May 2016

A Study on Automated Process for Extracting White Blood Cellular Data from Microscopic Digital Injured Skeletal Muscle Images

Bibek Karki
University of Nevada, Las Vegas

A STUDY ON AUTOMATED PROCESS FOR EXTRACTING WHITE BLOOD CELLULAR DATAFROM MICROSCOPIC DIGITAL INJURED SKELETAL MUSCLE IMAGES

By

Bibek Jang Karki

Bachelor of Electronics and Communication Engineering
Tribhuvan University, Nepal
2012

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science in Engineering- Electrical Engineering

Department of Electrical and Computer Engineering
Howard R. Hughes College of Engineering
The Graduate College

University of Nevada, Las Vegas
May 2016

UNLV | GRADUATE COLLEGE

Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

April 29, 2016

This thesis prepared by

Bibek Jang Karki

entitled

A Study on Automated Process for Extracting White Blood Cellular Data from Microscopic Digital Injured Skeletal Muscle Images

is approved in partial fulfillment of the requirements for the degree of

Master of Science in Engineering- Electrical Engineering
Department of Electrical and Computer Engineering

Mei Yang, Ph.D.
Examination Committee Chair

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

Venkatesan Muthukumar, Ph.D.
Examination Committee Member

Yingtao Jiang, Ph.D.
Examination Committee Member

Barbara St.Pierre Schneider, Ph.D.
Graduate College Faculty Representative

Abstract

Skeletal muscle injury is one of the common injuries caused by high-intensity sports activities, military related works and natural disasters. In order to discover better therapies, it is important to study muscle regeneration process. Muscle regeneration process tracking is the act of monitoring injured tissue section over time, noting white blood cell behavior and cell-fiber relations. A large number of microscopic images are taken for tracking muscle regeneration process over multiple time instances. Currently, manual approach is widely used to analyze a microscopic image of muscle cross section, which is time consuming, tedious and buggy.

Automation of this research methodology is essential to process big amount of data. The objective of this thesis is to develop a framework to track the regeneration process automatically. The framework includes dynamic thresholding, morphological processing, and feature extraction.Based on the clinical assumptions, the threshold is calculated using standard deviation and mean of probable single cells. After thresholding, six parameters are calculated including average size, cell count, cell area density, cell count on the basis of size, the number of cytoplasmic and membrane cells as well as the average distance between cellular objects. All these parameters are estimated over the time, which helped to note the pattern of change in leukocytes (White blood cells) presence. Based on these results, a clear pattern of leukocyte movement is observed. Our future work includes deriving the mathematical equations using regression model on the data set of increased time points.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Mei Yang for providing me with all the guidance, motivation, and supervision to complete my research work. I could not have imagined having a better advisor for my research work and other academic objectives.

I am also grateful to my thesis committee members Dr. Yingtao Jiang and Dr. Venkatesan Muthukumar, for being part of the committee and providing advice and encouragement.

I am very grateful to Dr. Barbara St. Pierre Schneider from health science department for her continuous support on my work. Her guidance and immense knowledge on the clinical part of muscle injury helped me during the period of research and writing thesis.I am thankful to Hananeh Derakhshan for helping to capture the microscopic digital images. The images were obtained from samples were generated from a project supported by the Department of Defense, Air Force (Grant # FA7014-10-2-0001). Review of material does not imply the endorsement of factual accuracy of opinion by the Department of Defense, Air Force.

Finally, I would like to thank all my friends, family and seniors for the help and continuous encouragement for completing this thesis.

Bibek Jang Karki

University of Nevada, Las Vegas
May 2016

Table of Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

Skeletal muscle occupies nearly half of the human body weight in health adults. High-intensity sports activities military and natural disaster-related injuries can cause skeletal muscle injury. The injured muscle goes through different stages of healing which includes degeneration, inflammation, regeneration and fibrosis. After the injury, monocytes and macrophages initiate repair as a part of an inflammatory response, which is the activation of muscle regeneration [2].

Skeletal muscles are composed of individual muscle fibers.A skeletal muscle injury composes the majority of sports, military and natural disaster-related injuries. Muscle injuries are the result of direct processes (involves lacerations and strains) and indirect processes (involves ischemia and neurological dysfunctions) [3]. Also, they can be induced by a pressure applied to skeletal muscles, which interrupts blood flow damaging cell membranes [4]. Recovery of skeletal muscle injuries mostly follows three phases: acute inflammatory and degenerative phase, repair phase and the remodeling phase.

Since skeletal muscle injury constitutes the majority of injuries, research to find possible therapies via automatic framework is essential. Several research projects are ongoing to study muscle inflammation and track muscle regeneration process using microscopic images.The white blood cells (WBC) play an important role in the regeneration of injured muscle fibers by repairing the damaged fibers. Therefore information extraction of WBCs is valuable. Microscopic analysis of leukocytes (WBCS) s one of the most used laboratory investigation for assessing muscle regeneration. These microscopic images are preprocessed and thresholded to identify leukocytes presence. Also, analysing size, shapes, numbers and sometimes even the color of cells or organisms (parasites) can help to identify parasites, degree infection and appropriate medication [5].

1.1 Overview of Muscle Regeneration Process Tracking

Muscle regeneration process tracking is the act of monitoring injured tissue section over time, noting cell behavior and cell-fiber relations (see Fig. 1.1). Cell concentration around injured part over time gives cell behavior which can be related to healing mechanism. Also, cell and fiber relation over time can be estimated and analyzed and relate it to healing procedure.

The microscopic images of the injured muscle section are taken at different point of time, to assess the stage of healing. The sections taken are stained with hematoxylin and eosin for histologic assessment of fiber damage and cell infiltration. The analysis on the change of fiber and cells morphology over healing time will help to investigate the healing process and find effective therapies. Since there can be different types of injuries, animal models, and antibodies, processing and analysis have to be performed on big image data. Automation helps in assess with the high degree of certainty and allow statistical analysis in large scale as compared to a manual approach.

Biomedical image data mining requires proper preprocessing (ie. image enhancement and segmentation, feature extraction, classification). However, gray level histograms are not always bimodal, methods other than valley seeking are thus required to solve thresholding problem. The variation in shape, size and imbalanced illumination in microscopic images makes the thresholding more difficult.

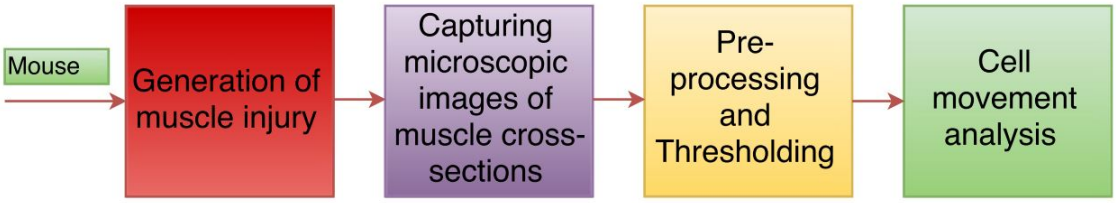


Figure 1.1: An overview of muscle regeneration tracking process

Animal models are arranged to study a pathophysiology of muscle inflammation and regeneration. The mice, which are free of ectoparasites, helminth endo parasites, antibodies to 19 murine viruses, were randomly assigned (24,48 hr) injury group. They are kept on a crush injury device platform to induce injury on the medial or lateral surface of the gastrocnemius muscle. After the generation of injury, muscles are allowed to recover for 24 to 96 hrs. These are time points when images of muscle cross sections are recorded using a microscope. In the meantime, regular assessment on the daily basis by a veterinarian is carried out to check unrelieved pain such as piloerection of fur, over grooming of injured limb, and abnormal posture [4].

These images are thresholded with different threshold values calculated using standard deviation and

mean of probable single cells. First, the image is converted to binary using Otsu threshold and single probable cells are identified. The assumptions made by the experts on the field are implemented mathematically to find single cells. A new threshold is estimated using the mean and standard deviation of gray values of single cells.The binary image obtained is further processed using morphological operation such fill, dilation depending upon the algorithms used. The shape, size, and intensity values are recorded for the further analysis of leukocytes movement.

1.2 Motivation

Microscopic images are one of the most important resources to use for diagnosis and decision making on diseases. It requires analyzing the big amount of images and samples to assess the problem and discover better recovery procedure. Currently, the diagnosis and analysis on these images are mainly performed manually with the help of medical personnel.

Since manual classification requires direct visual inspection, the process is prone to error and directly depends on the experience of a medical researcher. In addition, the medical devices used for cell analysis are expensive and they may not be found in every medical laboratory. This ineffective technique is highly time-consuming and tedious. And also, there is a possibility of different assumptions and interpretations on the same sample by different medical researchers. A medical personnel s data input may not have uniformity. As we know medical studies require data with high accuracy and reliability for decision making and interpretation.

