Medical Image Processing

Detection & Segmentation of cells in a Pap Smear test

By Kimiya Mousavi & Miguel Manguerra

TABLE OF CONTENTS

Introduction

- Pap smear screening tests detect potential cancer by examining cells from the cervix under a microscope.
- These tests are performed manually by healthcare professionals.
- Factors like cell overlap, poor cytoplasm contrast, and the presence of blood or mucus affect the accuracy, speed, and cost of pap smear tests.
- The primary goal of the project is to extract individual cytoplasm and nuclei boundaries from cervical cytology images that overlap.

Objectives

1. Obtain exposure to state of the art medical imaging literature.
2. Write Matlab code for automatic cell detection and segmentation of overlapping cells in Pap smear images.
3. Boundary extraction of individual cytoplasms and nuclei.

Implemented Study

“A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images” [1]

- Detects and segments nuclei and overlapping cytoplasm in cervical cytology images.
- Defines boundaries using a similarity metric and refined by reducing concavity (coarse refinement) and iterative smoothing (fine refinement).
- Method missed less than 5% of cells when the pairwise cell overlapping was less than 0.3.
- Method missed 7% of cells in a dataset of 810 images with 4860 overlapping cells.

ALGORITHM

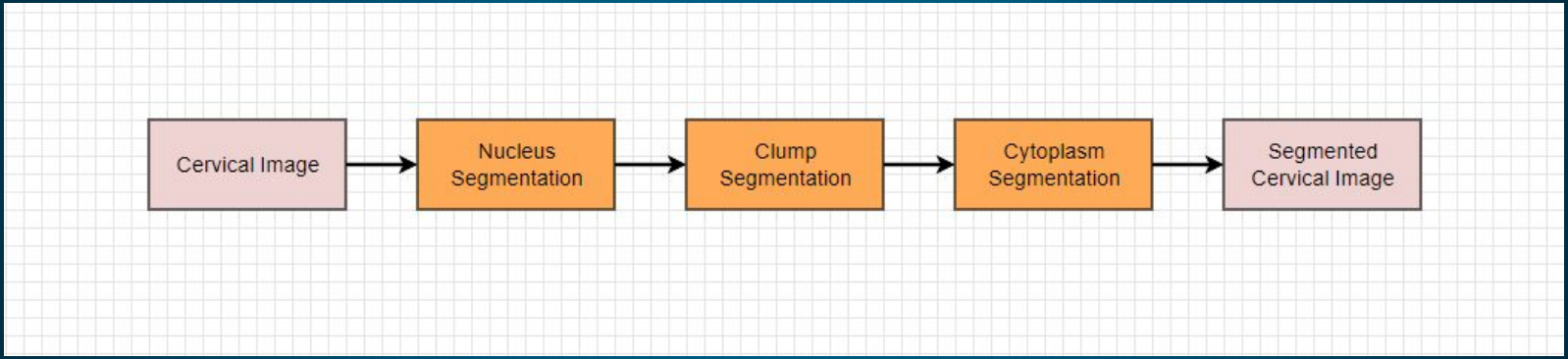


Figure 1 - Algorithm from implemented study

Nucleus Detection and Segmentation

- Input parameters:
  - EDF image
  - Minimum nucleus area
  - Nucleus solidity
  - Lower intensity threshold
  - Higher intensity threshold
  - Boundary intensity image
- Outputs a nucleus mask.

```
N ← ∅
I ← Wiener (I)
for t ← t1 to t2 step 10 do
    B ← I ≤ i
    foreach region r in B do
        if size of r < m or solidity of r < s then
            remove r
        else if binary mask of r ∩ N = ∅ then
            N ← N ∪ r
        else
            /* r overlaps with regions r1, r2, ..., rn
               in N                                    */
            if solidity of r ≥ solidities of all
             regions r1, r2, ..., rn then
                N ← N ∪ r
            else
                remove r
            end
        end
    end
end
foreach region r in N do
    dilate r
    if average intesity of r − average intensity of
     outer boundary < d then
        remove r
    end
end
```

Figure 2 - Algorithm for Nucleus Segmentation [1]

