VIP IPA: Nuclei Detection and Counting

Authors: Tikhon Pachin, Yu-Hsuan Lin, Annapoorna Prabhu, Aubrey Gatewood, Nikhil Manglore, Niharika Narra, Lydia Hanna
Mentors: Prof. Edward J. Delp, Prof. Carla Zoltowski

PURDUE UNIVERSITY

INTRODUCTION

Nuclei segmentation on microscopy images has been a topic of interest for over 10 years. There are multiple different algorithms that have been developed to target the issue. This project focuses specifically on Watershed segmentation with some preliminary image processing to segment and count nuclei. Nuclei segmentation is an important step in cancer prediction and detection.

The project was split into proof of concept and algorithm altering. Proof of concept utilized some preexistent tools to segment the nuclei. Algorithm altering team implemented and adjusted the Watershed algorithm from scratch.

The output of nuclei segmentation is an image that shows the borders between the cells and a number that corresponds to the number of cells in the image. This output is a preliminary processing method for further research that may be conducted on such images.

DATA

For the purposes of our project we used Broad Bioimage Benchmark Collection resources that provide open source ready to use microscopy images for testing purposes.

Source: https://bbbc.broadinstitute.org/

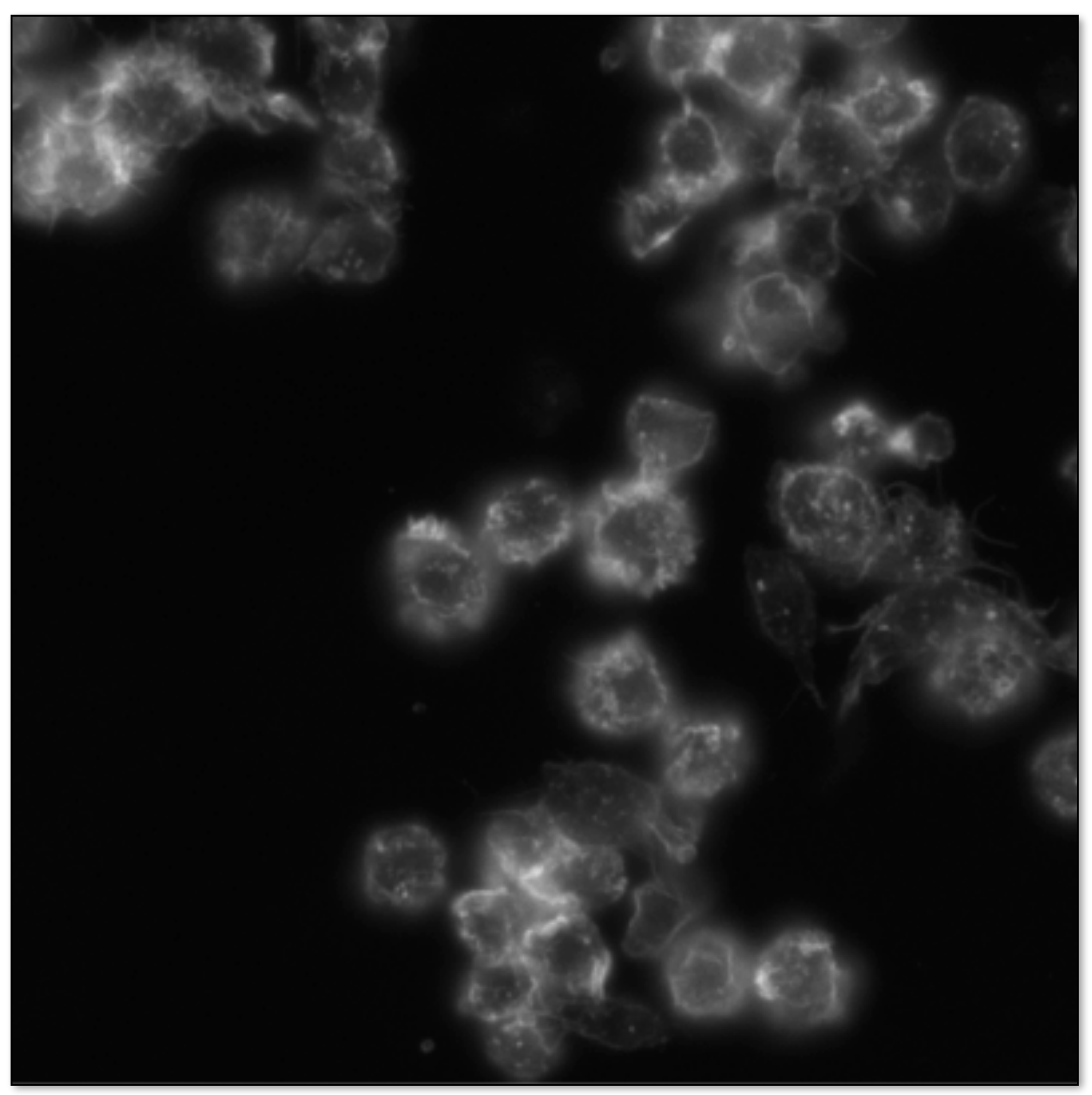
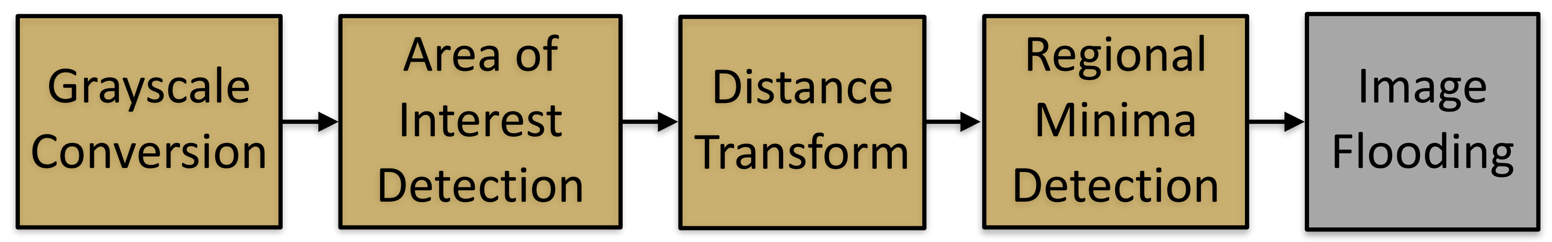