These problems motivated me to develop an automated muscle regeneration tracking process which can be more efficient, accurate, and less costly. Since there is a huge advancement in computing and camera technology, it is important for computer scientists and engineers like me to replace the manual process to a computer based automatic process. The Automation, in this case, includes collecting information of the interested region in an image by extracting object parameters (size,area,orientation) that help automatic diagnosis and evaluation of treatment progress. This automated process can eventually support on discovering effective and speedy therapies.

1.3 Objective

The main objective of our research is to develop an automated method for tracking muscle regeneration process after the injury. These are the proposed tasks to achieve the objective:

1. **Dynamic thresholding and Cell detection:** An efficient and accurate method to threshold unimodal microscopic images will be developed. An approach to identify different types of cell will be

developed. The algorithm will be based upon intensity values, shape and texture

2. **Morphological processing and result analysis:** After the detection of cells in all differently timed images (24, 48 , 96 , 192 hrs), a method will be developed to calculate values of different parameters such as density, cell counts, average size, cell type counts. This information with aid of biological knowledge will be implemented to find possible healing equation for different proteins.

1.4 Outline

The rest of the thesis is organized as follows.
In chapter 2, we will briefly review the muscle regeneration process and review existing methods of thresholding microscopic images. We will discuss how the thresholded image is further operated analyzed morphologically and statistically to obtain data related to cell behavior and cell-fiber relations.

In chapter 3, we will discuss our approach to solve the thresholding problem for unimodal microscopic images. The method of using standard deviation and mean for setting threshold will be discussed. We will further show the implementation of different parameters such as density of cell area, the average size of a cell, cell type counts, and the average distance between cells. Also, the data analysis technique to extract medical information from the set of images will be presented.

In chapter 4, we will present and discuss the results of our work. The results of each parameter are analyzed and discussed on medical information that can be extracted from mathematical values. We will do a comparative study of how antibodies are performing on the basis of calculated parameters.

In chapter 5, the summary and conclusion of work will be given. Besides, we will discuss how this work can be improved in future.

Chapter 2

Literature Review

In studying status of muscle injury, different changes such as muscles mass, biochemical changes, and morphological parameters are evaluated. Different research work is carried out to study cell patterns for a number of health applications. Manual assumptions are implemented in framework development to detect stages of diseases, assessing healing process. Besides that, a comprehensive study of injury and healing procedure helps to find the best therapy.

Medical images go through preprocessing stages before implementing thresholding. Thresholding or segmentation is one of the important steps which removes noise or unwanted objects from an image. The thresholding is followed by morphological operations and evaluations of the region of interest (ROI). Also, use of classifiers is increasing to improve the accuracy and solve problems of conventional procedure.

The muscular diseases mostly manifest themselves as an abnormality in terms of changes in nuclei and cytoplasm of muscle fibers (muscle cells. Researchers [6] have developed an image processing framework which automatically analyzes images and quantify morphology of muscle fibers. They used to two different processing techniques to identify and analyze cytoplasm and nuclei. Their process includes adaptive thresholding, distance transform, peak finding and centroid calculation for nuclei channel and fill, skeletonization and morphological analysis after adaptive thresholding for cytoplasm channel.

2.1 Thresholding

Thresholding is a simple method to segment an image into two regions (black and white). It creates binary images from gray-levels by replacing pixel values with zero or one on the reference of a threshold value (T). Each pixel is replaced with black pixel if pixel intensity is less the threshold constant T or else with a white pixel (see Algorithm 2.1).

Algorithm 2.1: Thresholding

Data: Image (i,j)

Result: *Thresholded image (Imageth)*

```
foreach pixel of image do
    if Image(i,j) < T then
        Imageth(i,j) = 0;
    else
        Imageth(i,j) = 1;
```

Thresholding is a crucial task in developing automatic image processing framework, which provides very important information for further evaluation. It helps to find objects of interest which eventually helps on cell(object) tracking, shape, and size analysis. A numbers of segmentation algorithms have been developed to cover a wide range of areas and applications.

Different factors that affect the choice of threshold techniques are:

- Separation between the peaks in histogram.
- Intensity uniformity.
- Relative size of objects and background.
- Noise content in image.

Automated threshold estimation is a good thresholding choice when there is a clear valley between two peaks (modes) in image histogram. This iterative approach with small predefined parameter (delta T) will take infinite loops in case single mood histogram types (see Algorithm 2.2).

Algorithm 2.2: Automated Threshold Estimation

Data: Image (i,j)

Result: *Thresholded image*

`/* T- threshold value */`

1. *Select an initial global threshold T.*
2. *Segment image using constant T (A1 pixel group with intensity values ¿T, and A2 with intensity values ¡=T)*
3. *Compute average m1 and m2 for each group A1 and A2 respectively*
4. *Compute threshold ($T = \frac{1}{2}$ (m1+m2)*
5. *Repeat from step 2 to step 4 until the difference between successive iterations goes lower than a predefined parameter (δT).*
6. *Threshold with final value using Algorithm 2.1*

Otsus approach is one of the powerful methods of gray scale thresholding. It operates on the gray scale histogram making the computation fast. Otsus method makes assumptions on bimodal histogram distribution and uniform illumination [1].Chronic lymphocytic leukemia (CLL) cells are abnormal lymphocytes, which are slightly larger than the resting lymphocytes. The optimal thresholding (Otsu method), watershed segmentation and morphological operations can be used to identify CLL cells. Out of the tested 140 microscopic images, algorithms based on these methods give 99.92 % maximum accuracy for nucleus segmentation and 99.85 % maximum accuracy for cell segmentation [7].

Entropy-based thresholding uses information theory to find probability distribution on the basis of the gray level image histogram. This method works better on the noise images with gray level histogram possessing unimodality. The adjustable Tsallis coefficient q effectively works as a tuning parameter for the same class of images, assisting in automatic thresholding [8]. In 1985,Kapur and his team [9] used two probability distributions, for both foreground and background to compute a threshold. This algorithm uses a global and objective property of gray level histogram of an image.

Some researchers have used the different pre-processing technique to compute correct threshold value. Authors of [10] use a different method which adds and subtracts the processed image with the original image.

$$R_1 = \text{L+H}$$
$$R_2 = \text{R1-H}$$
$$R_3 = \text{R1+R2}$$

where, L-enhanced with linear contrast stretching, H- histogram equalized image.

R3 removes most of the non-interested cellular objects. After this step, Otsu thresholding correctly segments since the image R3 is bimodal. Their result was able to obtain accuracy between 85 % -98 %. Similarly, the work in [5] has been done to identify human intestinal parasites using image processing methods. The images are preprocessed using contrast enhancement, edge detection, median filtering and basic morphological operations. The F-SDTS classifier is used for the final detection of parasites type. Their results show 95% of accuracy on finding parasites ALO and TTO.

Image thresholding and segmentation algorithms are ad-hoc based; their efficiency differs from one image to another image and also is based on the region and properties of interest area [11]. A research group[12] has introduced a new cell cluster segmentation algorithm based on local and global thresholding. The maximum-likelihood algorithm is used for setting global threshold where the local threshold is used to improve the segmented region of clusters defined by the global threshold. Their experimental results show the average processing time of 2.62 seconds and 82.26 % improvement accuracy in segmentation.

A boundary break point connection can be used to connect cell area missed during segmentation.Then, the hole is filled using morhpological operation to recover the lost cell area. An increase in recognition rate is found and the process takes about 10 seconds [13]. A Research in [14] is conducted using Haematoxyling Eosin DAB (HED) color space followed by bilateral filtering, canny edge detection, and watershed algorithm to count lymphocytes in microscopic images. They obtain different range of accuracy after implementing each of these algorithms, canny edge, and watershed segmentation had 90.99% accuracy.

Beside global and adaptive thresholding, the classifiers are also useful for finding interested objects in microscopic images. Authors in [15] have used Local Bayesian classifiers for cell segmentation, wherein classifier are trained using clustered training image patches. A mixture of experts model is used to find likelihood of any new pixel to be a background. They experimented with four different types of microcopy images with 92.5 % accuracy. To solve the same problem of human error on counting cells, an algorithm [16] for color based segmentation of white blood cells is developed. A rough set is created for the objects classifying under(wlow) and upper(wupper) approximation on the basis of possibility. And, the rough k-means clustering based on three parameters (wlow, wupper, threshold(T)) is used for white blood cell segmentation to assist in acute leukemia detection.

