Automated Segmentation of Cervical Cells

Raymond Bamford
Bowen Song

Abstract – The method for segmenting cervical cells implemented and tested in this report comes directly from that found in Tsai et al. [1]. This method consists of three steps: pre-processing by bi-group enhancement, cytoplast contour detection by a k-means algorithm, and nucleus contour detection by the maximal color difference (MCD) method. This report contains an explanation of the implementation of this method, as well as a qualitative evaluation of the results. Difficulties and obscurities that were encountered during implementation and testing are discussed as part of this report.

I. Introduction

The Pap smear is a screening test used for detecting the development of cervical cancer in female patients. The test consists of taking a sample of cells from the uterine-cervix and placing them under a microscope for visual examination. By taking a microscopic image of the cervical cells, computer vision and image processing techniques can be applied to the image to automatically segment the cytoplasts and nuclei that are present. This automated segmentation can aid in the Pap smear diagnosis and save both time and money. Several challenges are present in this area of medical image processing, since there are large variations in cell structure, contrast, noise, and other obstructions present on Pap smear images.

II. Related Work

This cervical cell segmentation algorithm is a replication of the cytoplast and nucleus contour CNC detector method proposed by Tsai et al. [1]. The latter sections of this report explain the steps taken for implementation and the results that were observed.

III. Proposed Approach

The CNC detector consist of three sub-algorithms. The CNC detection begins with a pre-processing algorithm that implements a median filtering and bi-group enhancement, followed by a cytoplast contour detection method that uses a k-means algorithm, then concluded by a nucleus contour detection approach that utilizes the MCD method. Each sub-algorithm has its own equations and steps associated with it as detailed in the following sub-sections.

A. Bi-Group Enhancer Approach

This pre-processing approach applies a median filter prior to the bi-group enhancer algorithm. For the median filter, a nxn window W_{ij} is made around each pixel p on the image, with p being the center pixel. The pixel p is assigned a new color equivalent to the median color of the window. The newly filtered image is called f_t. Let C_{ij} be an array containing all the colors of the pixels of W_{ij} around p of the new image f_t. If C_{ij} is sorted in ascending order, the color c' of pixel p is given by the equation below, where $\text{mid} = (m^2 + 1)/2$.

$$c' = \begin{cases} \frac{1}{\text{mid}} \sum_{i=1}^{\text{mid}} c_i, & \text{if } \frac{1}{\text{mid}} \sum_{i=1}^{\text{mid}} c_i \leqslant c \leqslant c_{\text{mid}}, \\ \frac{1}{\text{mid}-1} \sum_{i=\text{mid}-1}^{m^2} c_i, & \text{if } c_{\text{mid}} < c \leqslant \frac{1}{\text{mid}-1} \sum_{i=\text{mid}+1}^{m^2} c_i, \\ c, & \text{otherwise.} \end{cases}$$

To summarize, for each pixel with a relatively higher value the pixel value is replaced by the mean of the interval $[C_{mid}+1 \quad C_{m^2}]$, and each pixel with a relatively lower value is replaced by the mean of the interval $[C_1 \quad C_{mid}]$. Pixels with colors that are outside of the interval $\left[\frac{\text{mean}([C_1 \quad C_{mid}])}{\text{mid}} \quad \frac{\text{mean}([C_{mid+1} \quad C_{m^2}])}{\text{mid}-1}\right]$ are left unchanged.

B. Cytoplasm Contour Detection Approach

The cytoplast contour detection approach uses the k-means clustering algorithm. The algorithm starts with picking the maximum value in the input image (C_n) and minimum value (C_i). An interval is constructed with the lower bound as $C_i + \epsilon$ and the upper bound as C_n. Two random values are chosen within this interval to represent the foreground color (lower value) and the background color (higher value). Every pixel in the image with a value inside this interval is classified into one of these two groups according to the relative distance to these two representative values. Once an initial background and foreground have been made from the two groups, new representative values are computed as the average intensity of each group. The new representative values are now used to re-classify (by the relative difference in color) the two groups of the image and the algorithm repeats until the representative values no longer change.

C. Nucleus Contour Detection Approach

This approach uses the maximal color difference (MCD) method to detect the nucleus contour. The algorithm requires an input contour, S. Tsai et al. recommends using the cytoplast contour from the previous approach and shrinking it by t pixels inward [1]. There are three possible scenarios as presented in Figure 1.

