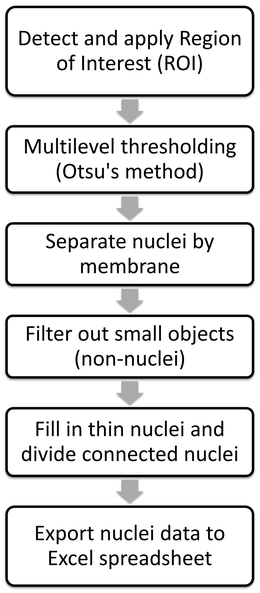
For this component of the project, we designed an image processing system which analysed microscope images and segmented cellular components, with current focus on nuclei. Statistical analyses of the data could identify subtle but consistent differences between healthy and cancerous tissues.

Methods

In order to process and analyze these images in the context of high-content screening (HCS), computational tools must be utilized. Using MATLAB, a commercial computing environment, the image processing system was built from a fundamental level. To have the option of refining the source code can increase efficiency and precision of image analysis.

- The Image Processing Toolbox™ provided necessary functions for the algorithm
- Datasets comprised of 3D microscopic images stained with fluorescent markers: DAPI (blue) for nuclei, green for cell membrane, red for peroxisomes
- The images included the region of interest (ROI) set manually by the researcher

The logical framework of the overall algorithm is shown and described below. We began development with a preliminary algorithm and made improvements as necessary.



- ROI defined area to be manipulated by algorithm
- Gray scaling by multilevel thresholding for DAPI marker to detect nuclei
- Identified connected components using `bwconncomp`
- Membrane marker (green) used as the bounding region for individual nuclei
- Objects smaller than threshold removed
- Irregular shaped nuclei reshaped circularly
- Individual and average sizes of nuclei exported

Results

After running various image sets through the algorithm, we noticed patterns of errors in the output. Nuclei were correctly identified; however, non-nuclei were also marked as well. We made improvements to the preliminary algorithm in the revised algorithm. Clean-up of the image was an important goal along with increasing robustness of the algorithm.

Preliminary algorithm

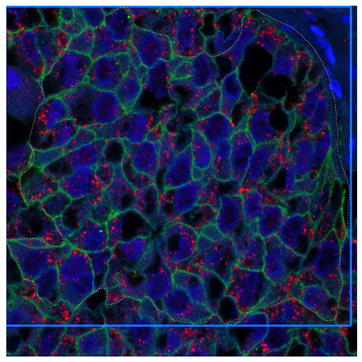


Figure 1. Middle slice of microscope image stained with fluorescent markers

- **Figure 1** shows the initial image to be processed
- The ROI is indicated by the dotted line
- The intensity of the coloured markers is used to separate the cellular components

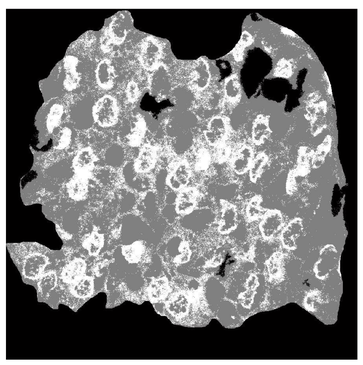


Figure 2. Image after multilevel thresholding for DAPI (blue) marker

- **Figure 2** is produced after multilevel thresholding for DAPI (blue) using Otsu's method
- High intensity areas of blues are marked white, low – grey, zero – black
- Non-nuclei stained blue were also included as white by this method
- Division by membrane marker is applied as an attempt to divide grouped nuclei

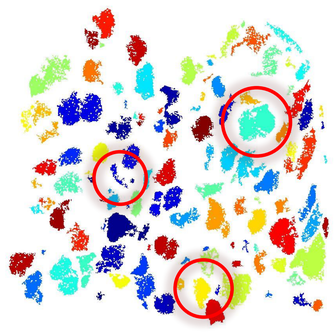


Figure 3. Image output of algorithm detecting nuclei (with errors)

- **Figure 3** shows the image output after the preliminary algorithm
- Common errors were non-nuclei objects, thin/hollow nuclei, and adjoined nuclei
- We made further improvements focused on resolving these errors