Microscopic image histograms are mostly unimodal, the popular thresholding method like otsu [1] alone generally provides poor results [15]. Thresholding becomes difficult due to one peak (see Fig. 2.1 to Fig. 2.6). A research team [17] uses standard deviation and mean of the image pixels to compute global threshold. They use final thresholded image for counting hepatocytes in microscopic images. Their automatic counting algorithm has achieved an average accuracy of 95 % for single cell counting and 88 % for total cell counting.

The global and local threshold values can be estimated using Maximu m likelihood thresholding technique. First the global threshold is set,then the local thresholding is done for each cluster. An algorithm based on RANSAC is used to estimate the local threshold. This method developed for a real time cell density (in-situ microscope) show an average processing time of 2.62 seconds and 82.26 % of improvement in the segmentation accuracy [18].

we developed a method that uses Otsu thresholding followed by morphological operations and statistical analysis of gray values. The possible cells are used for the analysis and manual assumptions are implemented in our algorithm to find possible singe cells. The final threshold is then estimated as the mean+ standard deviation from these cells.

Figure 2.1: A bimodal sample image

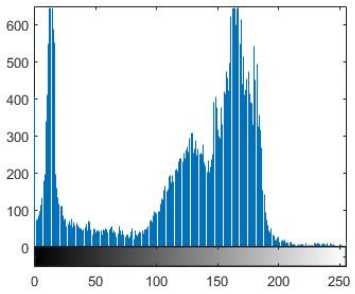


Figure 2.2: Image histogram

Figure 2.3: Binary image after Otsu [1] thresholding

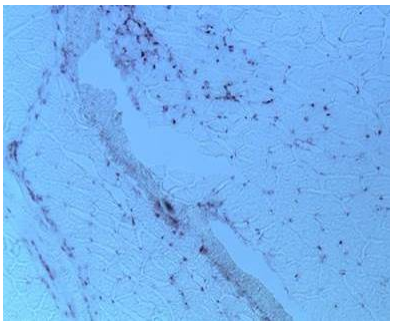


Figure 2.4: A unimodal microscopic image

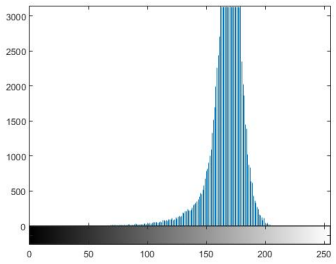


Figure 2.5: Image histogram

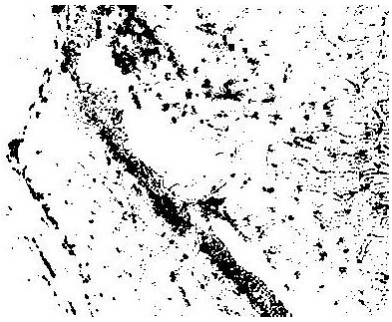


Figure 2.6: Binary image after Otsu [1] thresholding

2.2 Morphological analysis of microscopic images

Morphological operations are implemented over binary image sets, obtained through thresholding. Basic operations vary from image preprocessing, skeletonizing, thinning, thickening of objects structures to quantitative (area, perimeter) description of objects. Quantitative analysis of objects (cells, fibers) is of great importance to researchers in muscle related injuries. An automated image processing methods extracts objective and quantitative results that can be used for comprehensive statistical analysis as compared to manual procedure [19]. The thresholded images can be processed with morphological operations (close, bridge, fill) and cell counting can be performed. A reseach work to count dead and live hepatocytes (livercells) in cultures use mathematical morphological operations to identify cells. Their accuracy is 95 % for single cell counting and 85 % for total cell counting as compared to manual count[17].

Basic morphological operations are [20]:

- Erosion

  This operation $A \ominus S$ is used to shrink an object in a binary or grey image. As in dilation, the shrink amount depends upon size and shape of the structuring element. Pixels are taken awaya from both inner and outer edges of objects.

$$A \ominus S = \ z \mid (S_z) \subseteq A$$

where,A - the original image, S - the structuring element, z - displacements (x,y) locations

- Dilation

  This operation $A \oplus S$ is used to increase the size of the connected component or an object in a binary image. The amount growth directly depends on size and shape of the structuring element. The growth is implemented by adding pixels near edges.

$$A \oplus S = z \mid (S_z) \cap A \neq \Phi$$

  where, A - the original image, S - the structuring element, z - displacements (x,y) locations

- Opening

  This operation $A \circ S$ uses both erosion and dilation which basically separates the objects. It removes small objects and repeated operation does not change the results.

$$A \circ S = (A \ominus S) \oplus S$$

- Closing

  This operation $A \bullet S$ is dilation followed by erosion that connects closer objects and fills small holes.

$$A \bullet S = (A \oplus S) \ominus S$$

The method described in [21] uses labview and Matlab, labview for nucleaus segmentation to calculate the statistical parameters such as mean and standard deviation. Algorithms include multi thresholding, morphological erosion and finally a sobel edge detector for segmentation. Their method evaluates geometrical features (area, perimeter) of segmented cells to classify cell as a blast or normal cell.

Varying sizes of filter and structuring element can be used to remove and reconstruct details in an image. A closing filter with disk shaped structuring element of radius six pixels was used after the median filtering (5*5). It helps the research to remove non-important details in the images [22].The morphological evaluation after intensity based segmentation helps to detect nucleus and cytoplasm in the blood image. This method show 92 % accuracy in the nucleus and 78 % for cytoplasm segmentation.

A robust and accurate segmentation method is developed by researchers [23] using morphological dome extraction method , followed by watershed segmentation. The original image is subtracted with constant value h to compute gray scale reconstruction. This reconstruction consists of h-domes and the marker image where all cells from the original image are uniquely represented as a single connected object.

The use of automatic image processing framework helps to analyze big data in short time. A research team has developed an approach to analyze a high number of biomedical images which can be further extended to accommodate clinical or medical record data. This framework can help in mining image data and machine-aided intelligent diagnosis, a crucial step in implementing electronic healthcare system [24]. We have used morphological operations like dilation and erosion to recover lost cell area and remove smaller non-cellular parts.

Chapter 3

Methodology

In this chapter our approach for automatic muscle regeneration process tracking is described in microscopic images.

A complete algorithm (see Fig. 3.1) includes image enhancement and dynamic thresholding for images of each time points. All the obtained images are further morphologically processed as per the requirement to calculate area, the distance between cells, cell counts etc. The morphological process includes removing non cellular area, fill, dilation, erosion. The final result is analyzed to see the change in cell shapes, size, location and relate with the performance of protein.

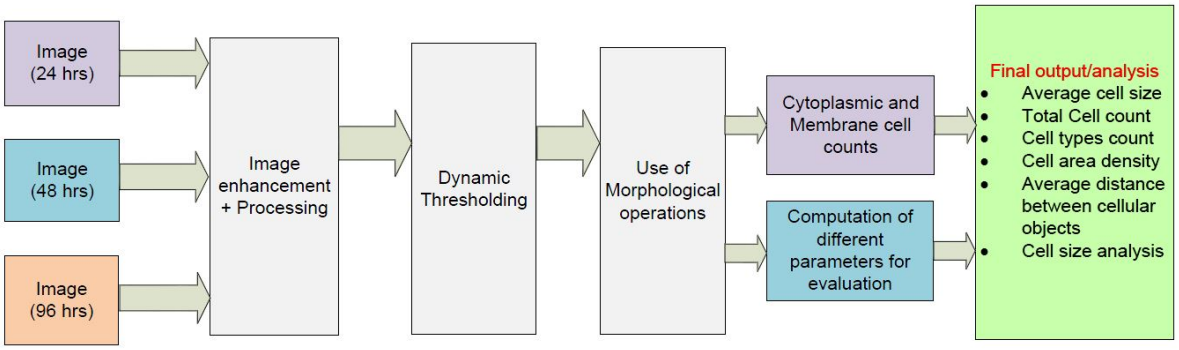


Figure 3.1: A block diagram for automated muscle regeneration tracking on microscopic images

3.1 Dataset

Dataset was collected from School of Nursing, University of Nevada, Las Vegas. The muscle injury is induced in mouse muscles and tissue section microscopy image is taken after different clinical procedures. These images taken at different time and using different antibodies are used for the further image analysis.

Mouse muscles are used for the study of muscle regeneration since the injury has to be made on the muscles. Before the start of the study, the mouse is kept for at least 5 days to acclimatize. Mice are kept individually in static polycarbonate micro isolation cages. The different mice are used to get muscle cross section image for different time. These experiments were performed at the University of Nevada, Las Vegas by the team of Dr. Barbara St. Pierre Schneider . The detail method and procedure can be found on the research paper [4]. Now to create injury, at first the crush injury device is calibrated between the mice. Then, mice are weighed and given buprenorphine (0.05 to 0.1 mg/kg SC; Buprenex, Reckitt Benckiser Pharmaceuticals, Richmond, VA). After 20 mins, the mouse is anesthetized using isoflurane via the nose cone. The crush injury is then induced on the medial or lateral surface of the gastrocnemius muscle.

