Technical test report

Objective

The objective of this project is to implement a watershed-based segmentation algorithm to segment cell cytoplasm (actin) given the corresponding nucleus (DNA). The project is a part of the analysis pipeline that aim to retrieve and quantitatively characterizes morphologies in Human HT29 colon-cancer cells.

Method

Based on the discussion with the panel, the seeded watershed algorithm will be used for segmenting cytoplasm using seeds derived from corresponding nucleus. The general workflow is described as follows:
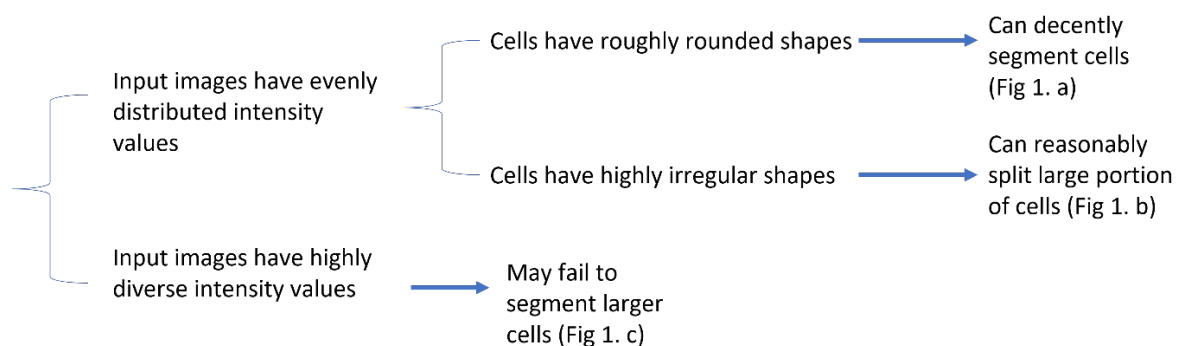
- **Preprocessing**: the input images are first normalized and then the contrast is enhanced.
- **Defining seeds**: the global 'Otsu' thresholding is applied to the DNA image then a connected component labeling algorithm is applied to segment nucleus as seeds.
- **Watershed segmentation**: the global 'Otsu' thresholding is first applied to the actin image, and morphological operations are implemented to extract the background. Then the original actin image is sharpened followed by the calculation of the gradient. Finally, the seeded watershed algorithm is applied on the gradient image using seeds from the previous step.

As for the output, a label map image and an overlay of label map and actin are produced for each of 56 fields of view. The label map image can be used for further quantification and the overlay image is for visualization.

Results and Discussion

Note that without quantitative assessment, results evaluation and discussion are based on visual observation only.

A simple diagram below shows a summary of segmentation results given different situations with corresponding examples in Fig 1.

Input images have evenly distributed intensity values

Cells have roughly rounded shapes → Can decently segment cells (Fig 1. a)

Cells have highly irregular shapes → Can reasonably split large portion of cells (Fig 1. b)

Input images have highly diverse intensity values → May fail to segment larger cells (Fig 1. c)

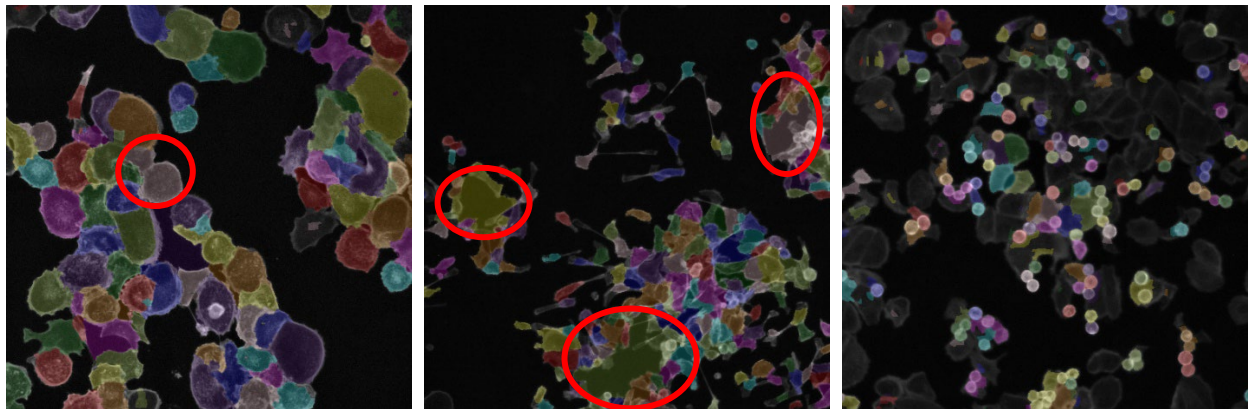Fig 1   (a)   (b)   (c)

For the current processing pipeline, several limitations still exist and are discussed as below:

1. **Failing to differentiate foreground/background:** the pipeline sometimes fails to recognize some smaller portions of background, an example is shown in Fig 1 (b) in the regions marked by red ovals. This is because the morphological operations applied to extract the background are not robust enough.
2. **Failing to split cells closely connected to each other:** this usually happens when the boundary of connected cells has relatively low contrast. Since the watershed is based on gradient image, low contrast of boundaries makes local maximum less accurate, causing leakage in the segmentation. An example is shown in Fig 1 (a) marked by the red oval.
3. **Failing to segment larger cells:** this happens when there are both large and small cells in the image and their intensity values are highly diverse (usually small cells have much higher intensities). This causes the local maximum to fall in the boundaries of small cells and fail to expand to surrounding larger cells. An example is show in Fig 1 (c).

Some thoughts on future work

1. A post-processing step might be added to fine-tune the segmentation, for example, a region-growing/level set algorithm to expand under-segmented cells.
2. A better normalization/rescaling technique to make the diverse intensities more evenly distributed between large and small cells.