**White Blood Cell (Leukocyte) Identification Progress/ Rough Draft**

1. **Introduction**

Hematologists often investigate blood smears to diagnose diseases by detecting and identifying white blood cells (WBCs). Traditional techniques to do this are somewhat time-consuming and serval methods do exist to automate this process. Companies like CellaVision [1] offer customers proprietary software and equipment to automate the laboratory processes of WBC identification and a simplified solution to perform classification. This paper investigates several methods to automate WBC identification and proposes a new algorithm to detect WBCs from nuclei locations. The algorithm would be useful for lab assistants and hematologists who inspect large sets of blood smears and prefer to use an alternative to current methods such as CellaVision.

*Background*

There exist multiple types of WBCs, each of which can be featured given an image of a blood smear. An example is shown below in figure 1. In the image, there are 6 large purple blobs. These are the white blood cells; however, not all of these cells are the same. There are five different types of white blood cells; each can be distinguished based on their appearance: (1.) Neutrophils, (2.) Lymphocytes, (3.) Monocytes, (4.) Eosinophils, and (5.) Basophils. The different types are labelled in figure 1.



1

32

2

4

1

5

Figure 1 - Different Types of Leukocytes and Test Image [2]

From the above figure, there are 6 large purple blobs, showcasing the five different types of White Blood Cells that exist. Neutrophils can be identified by their pinched-in nucleus (darker purple region) which may have a horseshoe-like appearance. Lymphocytes contain a large, generally round and uniform nucleus. Monocytes are distinguishable based on their kidney bean shape of the nucleus and large amounts of cytoplasm surrounding the nucleus. Eosinophils are similar to Neutrophils in appearance, but with large granules in the cytoplasm. Basophils are the easiest to distinguish as they are characterized by their multiple, large granules. These five types of Leukocytes can then be split into two distinct groups based on the granules;

1.) Granulocytes - Basophil, Eosinophil, and Neutrophil

2.) Agranulocytes - Lymphocytes and Monocytes [3]

1. **Related Work**

Automated Leukocyte identification is a popular topic as current techniques for identifying white blood cells and investigating blood smears are rather time consuming. In Mohammed et al [4], a nuclei segmentation algorithm was proposed using gray scale contrast enhancement and filtering. Sadehian et al. investigated the application of an active contour model for to perform nucleus segmentation and Zack thresholding to identify cytoplasm locations [5]. Prinyakupt and Pluempitwiriyawej investigated segmentation and WBC categorization through the use of linear and naïve Bayes classifiers with preprocessing focusing on nuclei segmentation using the RGB color space [6]. Common image processing topics and methods used to perform Leukocyte identification and segmentation include: edge detection, morphological opening and closing, thresholding (including global thresholding / Otsu’s method), contrast enhancement, and histogram equalization. The methods used and steps of execution are extremely important when detecting WBCs from a blood smear as each play a significant role in the procedure.

Initial approaches in this paper first utilized the nuclei segmentation and WBC identification algorithms from [4] and [6]. However, preliminary tests indicated that they were not as robust as they claimed to be. One image may work great in one, but terribly in the other (and vica-versa). Additionally, the only noticeable adjustably parameter in both algorithms was the structuring element size for the morphological processes. It is desired then, that the proposed algorithm in this paper can be tunable in such a way that a user may only need to manually inspect one to three images, determine the corresponding tuning parameters, and run the rest of their data set without worry.

1. **Data Collection**

The blood smear images used for the duration of this procedure were obtained from various online sources, with the largest data set being contributed by Mostafa Mohamed from Mathworks File Exchange [7]. Other sources for images include The American Society of Hematology (ASH) Image Bank [8] and the Bloodline Image Atlas [9]. While developing the algorithms used, only a handful of images were consisted, whereas testing featured a random batch of 40 – 50 images. The development images where chosen based on their clarity or if certain features were present (either optimal or suboptimal). During testing, the algorithm would be tweaked using the random sample set from Mohamed’s data bank. Several hand selected images from blood line were also to check for robustness. The images from bloodline needed to be hand selected due to the inclusion of unwanted features in some images (eg. strange cross hairs, orange tint, already segmented images etc.).