After the generation of crush injury, they are given time for recovery in the animal housing facility for 24 or 48 h. Extra doses of buprenorphine are given every 12 h until the time of euthanasia. Besides this, they are regularly monitored by the veterinarian to check unrelieved pain, over grooming of the injured limb, and abnormal posture.

Next step is to take the microscopic images for leukocyte analysis using a microscope (fig). Mice are anesthetized with isoflurane, and the plantar flexor muscles are collected and kept frozen at -70 degrees. After collecting tissue (muscles), mice are euthanized under anesthesia by cervical dislocation, and the tibia and fibula are dissected to check the fractures. Then, the muscle cross sections are made using hematomas using a cryostat (model CM1850). These cross sections are applied to poly-L-lysine coated slides and kept at -70 degree for immunolabeling. They are also stained with hematoxylin and eosin to check damaged fiber and infiltration of cells.

Leukocytes are then immunolabeled with different five primary antibodies such as CD68, F4/80. Muscle sections are air dried and fixed in room temperature acetone. These section after completing various clinical procedure are used to take microscopic images.

3.2 Thresholding

There is no any universal method for thresholding method applicable multiple images. Dynamic thresholding (see Fig. 3.2) is used to set different thresholds for different images with the aid of statistical parameters such as mean and standard deviation.

All the images of tissue section are taken using the same microscope. During this process, chemicals are used for staining the sample and the whole process is manual. The manual error and the inconsistent use of chemicals can cause a change in intensity even for the same section of tissue. The dynamic threshold value can help to figure out cell area correctly.

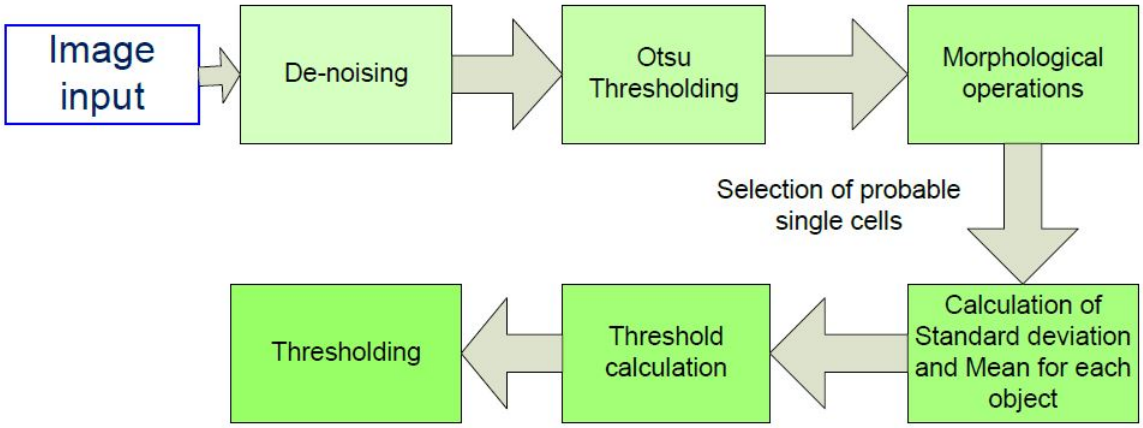


Figure 3.2: Block Diagram of Dynamic thresholding

Different thresholds are required for different images due to the change of intensity. The unique threshold value of each image is calculated using gray values followed by calculation of mean and standard deviation. Before calculating the threshold value, following assumptions are considered as advised by the personals working on this field:

- Area of single cells.
- Medium intensity cells are most probable ones.
- Constant intensity throughout the cell area.
- Possible circular shape.

A different dynamic threshold value (see Algorithm 3.1) is estimated for each image. During the calculation of dynamic threshold, before the calculation of statistical parameters (mean and standard deviation) Otsu's method is used for initial thresholding.

3.2.1 Otsu Thresholding

Otsu's method is based on the simple concept of finding the threshold that minimizes the weighted within-class variance [1]. This method works perfectly on bimodal histograms or an image with background and foreground.

The class probabilities are given by:

$$q_1(t) = \sum_{i=1}^{t} P(i)$$
$$q_2(t) = \sum_{i=t+1}^{I} P(i)$$

And the means for each class are estimated as:

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)}$$
$$\mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)}$$

The weighted within-class variance is:

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

In the last equation, use the every possible value of t [1,256], and pick the value that gives minimum estimation of $\sigma_w^2(t)$.

Algorithm 3.1: Dynamic Thresholding

Data: Image (i,j)

Result: *Dynamic thresholded image (Imageth)*

`/* Extent - ratio of pixels in the region to pixels in the total bounding box */`

convert to gray level image

Convert gray image to binary image(BW) using Otsu method

Remove connected objects with ***area<20*** *or* ***extent>0.5***

foreach *connected objects* **do**

    Find corresponding co-ordinates and gray values in gray level image

    Find mean (μ) and standard deviation (σ) of gray values

Remove the objects with standard deviation (σ>10)

Considering only the mean values of selected objects, find another mean (μ1) and standard deviation (σ1) of mean values

Estimate a threshold value as (T=μ1+σ1)

3.3 Computation of parameters for analysis

The area,shape, and texture of a cellular object gives information of regeneration stage. We will calculate these parameters for each image and see the change or pattern of change in the parameters, which eventually gives an idea of regeneration progress. At first, the smaller objects (area<20 pixel^2) are removed from all thresholded images before computing these parameters (value of area is in pixel^2).

3.3.1 Cell area density

Cell area density gives the amount of space taken by cell in certain area of muscle section. We have raster scanned whole tissue section image with varying box size (c), and calculated cell area density for each pixel. A density value for each pixel gives the cellular presence around that specific pixel. First, it helps to see the amount of high density pixels which denotes the presence of high number of cells in that area. Second, it helps to find high cell concentrated area in the muscle section. A varying parameter (c) is used in our method (see Algorithm 3.2), which changes the size of box to be considered around pixel. If a pixel (20,20)

and a box of size is 20 is chosen, it will consider a square of area 400 to calculate cell area density. The corner pixels will be as shown in Fig. 3.3. The change in box size changes the density value as per the cell cluster presence in that area. For higher clustered area, high density pixels are obtained for higher box size as well. But, in the case of uninjured muscle very pixels have higher density with increase in box size due to the less clustering of cells.

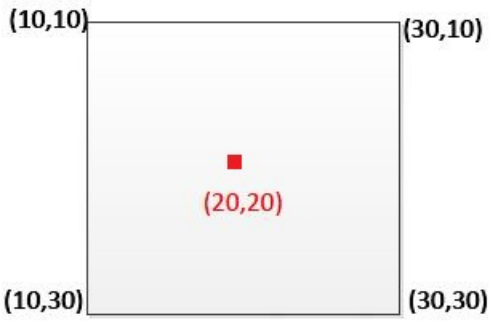


Figure 3.3: A sample pixel to show the cell area density calculation

Algorithm 3.2: Cell area density calculation

Data: Binary image (i,j), Box size (c)

Result: *Cell area density for each pixel*

`/* m=X co-ordinate of pixel, n=Y co-ordinate of pixel, c =` $\frac{1}{2}$`*size of box, area = cellular area, totarea = area of box considered */`

foreach *pixel in image* **do**

  `/* Special case for corner pixesl */`

  Calculate the total area to be considered (totarea= (min(x,m+c) - max (m-c,1)+1)(min(y,n+c)-max(n-c,1)+1))*

  if *if m<c+1——n<c+1——m>(x-1-c)——n>(y-1-c)* **then**

    for *a=max(m-c,1):1:min(x,m+c)* **do**

      for *b=max(n-c,1):1:min(y,n+c)* **do**

        if *(im(a,b)==0)* **then**

          area= area+1;

  else

    Calculate cellular area considering all the pixels inside box

    if *(im(a,b)==0)* **then**

      area= area+1;

  Calculate cell area density($D = \frac{area}{totarea}$*)*

3.3.2 Average size of objects

The area of all other total cellular objects (n) are considered to calculate average size of objects:

$$\text{Average Size} = \frac{\sum_{i=0}^{n} A_i}{n}$$

where, n = total number of objects

A_i = Area of i_{th} object which is calculated using a pixel as an unit.

The high average size of objects denotes multiple number of cells are close to each other or even attached. This shows some stage of regeneration process when cells are closer for repairing damaged fiber.

