EBI, Technical Test

Yamna Ouchtar

1 Introduction

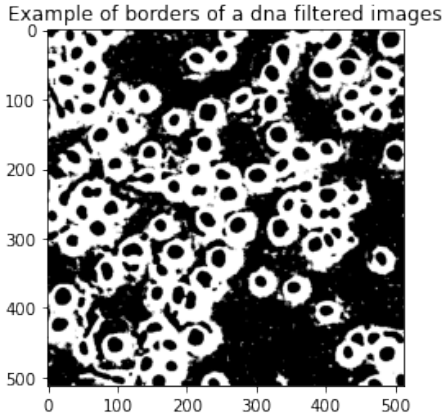
We aim to compare the different morphologies using an image dataset made of 14 samples of HT29 human colon cancer cells. To study the morphologies, the first step is to visualize and find the different cells in these images. To do this, we first need to separate the images into 3 categories: DNA, actin and ph3 images. Then, we will use a watershed segmentation on the actin images containing the cell cytoplasm using the markers obtained on the DNA images, which themselves contains the nucleus. Indeed, a nucleus represents a cell and therefore corresponds to a cytoplasm. As cytoplasm is difficult to segment, the use of nuclei will greatly improve segmentation.

2 Method and Results

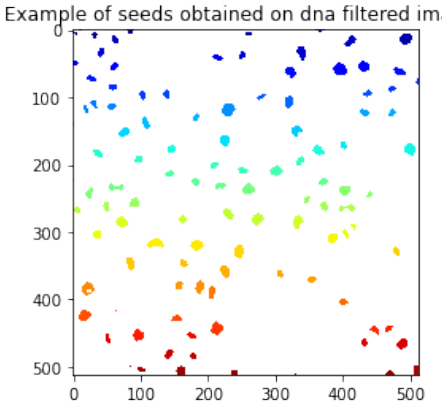
The first step is to open and read the images. Using skimage, we easily separated our images into 2 tables. The first one containing the DNA images and the second one the actin images. I then work on the DNA images first to obtain markers. By using a Gaussian filter and separating the background (Figure 1a), I obtained strong seeds (Figure 1b) where each seed corresponds to a nucleus. In order to improve the watershed segmentation, I then filtered the actin images using a sobel filter. I then applied the watershed algorithm on the images using the seeds obtained in the previous step. The results of the segmentation can be seen in Figure 1c. Using this segmentation, the aim is to measure the different regions and then produce a table containing all the different measurements. This table can then be used as input to a Kmeans method, to group the HT29 human colon cancer cells into 14 groups.

3 Discussion

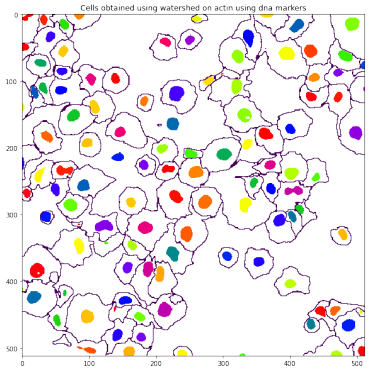
The segmentation obtained seems to be interesting and improved significantly the segmentation results compared to one without watershed markers. Nevertheless, it can be observed that some nuclei are not selected. This problem probably comes from the filtering step. Indeed, I put a Gaussian filter with a parameter of 3 and an uniform filter with a parameter of 10. But these two parameters are probably not the best. It's the same case with the mask I defined.



(a) Filtering



(b) Seeds



(c) Final Segmentation

In fact, I set the parameters to 10 by trial and error, but one can identify the perfect parameters by using for example a grid search. Similarly, I only put a sobel filter on the actin images, but using another filter can probably improve the final segmentation. Finally, we also observe some segmentation noise, which could be eliminated by removing small clusters or with a filter based on the pixel level.