In the preliminary results (section 5), the image from [2] was used for the initial construction of the algorithm, with current testing be performed using 35+ images from Mohamad’s data set (it is intended to use a batch of 50 while discarding any repeats). It was decided to use Mohamad’s set due to having both a large amount of images, but also consisting of images of strictly blood smears.

1. **Technological Approach**

The proposed algorithm attempts to improve upon the methods suggested in [4] and [6], and is heavily based off the pre-processing procedure in [6]. For both brevity and to prevent redundancy, a flow diagram and subsequent Pseudo code is presented below before breaking down each sub block and corresponding mathematic procedure (note: a more detailed version and explanation will be featured in the final iteration of the report and only what is believed to be the most import steps are outlined here) . It is noted that MATLAB and the corresponding Image Processing Toolbox was used for implementation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| Import Image  Conversion to LAB color space  Color Contrast Enhancement  Conversion back to RGB color space  Channel Operations  Morphological Operations  High Intensity Object Area Filtering  Result Indication   |  | | --- | |  | |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Figure 2 – Technological Approach Flow Chart

Figure 2 is the flow chart corresponding to the proposed algorithm. The pseudo code is as follows.

**Pseudo Code:**

1.) import blood smear image  
  
2.) convert to image to LAB color space  
3.) obtain normalized luminosity values  
4.) adjust luminosty using normalized value by contrast stretching  
5.) multiply adjusted value by max luminosity   
6.) replace the luminosity of the original image with the adjusted one  
7.) convert back to RGB color space  
  
8.) take difference betweem weighted blue and red channels, divide the result by the green channel  
9.) threshold channel operation result by a predefined value  
  
10.) morphologically open thresheld image  
11.) morphologically close image result from opening  
  
12.) for number of high intensity objects in morphologically processed image  
13.) if high intensity object area greater than threshold area  
14.) obtain centroid of object in image  
15.) place circle on original image at the centroid location  
  
16.) return result to user

Presently, only the operations of contrast enhancement and the channel operations are outlined here. A color contrast enhancement was performed in LAB color space, utilizing the luminosity of the image. Note, the enhancement technique used is outline in the tutorial section on Mathworks’ website [10] After converting the stretch image back to RGB color space, the channels are manipulated in such a way that results in obtain a binary image, where high intensity values are featured at WBC nuclei locations.

LAB color space allows all the colors of the spectrum to be accessed, where *L* is the value corresponding to luminance and *a &* *b* are the components for color [11]. Here, only the luminance value is considered. Therefore, once an image was converted, the luminosity values were first normalized so they were in the range from 0 to 1.

**[1]**

Where the maximum obtainable luminance is value and is the luminance value of the image of interest. The contrast stretching procedure then involves setting the luminance value to a range such that 1 % of the pixel intensities of the normalized luminance values are saturated at both the low and high ends, and scaling this result by the maximum value.

**[2]**

Using , the image is converted back to RGB color space where the channel operations occur.

The channel operations involve taking a weighted difference between the blue and red channels, and then dividing the result by the green.

**[3]**

Where *B* are the pixel intensities in the blue channel, *R* are the intensities in the red, and *G* are the intensities in the green. and are the weighting factors of the blue and red channels respectively. For the duration of the experiment and results presenting in this paper, weighting factors of and were used. These were determined through trial and error, and gave sufficient results highlighting the WBCs. The remaining operations used currently are more standard image processing techniques and will be explained in detail for the final paper submission.

1. **Preliminary Results**

The following section documents the current status of the proposed algorithm and the steps performed to arrive at the current status.

Following [6], the grayscale and the individual RGB color channels were investigated.