3.3.3 Cell counts

The numbers of cells are counted assuming each cell area would range from 20 to 500 pixel^2. Each cells in the range are counted as a single cell and for the objects with area >500 pixel^2. This range is determined using manual assumption, zoom level used while taking microscopic and images and the findings from previous research [4].

Cell count = Total area/ 500 pixel^2

Total number of cells = single cells + $\sum$ Cell count

The cells move towards damaged fiber to repair it, thus the high number of cells in certain area is sign of regeneration. The cell count is an important information to track the regeneration of muscle cell over time.

3.3.4 Cell size counts

Since the cells can be in different shape and size, cell size counts (see Algorithm 3.3) will count four different kinds of cells on the basis of area. These counts will done for all three timing images (ie 48 hr, 96 hr, 192 hr).

The range of area for each kind of cells:

A1 = Range (20-50)

A2 = Range (51-100)

A3 = Range (101-200)

A4 = Range (200+)

Algorithm 3.3: Cell size counts

Data: Binary Image (i,j)

Result: *Counts of different size cells*

```
/* A1 = Range (20-50), A2 = Range (51-100), A3 = Range (101-200), A4 = Range (200+),
   All area in pixel^2                                                              */
```

```
foreach cellular objects i do
    if 20<= area(i) & 50>=area(i) then
        A1=A1+1;
    else if (51<= area(i) & 100>=area(i)) then
        A2=A2+1;
    else if (101<= area(i) & 200>=area(i)) then
        A3=A3+1;
    else
        A4=A4+1;
```

3.3.5 Membrane and cytoplasmic cell counts

The cell types present in the muscles are cytoplasmic and membrane cells. The presence of cytoplasmic cells is a sign of regeneration in that area. Thus, finding the count of cytoplasmic cells is important to track the regeneration activity.

Cytoplasmic and membrane cells have different intensity value pattern from centroid to the extrema of object. Cytoplasmic cells have high gray level intensity at the center and it decreases on the corners, whereas membrane cells have low gray level intensity at the center and high at the ends. The cell inside red pentagon in fig 3.4 is probable cytoplasmic cell and the cells on green pentagon are probable membrane cells.

A method (see Algorithm 3.4) is developed to find the pattern of intensity change and identify cytoplasmic cells. For each object, the intensity pattern from centroid to 8 extremas of that object is considered to decide the kind of cell. Since the gray level value of centroid pixel in cytoplasmic cell is high, the object having centroid pixel gray value above threshold is considered cytoplasmic cells.

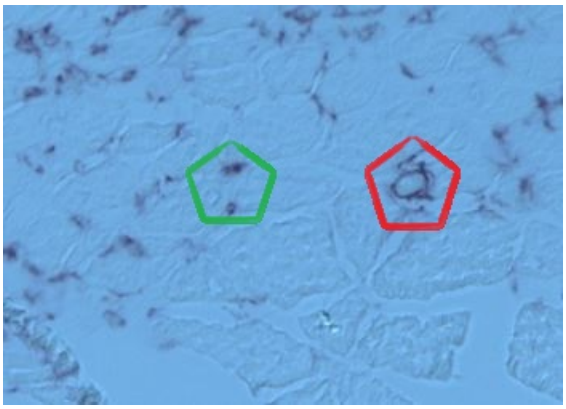


Figure 3.4: An image showing probable cytoplasmic and membrane cells in red and green pentagon respectively

Algorithm 3.4: Cytoplasmic and membrane cell count

Data: Binary Image (i,j)

Result: *Total number of cytoplasmic and membrance cells*

```
/* Ps(i) = Position of gray value in matrix, S(i) = Size of ith matrix set, ccyto =
   count of cytoplasmic cells, cmem= count of membrane cells                    */
```

```
1 foreach cellular objects do
2     Calculate centroid co-ordinate and 8 extrema co-ordinates
3     Find the gray level values from centroid towards extremas, 8 set of gray level values for each
      object;
4     foreach set of the 8 set of gray level values (i=1:8) do
5         find Ps(i) of gray value in matrix < threshold(T)
6         if Ps(i) > S(i)/2 then
7             ccyto = ccyto+1;
8         else
9             cmem = cmem+1;
```

3.3.6 Average distance between objects

The average distance between the objects is calculated using the centroid co-ordinates of two objects (see Algorithm 3.5). The distance is calculated with every other object in a selected area of muscle cross section.

This distance between the objects exhibit how closer are the cellular objects located. Also, the change in average distance over time gives the movement of cell. The decrease in distance shows, the cells moving closer which is a sign of regeneration in that section of cell. On the other hand, increase in distance is a sign of possible recovery or less regeneration.

Algorithm 3.5: Average Distance between objects

Data: Binary image (i,j)

Result: *Average distance between the objects*

Calculate the centroid co-ordinates of one object and every other objects

`/* n= total number of cellular objects, Totdist = total sum of distances calculated, tc= number of times distance calculated */`

foreach *objects n* **do**

  for *k=1:1:n-1* **do**

    Calculate x and y co-ordinates

    x1 = centroid(k,1); y1 = centroid(k,2);

    for *j=k+1:1:n* **do**

      x2 = centroid(j,1); y2 = centroid(j,2);

      Calculate distance (d) between objects

      $d = ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{\frac{1}{2}}$ *Calculate total sum of distances*

      Totdist = Totdist +d ;

  Find how many times distance was calculated

  for *n number of objects* (tc = $\frac{(n-1)*n}{2}$)

Calculate the average distance between objects Average distance = $\frac{Totdist}{totalcount}$

Chapter 4

Results and Discussion

This chapter is divided into two sections, the first section will show thresholding and intermediate results for CD68 antibodies. The second section contains six subsections which will show estimation of parameters for both CD68 and F480.

4.1 Thresholding

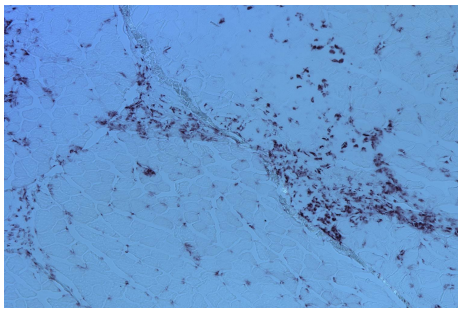
Thresholding includes multiple image processing method and statistical calculation. Thresholding results for CD 68 protein are shown in Fig. 4.1.

The gray level values of probable cells are used for calculation of mean and standard deviation (see sample values in Fig. 4.2). Though, few samples are shown in figure, 48 hr image used total of 416 objects and uninjured image used 58 objects to calculate final threshold(see Algorithm 3.1).

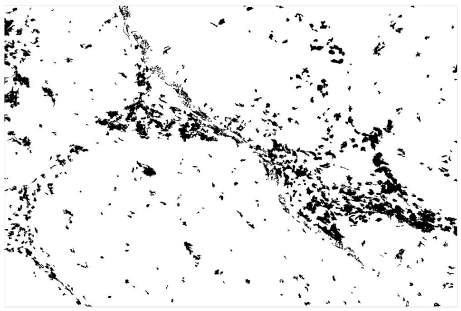
The final threshold values (see Table 4.1) are found close to the manually computed values.This threshold value worked considerably well with cd 68 images. But, for the images using other antibodies, the automatic threshold value included few non-cellular objects as well.

Sample Images	Automated Threshold	Manually computed
1	127¨	131
2	139¨	145
3	173¨	174
4	153¨	155
5	154¨	149

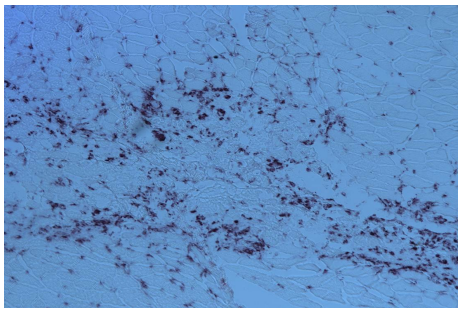
Table 4.1: Automated threshold value vs manually computed value



(a) An image of muscle cross section (48-hr)