Revised algorithm

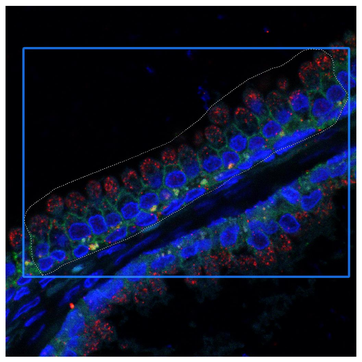


Figure 4. Middle slice of microscope image stained with fluorescent markers

- **Figure 4** shows the initial image to be processed
- The ROI is indicated by the dotted line
- Outside the ROI are extracellular regions not to be considered (e.g. basement membrane)
- Shape of some cells in image are affected by curvature of tissue

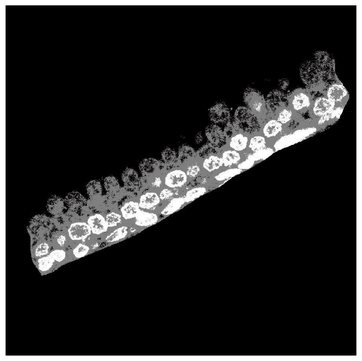


Figure 5. Image after multilevel thresholding for DAPI and noise removal

- **Figure 5** is produced after multilevel thresholding for DAPI in similar fashion as previously
- Noise removal is performed using the function `imopen`
- `imopen` first erodes the image then dilates it, essentially removing small or random signals
- Division by membrane marker is applied once again

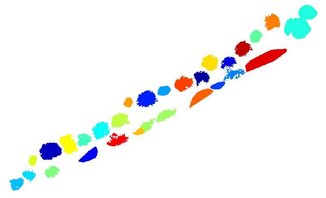


Figure 6. Image output of algorithm after adjustments for hollow or connected nuclei

- Filtering of small objects (<100 pixels) is performed
- Opening of the image is performed to identify hollow or adjoined nuclei
- Hollow nuclei are filled in and adjoined nuclei are partitioned
- **Figure 3** shows the image output after the revised algorithm

	A	B	C	D
1	Image	1	2	3
2	Number of nuclei	58	41	30
3	Median nucleus size	3091.5	2833	1249.5
4	Nucleus size	6247	5665	1040
5		2700	8977	771
6		1716	621	585
7		1259	3081	2455
8		7480	5129	997
9		3821	5932	1926
10		4466	2833	1088

Figure 7. Sample Excel spreadsheet for nuclei data

- Relevant data is exported and saved in an Excel spreadsheet (**Figure 7**)
- The number of nuclei, median and individual nucleus size are the current parameters for each image

Conclusion

Using the revised algorithm, nuclei are correctly identified in most image sets with few exceptions. Further improvement to the segmentation algorithm can be made to improve efficiency and accuracy.

Continuing with image analysis, the next steps in this component of the project are as follows:

- Segmentation of cells
- Counting of peroxisomes within cells
- Statistical analysis of the distribution of peroxisome counts

We can begin the segmentation of cells by applying a watershed algorithm. Using the location of individual cells we can determine the distribution of peroxisomes within cells.

Using the data collected through this image processing system, we look to identify statistical significant differences between healthy and cancerous tissues. From those differences, we hope to develop new methods for the diagnosis of early stages of cancer.

Acknowledgements

I would like to thank J. Levy for his guidance and patience throughout the whole learning process. Thanks to the Kim Lab at UofT for the image datasets. Finally, a thank you to UROP and uOttawa for making this experience possible and extremely enjoyable.

References

- Gonzalez, R., & Woods, R. (2009). *Digital Image processing using MATLAB®* (2nd ed.). United States: Gatesmark Publishing.
- Lodhi, I. J., & Semenkovich, C. F. (2014, March). Peroxisomes: A Nexus for Lipid Metabolism and Cellular Signaling. *Cell, 19*(3), 380-392.
- Shamir L, Delaney JD, Orlov N, Eckley DM, Goldberg IG (2010) Pattern Recognition Software and Techniques for Biological Image Analysis. PLoS Comput Biol 6(11): e1000974.doi:10.1371/journal.pcbi.1000974