Clump Segmentation

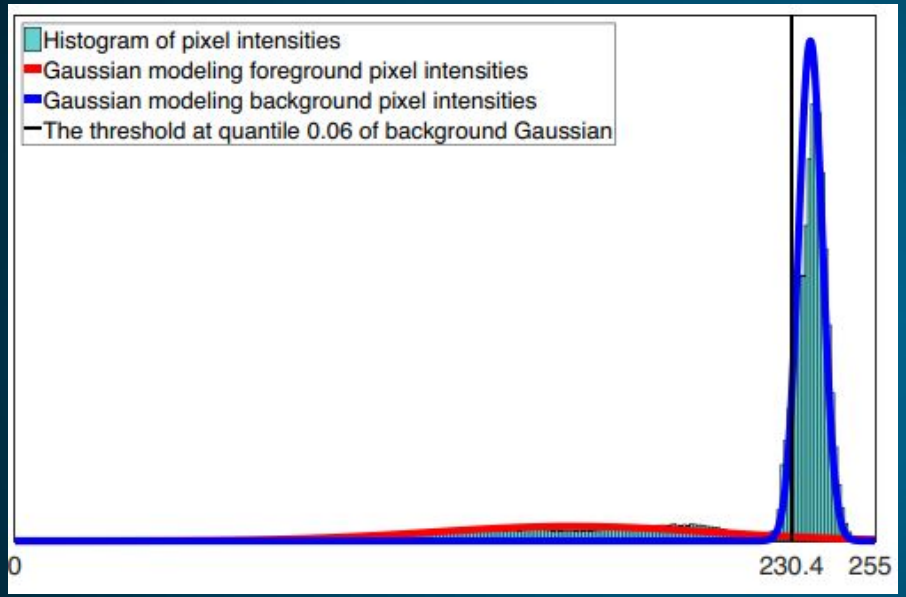


Figure 3 - Histogram of pixel intensities of an EDF cervical cytology image in the training dataset, the estimated Gaussian Mixture Model and the selected threshold. [1]

- Input Parameters:
  - EDF_image
  - Nucleus mask
  - Area
- Outputs a segmented clump mask.

Clump Segmentation

- Original implementation used Gaussian Mixture Model (GMM) alongside Expectation Maximization algorithm (EM).
- Using the EM, one gaussian estimates the foreground and another estimates the background.
- Then using the mean, standard deviation of the background alongside each pixel q, we can calculate the quantile function of the normal distribution which gives us threshold T.
- In our implementation, GMM distribution failed, so we thresholded the image using Otsu thresholding.
- Using T we binarize the image and discard any components that didn't contain nuclei or had a small area.

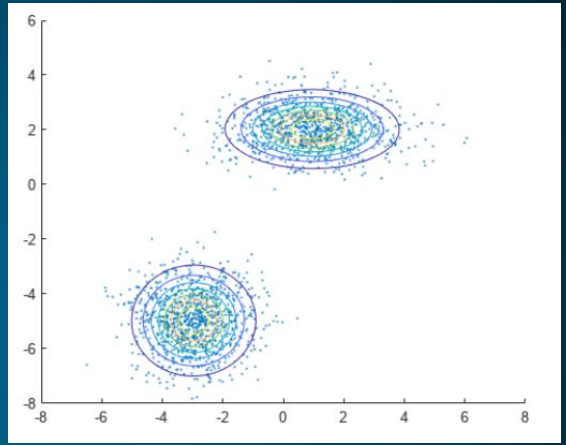


Figure 4 - GMM with EM algorithm plotted

$$Q(q) = \inf\left\{x \in \mathbb{R}, q \leq \frac{1}{\sqrt{2\pi}\sigma_b} e^{(} - (x-\mu_b)^2)/(2\sigma_b^2)\right\}$$
$$= \mu_b + \sqrt{2}\sigma_b \operatorname{erf}^{-1}(2q-1),$$

Figure 5 - Quantile function of the normal distribution used as threshold T

Cytoplasm Segmentation - Boundary Approximation

- Input Parameters:
  - EDF_image
  - Nucleus mask
  - Segmented clump mask
  - Width
  - Alpha
  - Beta
- Outputs a cytoplasm mask.

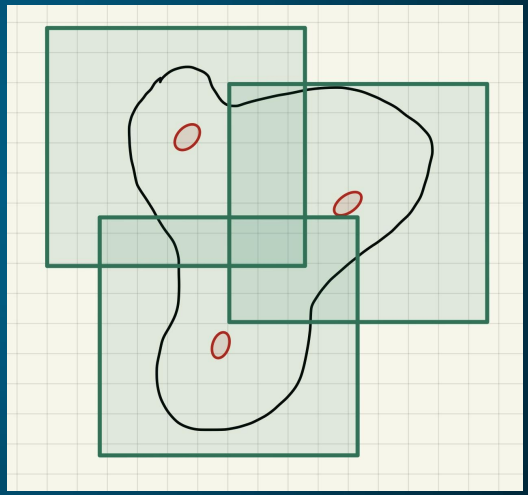


Figure 6 - Generating a Square of Width W on each Nuclei

Cytoplasm Segmentation - Boundary Approximation

- Find all nuclei within each cytoplasm clumps.
- Construct a square of width W around each nuclei of interest.
- Compute the closeness between grid squares i,j and i',j'.
- Using the closeness, compute the likelihood.
- With the likelihood matrix, compute the likelihood that each nucleus belongs to a cell.
- Cytoplasm boundaries are assigned based on likelihood that nucleus belongs to that cell is greater than the sum of the likelihood it belongs to other cells.