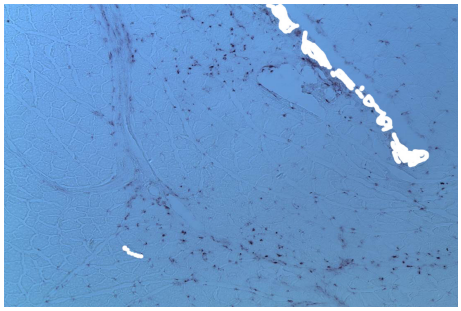


(b) Binary image after dynamic thresholding



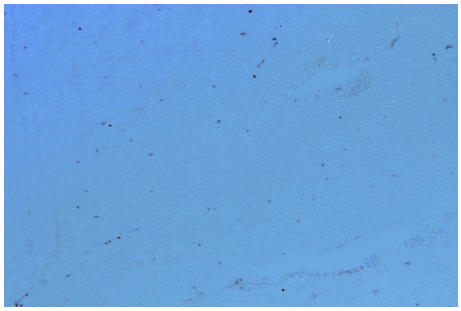
(c) An image of muscle cross section (96-hr)

(d) Binary image after dynamic thresholding



(e) An image of muscle cross section (192-hr)

(f) Binary image after dynamic thresholding



(g) An image of muscle cross section (uninjured)

(h) Binary image after dynamic thresholding

Figure 4.1: Results of dynamic thresholding in CD68 microscopic images

Object	Mean	Standard Deviation
1	76.13575	8.955933787
2	81.25325	7.61299937
3	81.65714	7.916463857
4	83.87652	5.592182043
5	86.00729	6.503608018
6	86.56253	8.820043357
7	87.69444	9.457109564
8	87.85253	7.224831694
9	88.30129	5.064582905
10	88.40948	9.177145526
11	89.65464	6.353094361

(a) A sample set of mean and standard deviation values(48-hr)

Object	Mean	Standard Deviation
1	56.08642	10.1482003
2	58.59868	10.28451228
3	63.26172	10.09270989
4	75.01303	9.172725271
5	79.7352	7.405893823
6	81.37409	9.954143709
7	82.565	8.628512299
8	85.60637	9.203925227
9	85.98825	7.095032404
10	86.03978	9.061008341
11	91.6748	6.82263784

(b) A sample set of mean and standard deviation values for uninjured muscle

Figure 4.2: Results of dynamic thresholding in CD68 microscopic images

4.2 Computation of Parameters

The algorithm was implemented in two antibodies CD68 and F480 to track the stages of regeneration. The results for each parameter are presented in this section.

4.2.1 Cell area density

The whole image is raster scanned with varying box size. Each pixel will take consideration of area around it as per the size of the box. Then, the cellular density is calculated for the corresponding box. Each pixel will have density value which denotes the presence of cells or cluster of cells.

The information about cell clustering can be extracted using the relation between density value and box size. If the density value stays high even after the increase in the size of the box, it is a sign of the presence of big clusters of cells. When the pixels above 60% are considered , for CD68, 48-hr, and 96-hr time point has high-density value while increasing box size from 10 to 30 (see Table 4.2, 4.3, 4.4). The average density and pixel count largely decreases for the uninjured case since there is no cluster and regeneration. For F4/80, the cell clustering is high at the 96-hr time point, a sign of active regeneration (see Table 4.5).

Time	Average Density	Pixel count
uninjured	127¨	237
48	67.73¨	86146
96	79.72¨	37647
192	73.43¨	1844

Table 4.2: Cell area density considering pixels area density above 60% and box size 10 (CD68)

Time	Average Density	Pixel count
uninjured	60¨	1
48	74.47¨	51802
96	69.68¨	14062
192	64.73¨	589

Table 4.3: Cell area density considering pixels area density above 60% and box size 20(CD68)

Time	Average Density	Pixel count
uninjured	NA¨	NA
48	68.78¨	23019
96	64.97¨	1697
192	NA ¨	NA

Table 4.4: Cell area density considering pixels area density above 60% and box size 30(CD68)

Time	Average Density	Pixel count
32	65.23¨	5223
96	73.42¨	13712
192	64.73¨	2121

Table 4.5: Cell area density considering pixels area density above 60% and box size 20 (F480)

For all the figures in 4.3, the x-axis ranging from 0 to 100 is a density value and a y-axis is a number of pixel for corresponding density value. The gaps in the plot are increasing over the time and high in the case uninjured muscle. This shows the absence of few kind of cell clusters in the image. These graphs were generated for CD68 protein with the box size of 20.

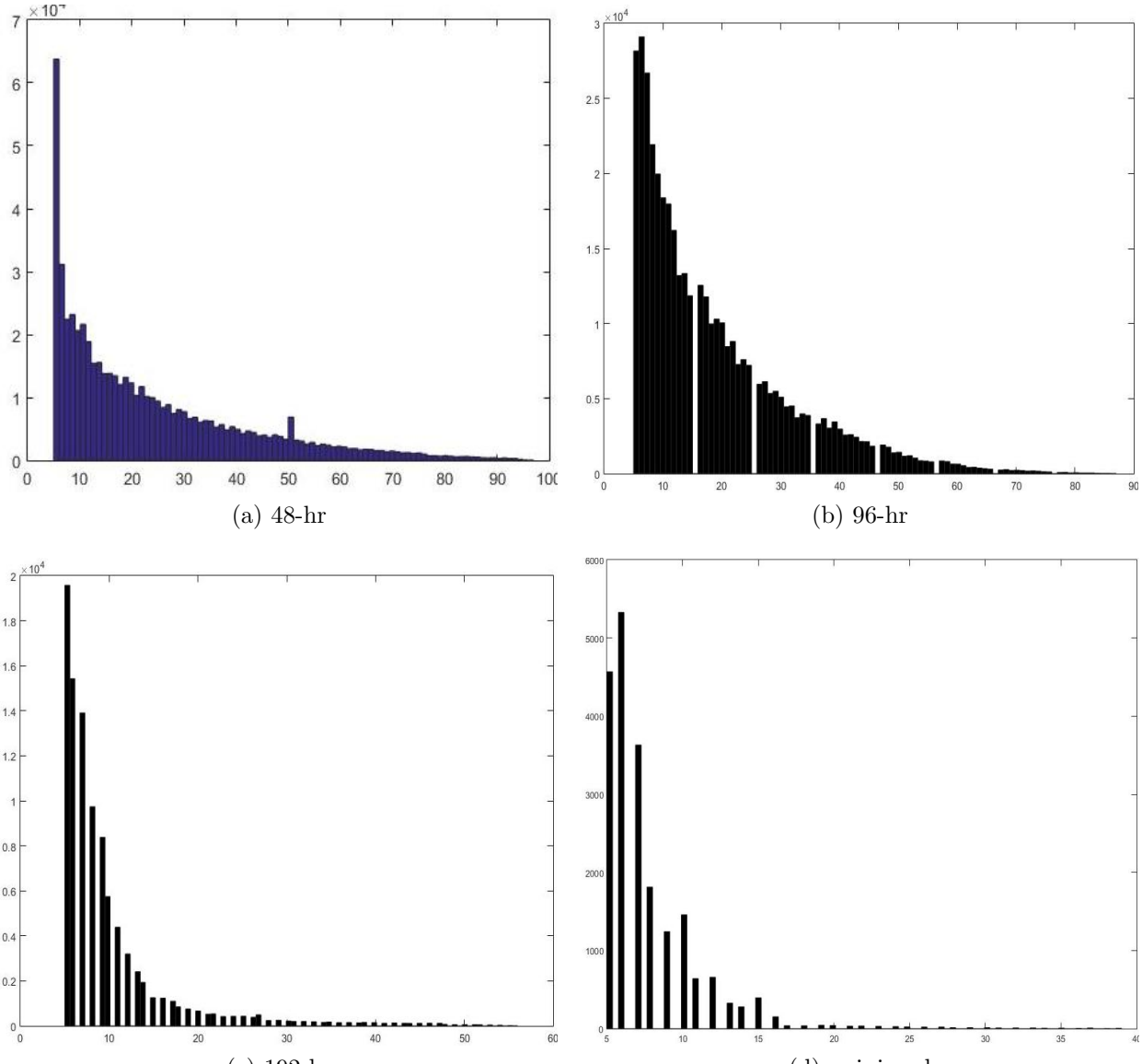
4.2.2 Average size of objects

In the case of CD68, the average size is decreasing uniformly over time. The size was highest at 48 hours, which is probably the peak of regeneration. The time point 300 here is the average size of uninjured tissue.

Whereas for F480, the peak is at 96 hours. This protein might have started to repair damaged fibers at 96 hours, which is very late as compared to CD 68 (see Fig. 4.9). The absence of some range of density values in an image is noticed with the gaps in the plot. These plots are for CD68 images with the box size of 20.