Image1. Dataset Example Image

METHODS

This project utilizes **Watershed segmentation**. The images are processed by considering them as topographical maps. The following methods are implemented to achieve the result:

- **Grayscale conversion:** converts a colored image into values of intensity for each pixel.
- **Area of interest detection:** utilizes filters such as Otsu's thresholding, dilation, erosion, opening, closing, exposure to detect all the areas where cells are present.
- **Distance transform:** converts the areas of interest into a topographical map.
- **Regional minima detection:** detects all the basins on the topographical image.
- **Image flooding:** fills up the image with 'water' until it hits the predetermined ceiling to detect where the 'watersheds' are, i.e. the borders of each cell.



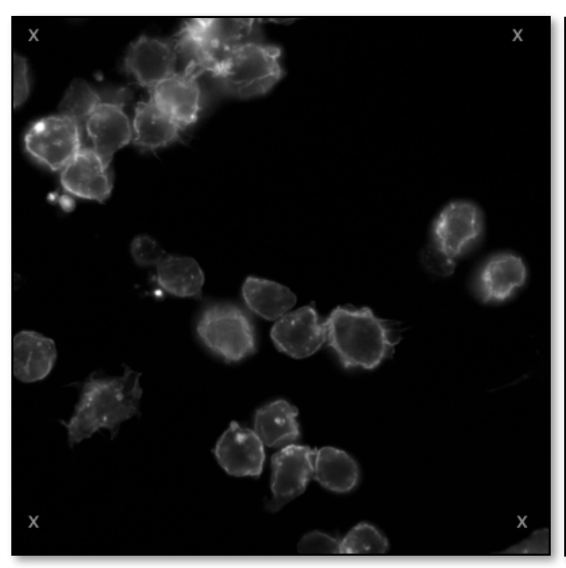


Image2. Original Image

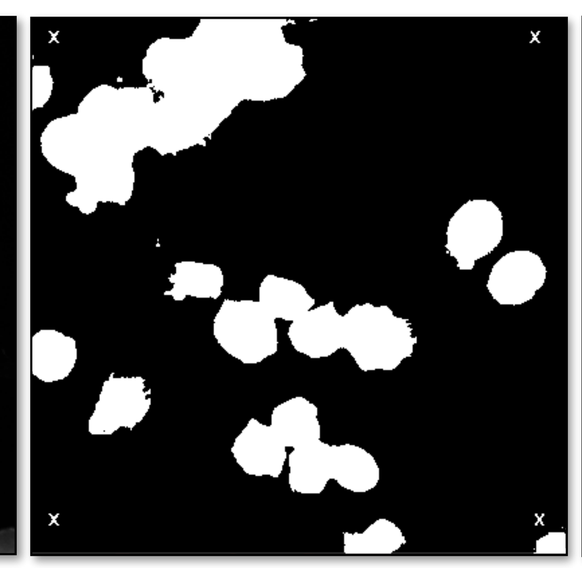


Image3. Areas of Interest

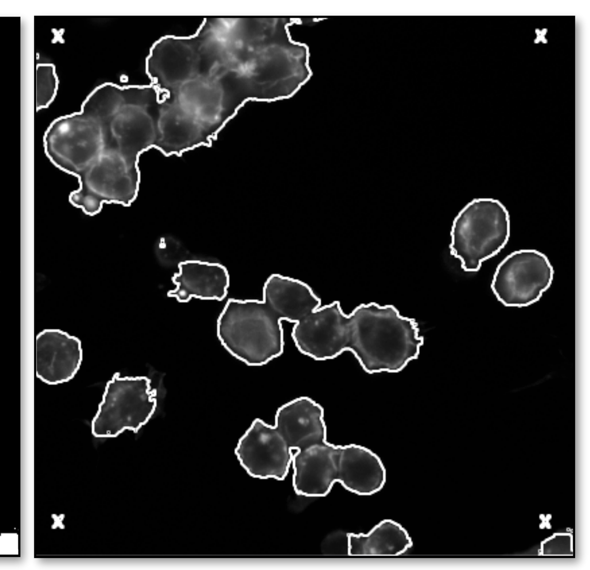


Image4. Areas of Interest vs Original Image

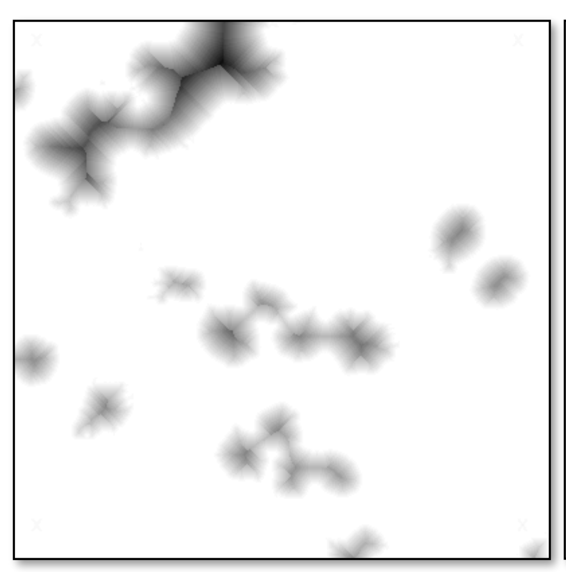


Image5. Euclidean Distance Transform

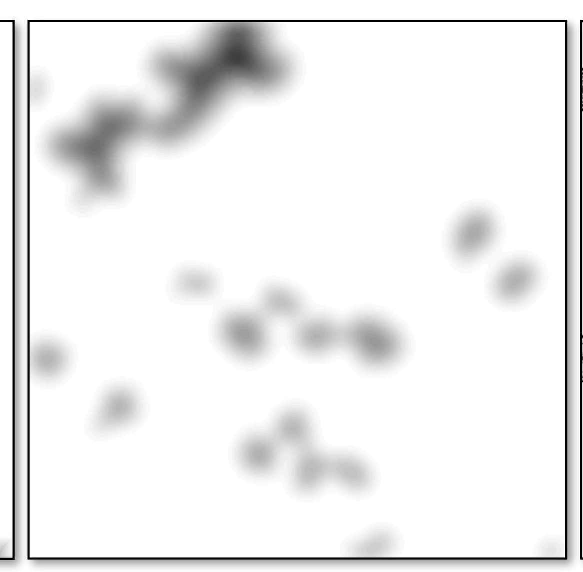


Image6. Gaussian Blurred Euclidean Distance Transform

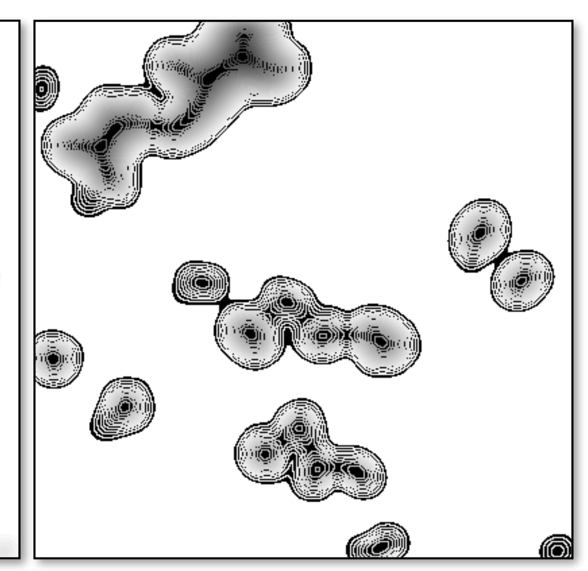


Image7. Regional Minima

DISCUSSION

The team was split into two sub-teams. The first team was in charge of implementing the Watershed segmentation from scratch. The other team was in charge of creating the proof of concept by utilizing external sources and libraries such as OpenCV.

- **Grayscale conversion.** ITU-R Recommendation BT.709: $\mathbf{Y = 0.2126R + 0.7152G + 0.0722B}$.
- **Area of interest detection.**
  - Gaussian blurring is used as noise reduction method. The noise prevents accurate detection of the areas of interest. The algorithm assigns weights to pixels around the target pixel for intensity calculations.