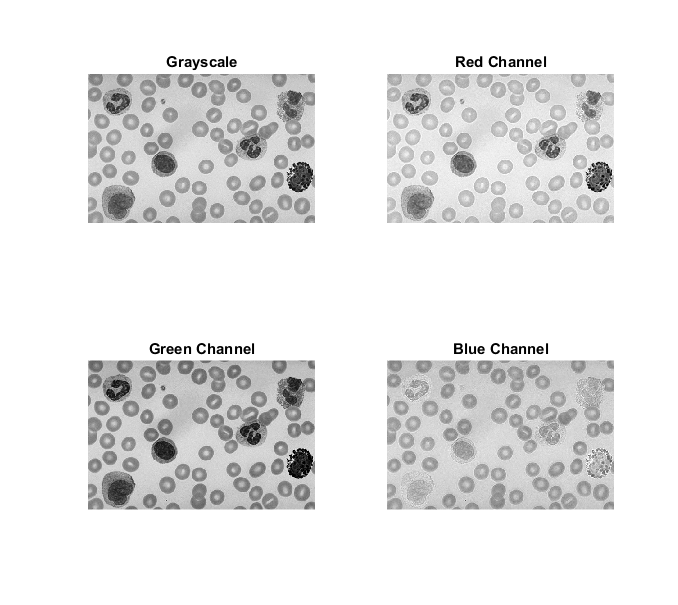
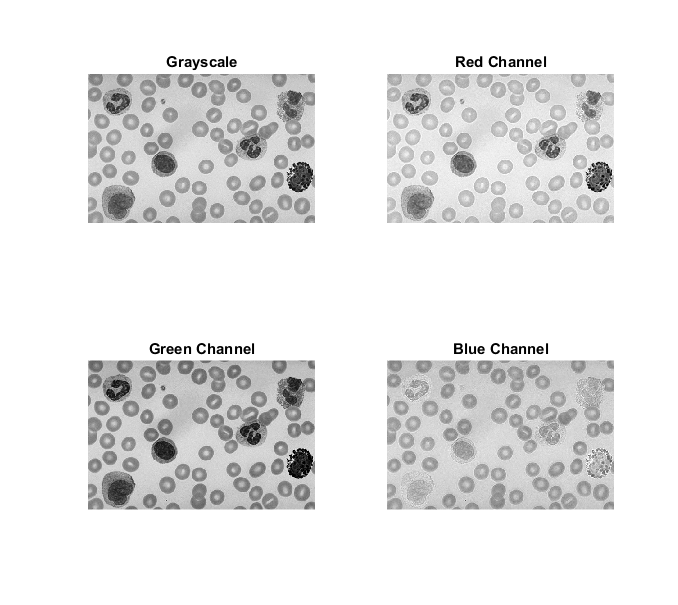


Figure 3 - RGB and Grayscale Channels for Test Image

Figure 3 contains the grayscale, red, green, and blue color channels of the image shown in figure 1. It is observed that in the test image, the WBCs contain large intensities in the blue channel; however that is true for most of the image. This is indicated by the fact that the majority of the image is made up of lighter shades of gray in the blue channel image. Conversely, in the grayscale, red, and green channels, the WBCs are characterized by lower intensity values. The intensity behavior can be attributed to the violet color of the WBCS from the original image. It is noted that the dark characterization however is only true for the nuclei, not the cytoplasm.

Using a similar process as [6], nucleus segmentation was first performed by averaging the red and blue channel intensities. This was done because the green channel contains the lowest pixel intensities of the nucleus (as indicated in figure 2). However, after trying to replicate the rest of Prinyakupt and Pluempitiwiriyawej’s algorithm (PPA), the results were not as robust as initially anticipated. When using selected images from Bloodline and the data set by Mohamed, there were large inaccuracies. Results varied substantially from not being able to identify any white blood cells, to picking out way more than what should be present. This is demonstrated below with several images, without presenting the details of PPA.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Figure 4 - PPA Results

Based on the obtained results, it is observed that several errors occur. The over estimation of the WBCs can be attributed to the structuring element size of the morphological proceses, however, the underestimation is of concer. It is noted that the several of the images in figure 4 were extracted from Mohamed’s data set.

Investigating the underestimation results from figure 4, a random image from Mohamed’s data set was considered.



Figure 5 - Mohamed Random Image, PPA

Figure 5 is a random image from Mohamed’s image set. From observation, it is clear that there is one WBC is the image. Thus, following along with PPA, the RGB color channels for the image in figure 5 were inspected.





Figure 6 - RGB and Grayscale for PPA assessment

Similar to figure 3, the RGB color channels are presented for the image in question. The high hintensities of the blue channel are observered, however, the low intensity responses for both the red and green channels are not as strong as in figure 3. It is noted that this was not just a one time occurance, and was observed for many images using PPA. To next step in PPA was to take the color channel adverage and then perform histogram equalization on the result. The produced image for performtion this operation on image considered in figure 5 is shown below.