4.2.3 Cell counts

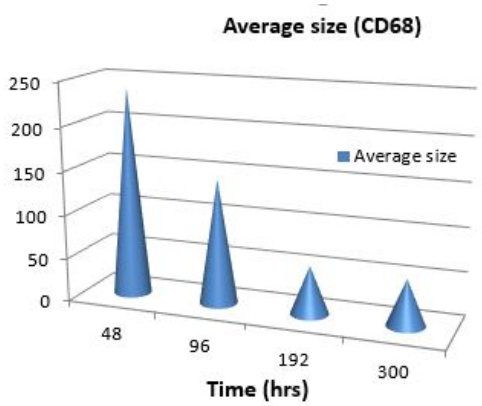
The high number of cells in an area is a sign of active regeneration. In the case of CD68, the number cells are decreasing over the time. The time point 300 is the uninjured/normal case, the graph shows cell



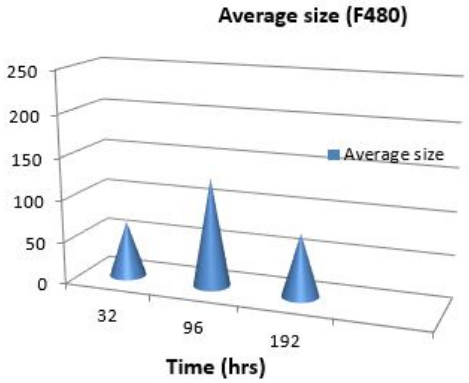
(a) 48-hr (b) 96-hr

(c) 192-hr (d) uninjured

Figure 4.3: A histogram plot of density values for each pixels in different time points



(a) CD68



(b) F480

Figure 4.4: A histogram showing change of average size of objects with time CD68(a) and F480(b)

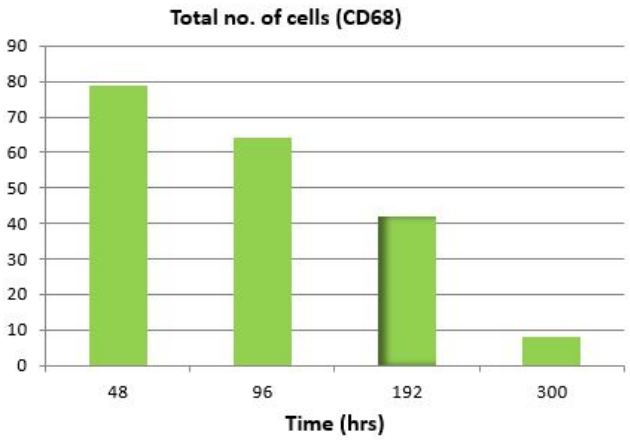
count getting closer to normal. For F480, cell count increases highly at the 96-hr time revealing the start of regeneration after 96 hours (see Fig. 4.5).

4.2.4 Cell size counts

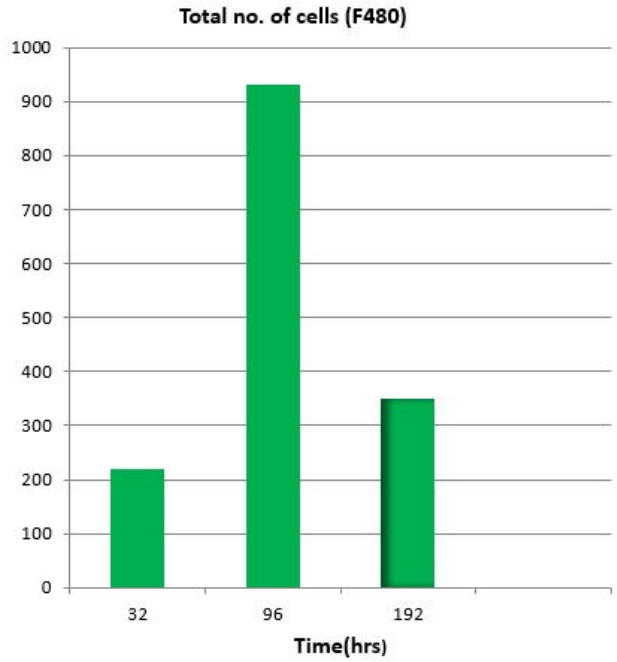
In the figures (see Fig. 4.6, 4.7), the x-axis shows increasing range of cell size such as 1 =Range (20-50), 2 = Range (51-100) 3 = Range (101-200), 4 = Range (200+) . The colors denote different time points.

For CD68, 96 and 48-hrs time point have the similar ratio of varying range of cellular objects. At, 192-hrs, there are mostly smaller cells and this pattern is close to the case of uninjured one (see Fig. 4.6).

For F480, bigger cells are present only at 96-hrs time point and also the number of cells is high as compared to that of CD68 (see Fig. 4.7).



(a) CD68



(b) F480

Figure 4.5: Cell count change over time in CD68(above) and F480(down)

4.2.5 Membrane and cytoplasmic cell counts

The high number of cytoplasmic cells in a certain area is a sign of active regeneration.

The decrease in cytoplasmic cell counts may indicate resolving regeneration and the return to normal muscle. For CD68, the cytoplasmic cell counts is decreasing over the time. But for F4/80, there are high counts in 96-hr time point and then it decreases. The highF4/80 counts occur at 96-hr may indicate an early time point of regeneration (see Fig. 4.8).

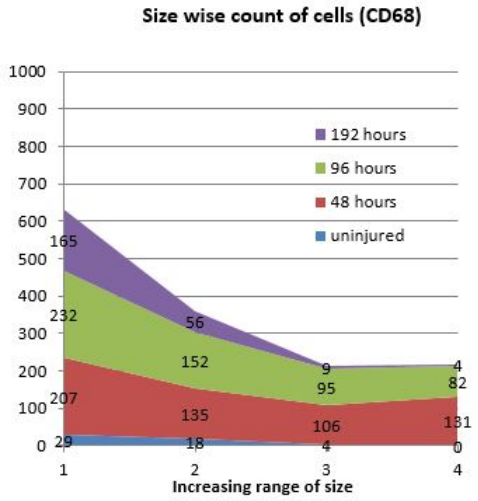


Figure 4.6: An area graph showing the amount of cells in different size and time

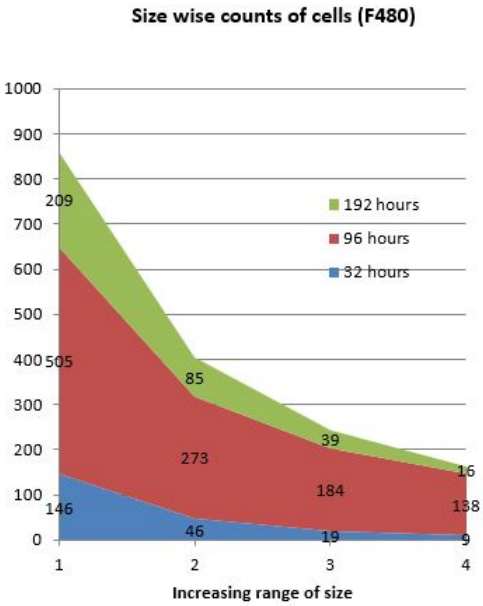
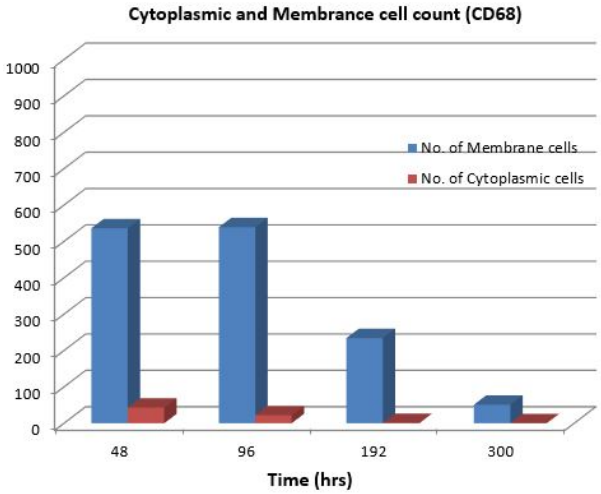


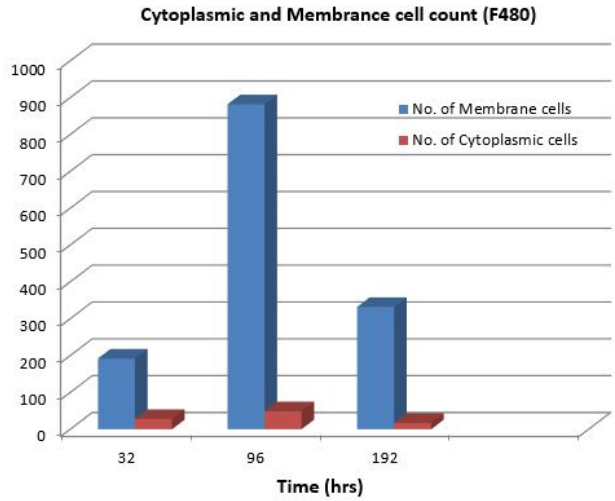
Figure 4.7: An area graph showing the amount of cells in different size and time (F480)

4.2.6 Average distance between objects

The average distance between the cellular objects is increasing with time in CD 68. This may be leukocytes completed their regeneration task and are not seen as time increases. Whereas, cells are closest at around 96 hours time point in F480. Since, centroid values are used to calculate average distance, large size cellular objects may have increased the distance in 96-hrs of F480 (see Fig. 4.9).