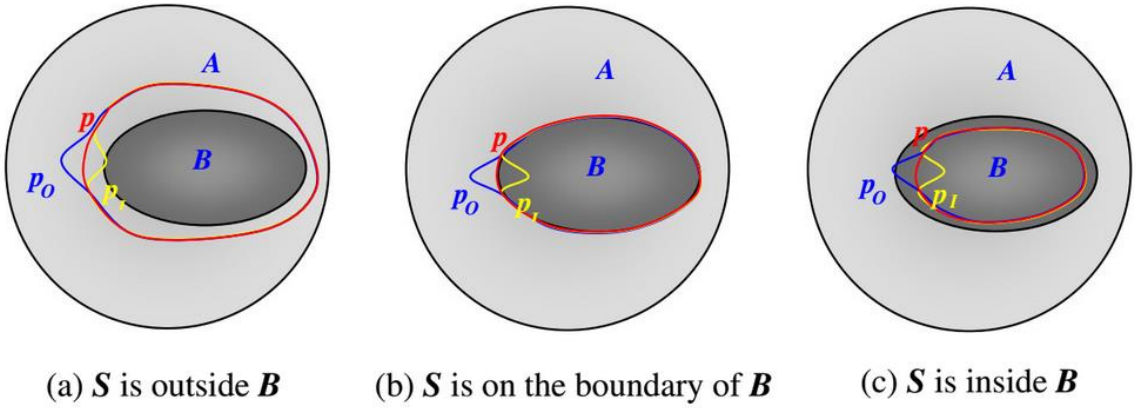

Figure 1: Three different locations of S.

The MCD method utilizes the algorithm below:

repeat
$D_O = \left|\frac{C_i n_i + C}{(n_i+1)} - \frac{C_o n_o - C}{(n_o-1)}\right|$, $D_I = \left|\frac{C_i n_i - C}{(n_i-1)} - \frac{C_o n_o + C}{(n_o+1)}\right|$,
if $D_O = \max(D, D_O, D_I)$ then
  $D = D_O$, $S = S_O$
  move p to p_O.
else if $D_I = \max(D, D_O, D_I)$ then
  $D = D_I$, $S = S_I$
  move p to p_I.
    try to move the next pixel on S
until no pixel on S is moved

If point p is a pixel located on the contour S, p_o is the pixel located 1 pixel away from p outside S and on the line that is normal to S at p, and p_i is the pixel located 1 pixel away from p inside of S and on the line that is normal to S at p. S, S_o and S_i represent the initial contour, the contour when p is moved to p_o, and the contour when p is moved to p_i, respectively. D, D_o and D_i represent the difference in color of each contour S, S_o and S_i respectively. Finally, n_o represents the number of pixels outside of S and n_i represents the number of pixels inside of S. n_o and n_i were re-calculated in this implementation each iteration, by summing the mask inside of the contour and the mask outside of the contour.

IV. Discussion of Results

The selected data base consists of 15 images of overlapping cervical cells with 512x512 pixels and 4 larger images of overlapping cervical cells with 1024x1024 pixels as test subjects. Although our database consists only of overlapping cell images, the algorithm that was implemented was designed on images with only two objects: cell and background. While implementing this algorithm, most of the testing was done on single cell images, to try to replicate the results of Tsai et al [1]. The original goal and plan was to replicate the algorithm first, and then apply it our database of overlapping cells to evaluate its performance and make any suitable adjustments. However, this original plan was abandoned, since the results of the original implementation did not yield the expected results showcased in Tsai et al [1].

A. Results of Bi-Group Enhancer

The goal of this bi-group enhancer algorithm is to eliminate noise in the image by first applying the median filter and then increasing the contrast of objects vs. background, while also distinguishing edges more effectively. The result of this pre-processing step will help greatly in the detection of the cytoplasts and nuclei. Comparing the result of this implementation and with the expectation, we found this implementation needs improvement since foreground cell image contrast is not what is expected from Figure 2. The work of Tsai et al. didn't specify more details for implementing the algorithm, so the problem encountered lies either in faulty interpretation, or additional information that is required in order to properly implement [1].

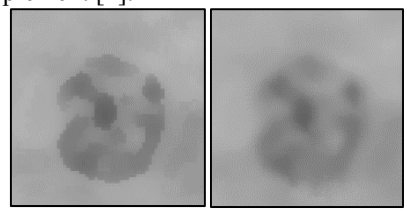

Figure 2: Bi-group output from Tsai et al. (left) [1]. Bi-group output from this implementation (right).