Figure 7 - RGB Avg. Histogram Equalization

Figure 7 is the obtained image after performing histogram equalization, as suggested by PPA. It is immediately apparent that PPA will not work and alternative approach is needed instead. The reason why such a poor image is obtained in figure 7 is believed to be caused by the both the blurriness of the blood smear and the weak contrast of the WBCs. As mentioned before, this is not unique to just this image and does occur in for many cases. Using this information, the preprocessing procedure of PPA was modified.

The proposed algorithm instead uses the idea of initially applying a contrast enhancement to the image, similarly to the solution introduced in Mohamed et al. [4]. Capitalizing on the strong violet colors of the nuclei, a color contrast enhancement was performed in LAB color space where the luminosity value was manipulated to adjust the image [10]. After adjustment, the resulting image is obtained.

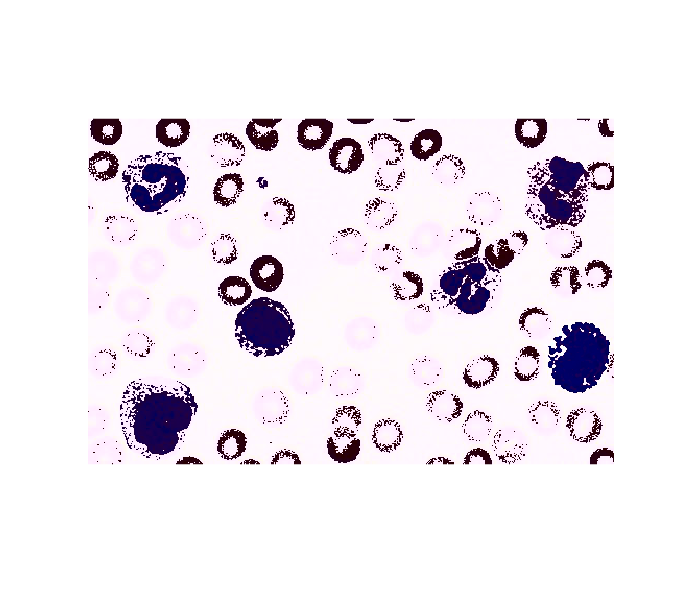
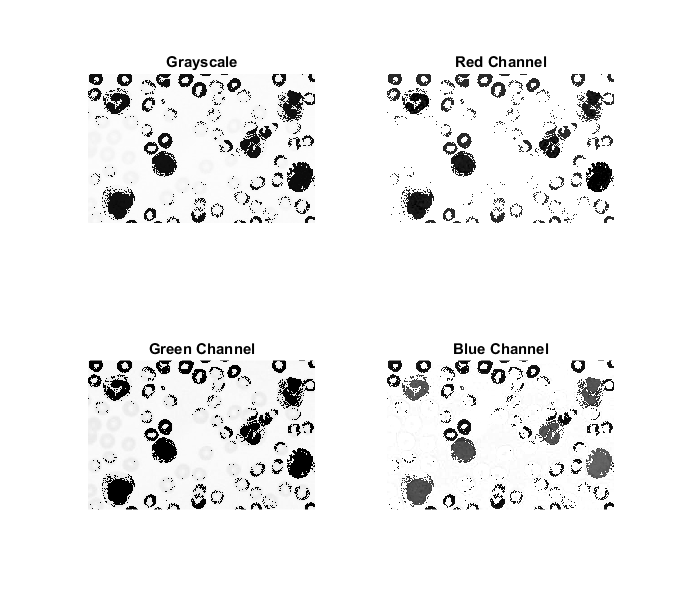


Figure 8 - Contrast Enhancement

Immediately, it is observed in figure 8 that the nuclei of the WBCs now appear to have a strong, dark blue color. It is noted the difference in color between the nucleus and the rest of the image is stronger than before. Similarly to PPA and due to the occurrence of the change in color, the RGB color channels were investigated once again. This is presented below.