$$C_{i,j}^{i',j'} = \sqrt{(i-i')^2 + (j-j')^2}.$$

Figure 7 - Closeness Measure [1]

$$L_{i,j}^{i',j'} = \exp\left(-\frac{C_{i,j}^{i',j'^2} + S_{i,j}^{i',j'^2}}{2\alpha^2}\right)$$

Figure 8 - Likelihood Estimate [1]

$$L_{i,j}^{m} = \frac{1}{m'}\left(\sum_{s=1}^{m'} L_{i,j}^{i_s,j_s}\right)$$

Figure 9 - Nucleus Overlapping Measure [1]

$$\beta L_{i,j}^{m} - \sum_{\substack{n=1 \\ n \neq m}}^{N} L_{i,j}^{n} > 0$$

Figure 10 - Cytoplasm Assignment [1]

Cytoplasm Segmentation - Coarse Refinement

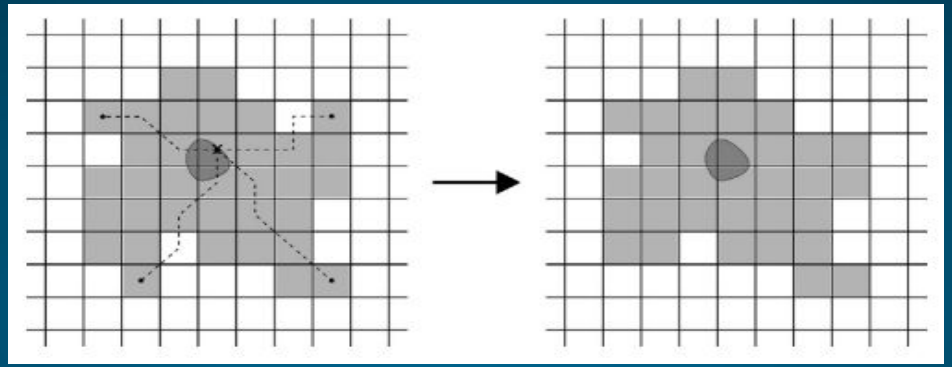


Figure 11 - the subimages not reachable by centroid subimage are removed. [1]

- This defines how “reachable” each pixel is from the nucleus centroid. If the pixel is determined to be not reachable (meaning there exists at least one grid square on the distance segment from nucleus to pixel that is not within the boundary), it is removed.
- Uses the Bresenham [2] algorithm to find the reachability of each pixel.
- Basically, only allows a limited amount of concavity in cytoplasm boundaries.

Cytoplasm Segmentation - Fine Refinement

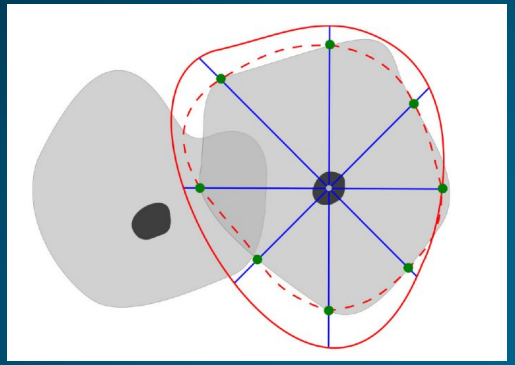


Figure 12: Approximate boundary (solid line) and result of fine refinement (dashed line) [1]

$$W_\theta = \left(\frac{1}{1 + \exp\left(-a\left(\frac{12}{-}\frac{|r_\theta - s|}{r_\theta}\right)\right)} \right)_{s=0}^{2r_\theta}$$

Figure 13: Weight Vector [1]

- Iterative process to find the boundary at a pixel level.
- Remove effect of nuclei on the boundary evolution by replacing each nucleus region's pixels intensity by the mean intensity of its outer boundary.
- In each iteration, the *weight* vector contains the values of the composite of a sigmoid function with thenormailzed distance of points on the radial from the boundary.