  - Otsu's thresholding is used to separate the background from the foreground. The foreground pixels get the intensity value of 255 and the background pixels get 0.
  - Dilation filter is used to overestimate areas of interest. Thus, the exact areas of interest will be selected within.
  - Exposure filter is applied to the overestimated areas of interest to bring out the foreground pixels beyond the threshold and keep the unwanted background pixels before the threshold.
  - Repeated Otsu's thresholding on the image yields accurate areas of interest for further processing.
- **Distance transform**. The areas of interest may be seen as a topographical map. The map is created by calculating the closest distances of feature elements to non-feature elements on the preprocessed image. Feature elements are the pixels with 255 values of intensity. Non-feature elements have 0 values of intensity. The three major ways of calculating the distances are Chessboard, City-block and Euclidean distance transforms.
- **Regional minima detection.** This algorithm is an important step that lastly precedes the flooding. The flooding of the image is not necessary for the cell counting, since it is only utilized to detect the edges of each cell. This algorithm simply scans through the image detecting the pixels that hold the minimum value within the kernel. It is important to note that if this algorithm is applied on the distance map, the resulted output will be over-segmented. Thus, distance transform post-processing is necessary. A simple smoothing operation yields the needed result.

Current output: the image that shows the 'true' and 'false' regional minima. The 'true' minima are the centers of each cell and they hold the lowest value within their surroundings. The 'false' minima are the rest minima on the image. The 'true' minima are visible on the images 8 through 10 and they represent the number of cells on the image.