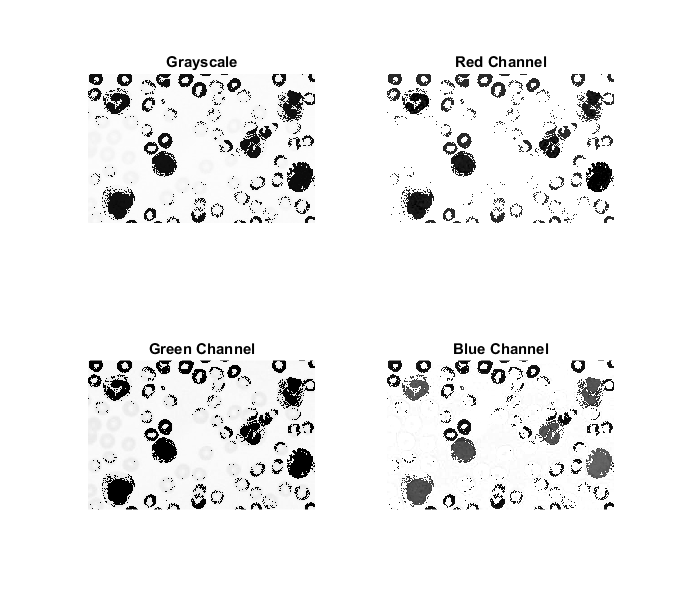


Figure 9 - RGB using Proposed

Figure 9 contains the grayscale rendition of the enhanced image in addition to each of the respected RBG color channels. The obtained results were not as expected as the blue channel does not feature as strong of a response that was hoped for at the nuclei locations. Instead a very strong intensity occurs in the area where neither red nor white blood cells lie. In addition, both the red and green channels contain very low intensities at nuclei locations, with high intensity values in the unoccupied areas. Using the information obtained from figure 9, a similar approach as PPA was deployed. As opposed to averaging the pixel values however, it was decided to weight red channel and divide the difference between the weighted red and blue channel by the green channel (as presented in the previous section). This resulted in having only high pixel intensity values of a grayscale image remaining at the WBC locations. Since operations were performed using a uint8 data type in MATLAB, results were typically bounded by 0 or 255. However, to ensure that the produced intensities formed a binary image, a threshold at a pixel intensity of 150 was performed. The threshold level was determined during initial implementation of PPA and it was decided to use the threshold for the proposed algorithm as well. The resulting image after channel operations and thresholding is presented below.

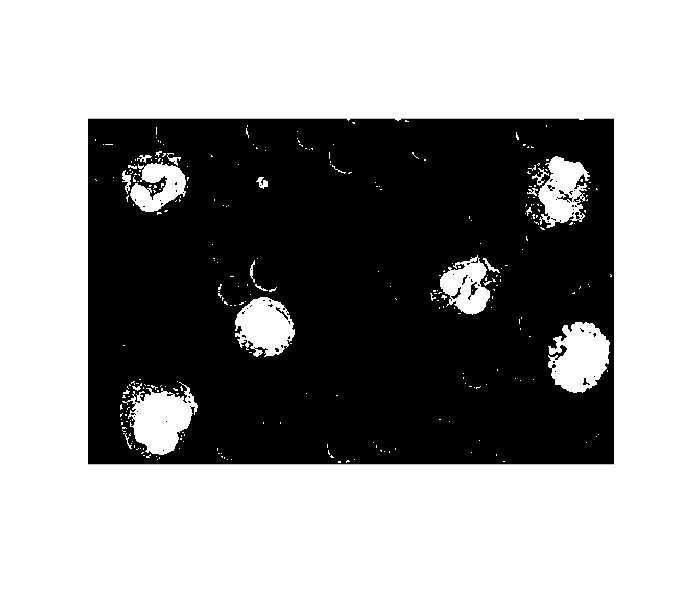


Figure 10 - Binary Image with Proposed

Figure 10 is the binary image after performing color channel operations and thresholding the enhanced image. From inspection, all 6 Leukocytes remain. Although the above image produces similar results to PPA, the strength of the proposed algorithm shines when using Mohamed's data set.