RESULTS

RESULTS - NUCLEI SEGMENTATION

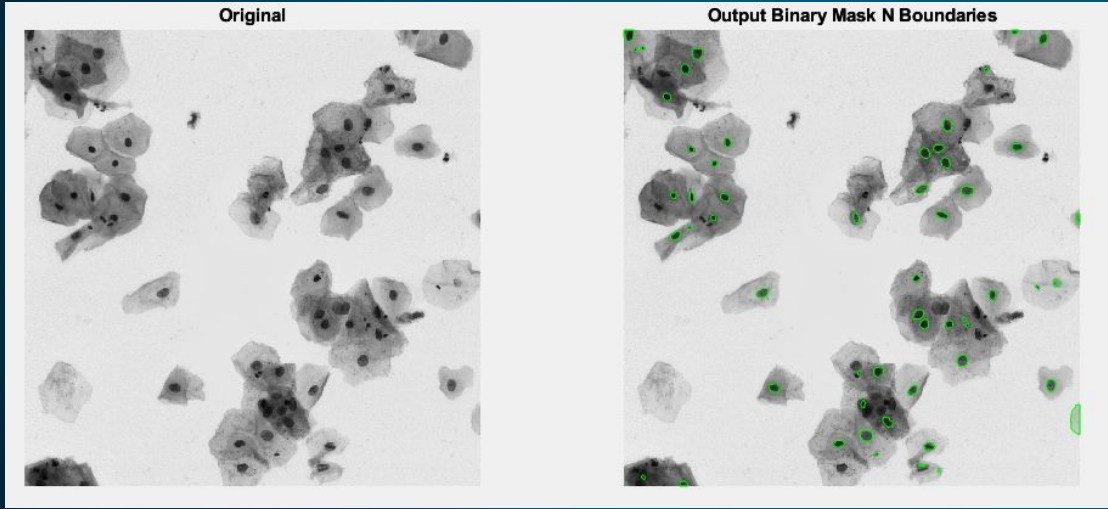
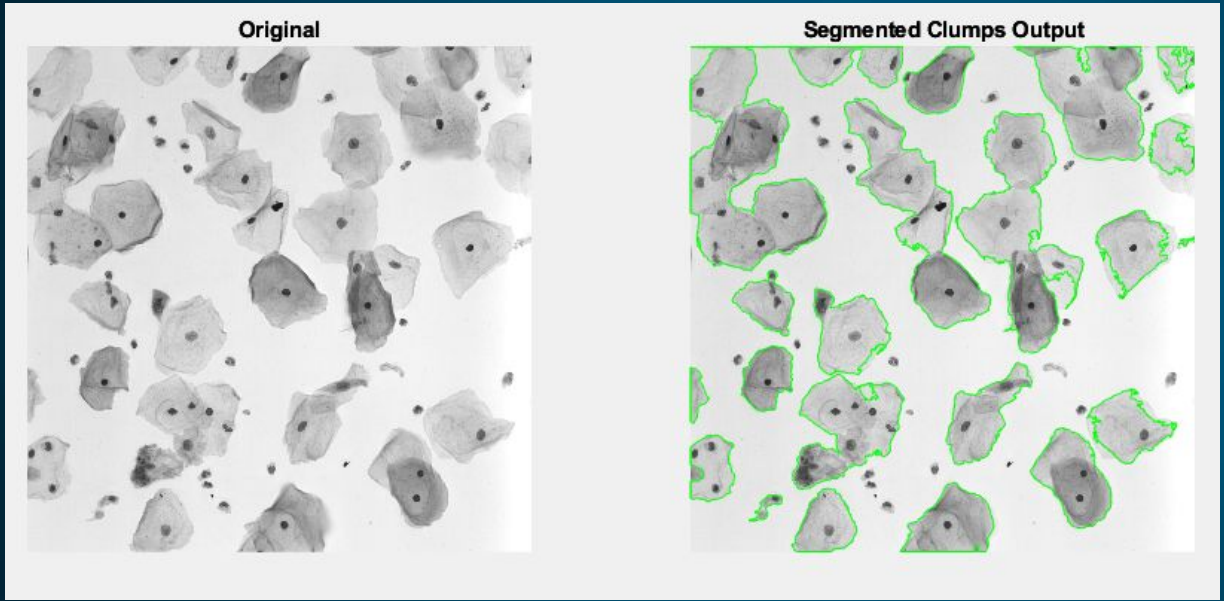


Figure 14: Output of Nuclei Segmentation

Image	# nuclei there should be	# nuclei found	% error
0	37	65	75.68
1	33	46	39.39
2	36	51	41.67
3	29	34	17.24
4	31	34	9.68
5	49	52	6.12
6	26	31	19.23
7	31	39	25.81
8	53	64	20.75
9	48	48	0.00
10	36	38	5.56
11	37	40	8.11
12	43	43	0.00
13	59	39	33.90
14	45	53	17.78
15	52	55	5.77

Figure 15: Results compared to ground truth with % error for each image in dataset

RESULTS - CLUMP SEGMENTATION



Clumps do not include the small, irrelevant nuclei.