B. Results of Cytoplast Detection by K-Means Algorithm

The K-means algorithm was implemented partially successfully. The detection of a single cell's cytoplast was accomplished (Figure 3), but the detection of multiple, overlapping cells was not ideal (Figure 4). The error here is believed to be from the utilization of the MATLAB imfill command, which is causing spaces between clusters of cells to be filled in, instead of just the cells being filled. No recovery action was achieved for this issue, as there was no time to implement and test an alternative method.

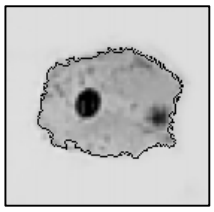

Figure 3: K-means algorithm applied to single cell.

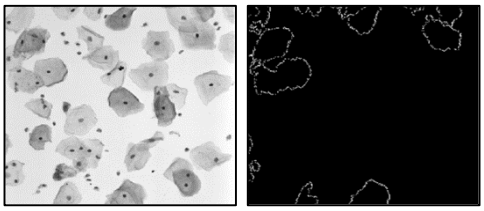

Figure 4: K-means algorithm applied to multiple, overlapping cells. Input (left). Output(right).

C. Results of Nucleus Detection by MCD Method

The MCD method was intended to take an initial contour and move and reshape it to converge onto the edges of greatest color difference: the edges of the nucleus (or nuclei). Unfortunately, this was not achieved in the implementation (Figure 5 and Figure 6). One reason for this could be from breaking the contour during the method. Once a pixel is isolated and on its own, the normal line through that pixel can no longer be calculated because it is not part of an edge. Attempts were made to fix this problem by working with masks as well as contours, instead of just contours so that moving pixels on the mask could be done in such a way there would always be a connected contour that can be deduced from the mask. The attempts to make this work were also unsuccessful. With more time, it would have been possible to attempt a different method altogether for detecting the nuclei. One idea that came from trying to implement the MCD method, is that when taking the color difference at each point on the image, if one collects these values and plots them, the resulting image clearly shows the edges around the nuclei. It may be possible to use this data to segment the nuclei.

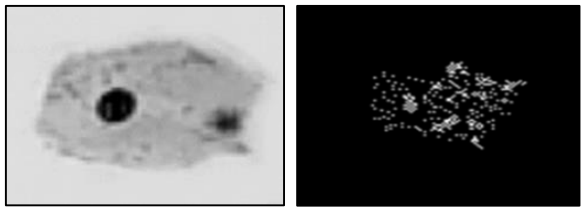

Figure 5: MCD method applied to single cell. Input (left). Output (right).

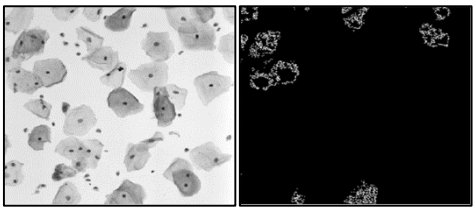

Figure 6: MCD applied to multiple, overlapping cells. Input (left). Output (right).

V. Conclusion

Overall, the methods applied in this approach to the cervical cell segmentation problem are difficult to implement, and perhaps need further explaining in order for a third party to replicate. The results obtained from the implementation were not sufficient enough to perform meaningful quantitative analysis. Also, with the amount of time spent trying to debug issues with the algorithm, there wasn't time for a thorough evaluation.

For future work, the bug that is present in the bi-group enhancer should be found and resolved to overall improve the results of the latter two methods. Also, additional work should be done on the nucleus detection method by formulating a method to ensure that the contours don't break. It would also be useful to research an alternative method altogether, since the MCD method was relatively time consuming.

Acknowledgement

Paper research & database download
- Dr. Alexandra Branzan Albu
- Raymond Bamford
- Bowen Song

Project progress presentation preparation and delivery
- Raymond Bamford
- Bowen Song

Algorithm implementation
- Raymond Bamford
- Bowen Song

Final project demo
- Raymond Bamford
- Bowen Song

Final project presentation preparation and delivery
- Raymond Bamford
- Bowen Song

Final project report
- Raymond Bamford
- Bowen Song

References

[1] M. Tsai, Y. Chan, Z. Lin, S. Yang-Mao, P. Huang, "Nucleus and Cytoplast Contour Detector of Cervical Smear Image", Pattern Recognition Letters 29.9 (2008): 1441-1453. Web.