Once the binary image was obtained, morphological transformations were implemented to remove any unwanted small objects or small holes from the nuclei, hence both the opening and closing processes. The constructed algorithm is designed to be robust so a user may be able to adjust the structuring element sizes involved during the morphological processes. This may be changed depending on the resolution of the imported image(s) and the relative size of the WBS(s). The effects of applying morphological opening and subsequent closing are present below.



Figure 11 - Morphological Processing

Figure 11 is the morphologically transformed binary image using the thresholded image as an input. Based on the image obtained and comparing it with figure 6, it is seen that while both pixel islands and holes have been removed, some distinguishing characteristics of the WBC nuclei have also been lost. It is desired to possibly address this issue before completion. Removing certain characteristics of the WBCs will be problematic when trying to perform classification of the different WBC types. Despite the lost in characteristics, the proposed algorithm is still able to perform WBC identification (no classification presently). The last step of the process in the algorithm is filtering out high intensity objects based on their relative area in the image and a tuning parameter set by the user corresponding to data set being used. If the high intensity object area is less than a certain area, the object is mapped to a low intensity and is not considered a WBC. Once this is done, the results are indicated on the original, imported image and relayed back to the user. The results of the algorithm are shown below.

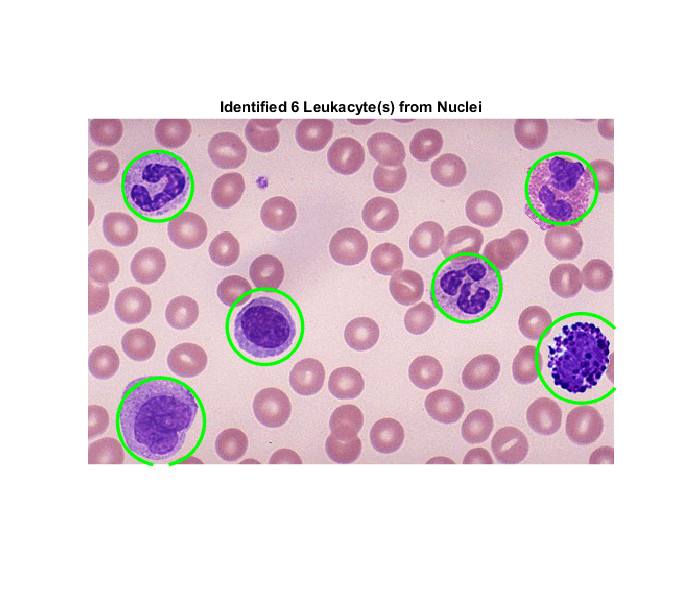


Figure 12 - Algorithm Results

From figure 12, the produced algorithm is capable of performing WBC identification given an image.

After some parameter tweaking, the proposed algorithm was initially tested using a selected set of images from Bloodline.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 13 - Selected Bloodline Test

Figure 13 demonstrates the proposed algorithm’s ability to identify WBCs if a different tint is used. The above results are acceptable.

In order to really evaluate the current performance of the proposed algorithum, a large data set from Mohamed’s image bank was used The images that were selected based on the ability to quickly import and construct a random batch from the image set. Stucturing sizes and parameters were also tunned after looking at results from three random images. The result of using a batch size of 42 images is shown below.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 14 - Mohamed Data Test

Figure 14 presents the preliminary results of applying the proposed algorithm to a random batch of Mohamed’s data set. Based on the obtained results, the proposed algorithm works fairly well for the above data set of 42 blood smear images. Out of the 42 tested images, only three appear to produce non ideal results. These are indicated above by the yellow outline of the images. The images do not produce desired responses feature underdeveloped WBCs, unique overlapping of nuclei and granules, as well as less lower intensity value for the nucleus location. Each of these issues will need to be inspected before completion.

1. **Plan for Completion and Further Work**

Listed below are tasks that are desired to be performed before project complete

* Quick and dirty segmentation using dilate
  + If results not acceptable, use Moore’s Boundary following algorithm
* Need to include some differentiation between cell types. Could possible use some feature detection, area of the high intensity objects in the binary image. May need to look at more literature. This is a significate portion of the proposed project, however preliminary results suggest that this may be more difficult than initially believed. In addition, the proposed algorithm is very similar to the preprocessing in PPA and feels that it is lacking in terms of content.
* Include some form of validation / statistical analysis of the obtained results to evaluate algorithms more than just, “this is good / bad.” Look up and possibly apply Sørensen–Dice index for similarity as this has come up numerous times in literature. Even if the similarity index is not included, there needs to be some form of evaluation.