Figure 16: Output of Clump Segmentation

RESULTS - CYTOPLASM SEGMENTATION

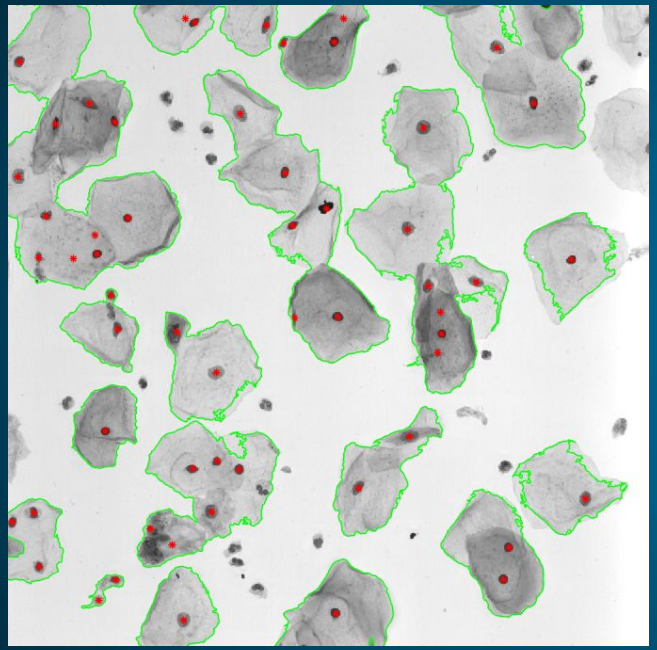


Figure 17 - Finding all nuclei within Cytoplasm Boundaries

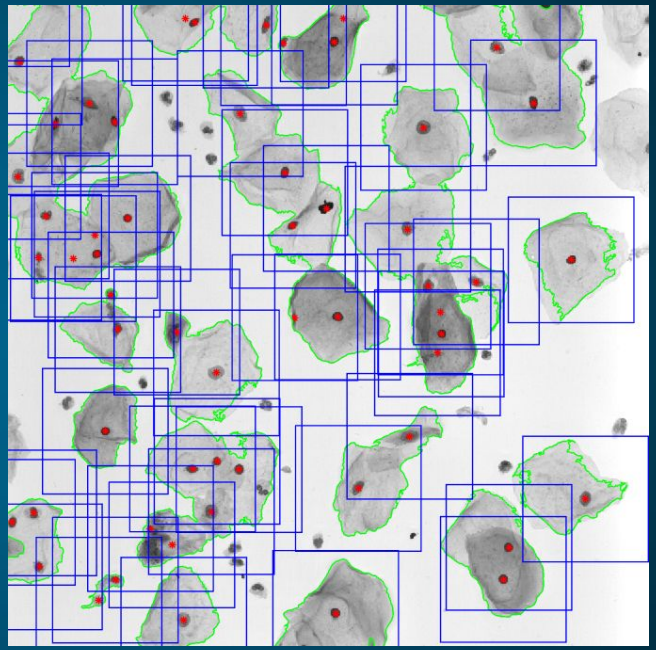


Figure 18 - Constructing Squares of width W on all Nuclei of interest

CONCLUSION - FUTURE WORK

- Successful detection of cytoplasm clumps and individual cell nuclei.
- Partial cytoplasm detection.
- The algorithm works quite well (has low error) for detecting nuclei and segmenting clumps.

Future work:

- Implement coarse refinement and fine refinement.

THANK YOU FOR LISTENING!

QUESTIONS?

REFERENCES

[1] H. A. Phoulady, D. Goldgof, L. O. Hall, and P. R. Mouton, “A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images,” Computerized Medical Imaging and Graphics, vol. 59, pp. 38–49, 2017.

[2] J. E. Bresenham, “Algorithm for computer control of a digital plotter,” IBM Systems Journal, vol. 4, no. 1, pp. 25–30, 1965.