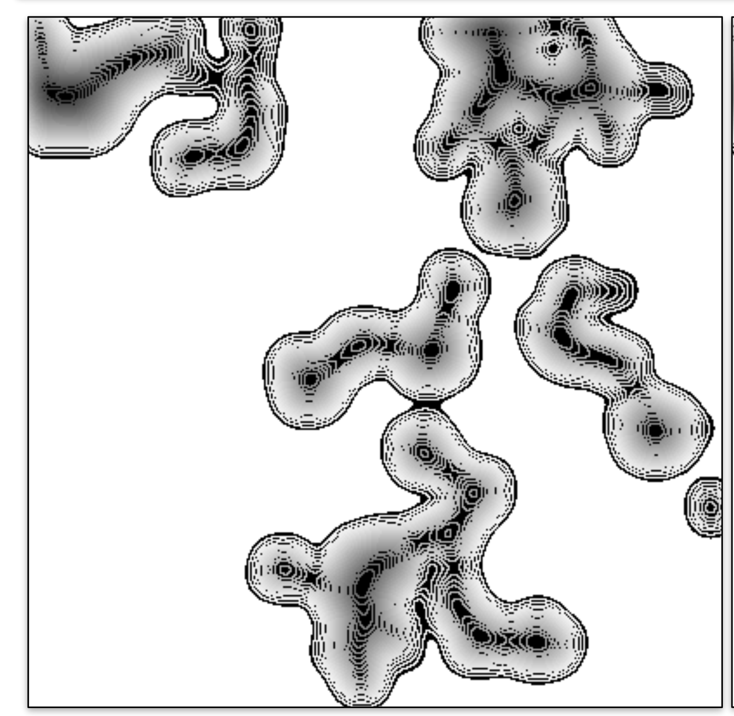


Image8. Euclidean Distance Transform

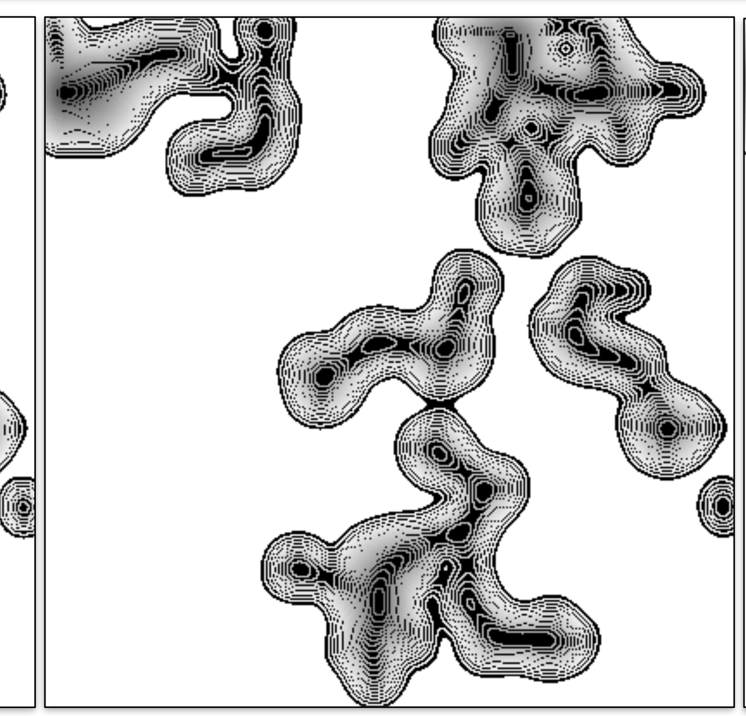


Image9. Chessboard Distance Transform

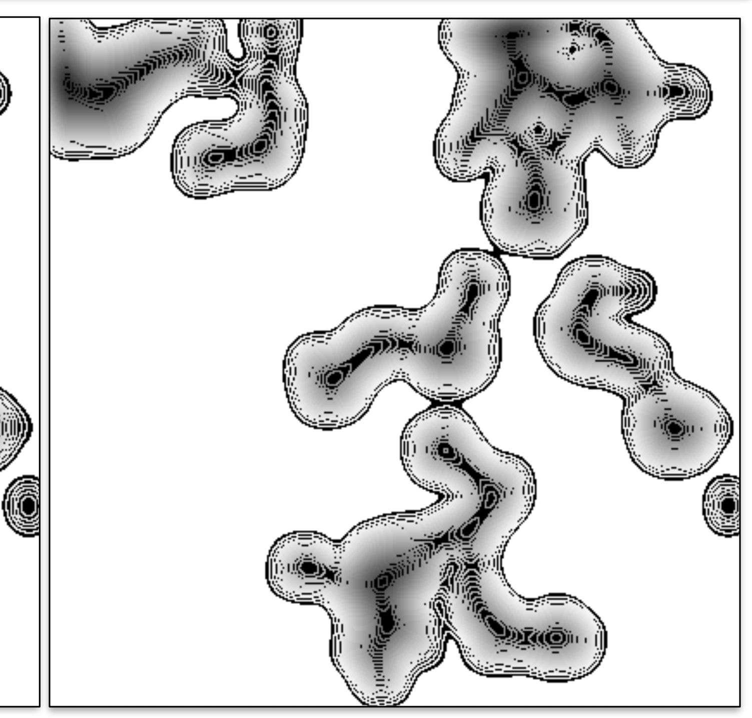


Image10. City-block Distance Transform

RESULTS AND FURTHER STEPS

The team has managed to complete the most crucial preliminary steps before flooding the image. The final step of the project is the flooding of the image. However, the final step is not necessary to determine the accuracy of the segmentation and counting. The current algorithm manages to separate the cells that are connected to each other but not heavily clustered. Some heavily clustered cells are also detected.

The next steps of the project are to flood the image to confirm the segmentation prediction based on the regional minima detection. If the prediction stands, we will continue with the algorithm optimization and improvement. One way to do this is to review different types of Watershed segmentation, since there are a number of them.

Purdue Undergraduate Research Conference Fall 2021