1. **References**

|  |  |
| --- | --- |
| [1] | "CellaVision," 2017. [Online]. Available: http://www.cellavision.com/en/. |
| [2] | [Online]. Available: http://library.med.utah.edu/WebPath/HEMEHTML/HEME100.html. |
| [3] | "Identify Normal Leukocytes," 19 December 2011. [Online]. Available: http://www.pathologystudent.com/?p=4776. |
| [4] | M. Mohamed, B. Far and A. Guaily, "An Efficient Technique for White Blood Cells Nuclei," *2012 IEEE International Conference on Systems, Man, and Cybernetics,* pp. 220-225, 14-17 October 2012. |
| [5] | F. Sadeghian, Z. Seman, A. R. Ramli, B. H. A. Kahar and M.-I. Saripan, "A Framework for White Blood Cell Segmentation," *Biological Procedures Online,* pp. 196-206, 2009. |
| [6] | J. Prinyakupt and C. Pluempitiwiriyawej, "Segmentation of white blood cells and comparision of cell morphology by linear and naïve Bayes classifiers," *BioMed Central,* 2015. |
| [7] | M. Mohamed, "An Efficient Technique for White Blood Cells Nuclei Automatic Segmentation," [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/36634-an-efficient-technique-for-white-blood-cells-nuclei-automatic-segmentation. |
| [8] | "ASH Image Bank," [Online]. Available: https://imagebank.hematology.org/atlas-images/list/. |
| [9] | "Image Atlas Categorical Index - White Blood Cells," [Online]. Available: http://www.bloodline.net/imageatlas/white-blood-cells/. |
| [10] | "Contrast Enhancement Techniques," [Online]. Available: https://www.mathworks.com/help/images/examples/contrast-enhancement-techniques.html. |
| [11] | "Lab Color," Mathworks, [Online]. Available: https://www.mathworks.com/discovery/lab-color.html. |

1. **Appendix – MATLAB Script for Current Proposed Algorithm**

function [] = wbcNuclei\_v06(im, OP, CL, tune, num)  
%  
% wbcNuclei: Identify Leukocytes based on nulcei.  
%  
% INPUT:  
% OUTPUT: [] - figure  
%  
  
srgb2lab = makecform('srgb2lab');  
lab2srgb = makecform('lab2srgb');  
imlab = applycform(im, srgb2lab); % convert to L\*a\*b\*  
max\_luminosity = 100;  
L = imlab(:,:,1)/max\_luminosity;  
% contrast stretching with luminance  
imlab\_adjust = imlab;  
imlab\_adjust(:,:,1) = imadjust(L)\*max\_luminosity;  
imlab\_adjust = applycform(imlab\_adjust, lab2srgb);  
R = imlab\_adjust(:,:,1); % red channel  
G = imlab\_adjust(:,:,2); % green channel  
B = imlab\_adjust(:,:,3); % blue channel  
nucl = ((1\*B)-(0.75\*R))./(G);  
level = 150;  
nuclBW = nucl > level;  
sqOpen = strel('disk',OP);  
nuclMorph = imopen(nuclBW,sqOpen);  
sqClose = strel('disk', CL);  
bw = imclose(nuclMorph,sqClose);  
r = regionprops(logical(bw)); % obtain properties of the resulting image  
% display the original image with the thresheld cross-corelation marked  
figure(num);  
imshow(im,[]);  
[m, n] = size(bw);  
cnt = 0;  
hold on  
for i = 1:length(r)  
 if r(i).Area>((m\*n)/tune)  
 n = 3.\*sqrt(r(i).Area/pi());  
 rectangle('position',[r(i).Centroid(1)-n/2,r(i).Centroid(2)-n/2,n,n]...  
 ,'Curvature',[1 1],'EdgeColor','g','LineWidth',2.5)  
 cnt = cnt + 1;  
 end  
end  
title(['Identified ',num2str(cnt), ' Leukacyte(s) from Nuclei'])  
end