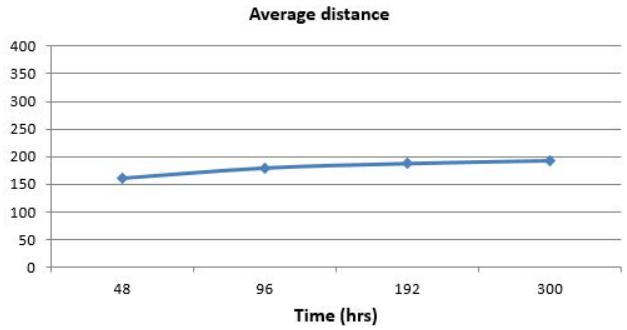


(a) CD68

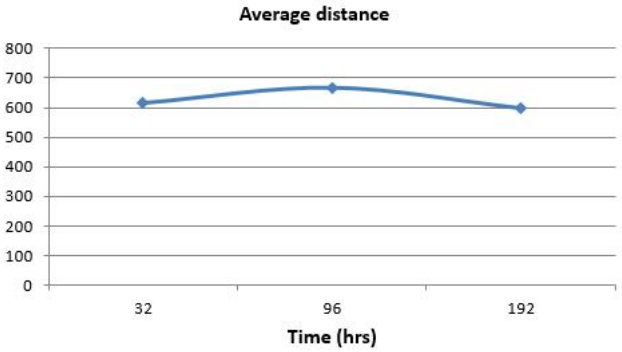


(b) F480

Figure 4.8: Total number of cytoplasmic and membrane cells in different time CD68(a) and F480(b)



(a) CD68



(b) F480

Figure 4.9: A curve showing change of average distance between cells with time CD68(a) and F480(b)

Chapter 5

Conclusion and Future Works

In this research, we were able to develop an automatic dynamic thresholding method though it doesnt work perfectly for all set of antibodies due to varying staining during clinical procedure to make muscle cross section slides. Our set of threshold values were close to the manually computed ones. Since this method is useful for unimodal microscopic images , it can be time saving and help to analyse a large amount of medical data. We implemented our thresholding method to get cellular objects in muscle section images. Our parameters calculated using thresholded image exhibited a pattern of change in shape,size, density of cells over time. This information is helpful to track regeneration in muscle. Also, there is a uniform increase of the average distance between cellular objects with time in CD68. It can be concluded that the number of cells is decreasing with time and they are moving away from each other from 48 hr in case of CD 68. Whereas, cells go closer and get clustered at 96 hours in case of F480 and only then pattern of CD68 is followed.

For future work, we can implement below mentioned two ideas to get mathmaetical modeling of protein performance.

- Intensity based thresholding alone cannot cover all the cell area or sometimes it may lose as well. Thus, we can collect more data and use classifiers to identify cells and fiber as well.
- We can increase the number of timings in which images are taken. This helps to model change of shape, size, density etc over time. A good number of data will help to extract a mathematical equation which can be very useful to compare protein performance and discover better therapies.

Bibliography

[1] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[2] S. Keeling, N. Deashinta, S. V. K. M. Howard, S. M. S, and B. S. Schneider, "Macrophage colony stimulating factor-induced macrophage differentiation influences myotube elongation." in *Biological Research for Nursing*, vol. 15, no. 1, 2013, pp. 62–70.

[3] C. Hughes, C. T. Hasselman, T. M. Best, S. Martinez, and W. E. G. Jr., "Incomplete, intrasubstance strain injuries of the rectus femoris muscle." vol. 23. Am. J. Sports Med, 1995, pp. 500–506.

[4] G. L. Dobek, N. Fulkerson, J. Nicholas, and B. S. P. Schneider, "Mouse model of muscle crush injury of the legs." vol. 63, no. 3. Comparative Medicine, 2013, p. 227232.

[5] K. H. Ghazali, R. S. Hadi, and M. Zeehaida, "Microscopy image processing analaysis for automatic detection of human intestinal parasites alo and tto," in *Electronics, Computer and Computation (ICECCO), 2013 International Conference on*, Nov 2013, pp. 40–43.

[6] C. H. Comin, X. Xu, Y. Wang, L. da Fontoura Costa, and Z. Yang, "An image processing approach to analyze morphological features of microscopic images of muscle fibers," vol. 38, no. 8, 2014, pp. 803 – 814.

[7] E. A. Mohammed, M. M. A. Mohamed, C. Naugler, and B. H. Far, "Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding," in *Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on*, May 2013, pp. 1–5.

[8] M. P. de Albuquerque, I. A. Esquef, A. da Rocha Gesualdi, and M. P. de Albuquerque, "Image thresholding using tsallis entropy." *Pattern Recognition Letters*, vol. 25, pp. 1059–1065, 2004.

[9] J. N. Kapur, P. K. Sahoo, , and A. K. C. Wong, "A new method for graylevel picture thresholding using the entropy of the histogram." in *Computer vision, graphics, and image processing*, vol. 29 Issue. 3, 1985.

[10] H. Madhloom, S. Kareem, H. Ariffin, A. Zaidan, H. Alanazi, and B. Zaidan, "An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold," *Journal of Applied Sciences*, vol. 10, pp. 959–966, 2010.

[11] A. A. Nawandhar, L. Yamujala, and N. Kumar, "Image segmentation using thresholding for cell nuclei detection of colon tissue," in *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, Aug 2015, pp. 1199–1203.

[12] F. Sadeghian, Z. Seman, R. A. Ramli, H. B. Abdul Kahar, and M.-I. Saripan, "A framework for white blood cell segmentation in microscopic blood images using digital image processing," *Biological Procedures Online*, vol. 11, pp. 196–206, 2009.

[13] X. Liang, H. Xu, and Y. Liu, "Recognition of living abnormal cells based on an optical microscope," in *Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO), 2012 International Conference on*, Aug 2012, pp. 18–21.

[14] D.-K. T. Le, A. A. Bui, Z. Yu, and F. M. Bui, "An automated framework for counting lymphocytes from microscopic images," in *Computing and Communication (IEMCON), 2015 International Conference and Workshop on*, Oct 2015, pp. 1–6.

[15] Z. Yin, R. Bise, M. Chen, and T. Kanade, "Cell segmentation in microscopy imagery using a bag of local bayesian classifiers." IEEE, 2010, pp. 125–128.

[16] S. Mohapatra, D. Patra, and K. Kumar, "Blood microscopic image segmentation using rough sets," in *Image Information Processing (ICIIP), 2011 International Conference on*, Nov 2011, pp. 1–6.

[17] H. Refai, L. Li, T. K. Teague, and R. Naukam, "Automatic count of hepatocytes in microscopic images," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2, Sept 2003, pp. II–1101–4 vol.3.

[18] E. Espinoza, G. Martinez, J. G. Frerichs, and T. Scheper, "Cell cluster segmentation based on global and local thresholding for in-situ microscopy," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, April 2006, pp. 542–545.

[19] S. Schiaffino and T. Patridge, "Skeletal muscle repair and regeneration. advances in muscle research," vol. 3. Springer, 2008.

[20] R. C. Gonzalez and R. E. Woods, "Digital image processing (3rd edition)." Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006, pp. 523–531.

[21] S. Mohapatra and D. Patra, "Automated leukemia detection using hausdorff dimension in blood microscopic images," in *Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on*, Dec 2010, pp. 64–68.

[22] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical and Biological Engineering and Computing*, vol. 44, pp. 427–436, 2006.

[23] L. M. Vincent and B. R. Masters, "Morphological image processing and network analysis of cornea endothelial cell images," in *San Diego'92*. International Society for Optics and Photonics, 1992, pp. 212–226.

[24] T. Goudas, C. Doukas, A. Chatziioannou, and I. Maglogiannis, "A collaborative biomedical image-mining framework: Application on the image analysis of microscopic kidney biopsies," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, pp. 82–91, Jan 2013.

Curriculum Vitae

Graduate College
University of Nevada, Las Vegas

Bibek Jang Karki

Degrees:

Bachelor of Electronics and Communication Engineering 2012
Tribhuvan University, Nepal

Thesis Title: A Study on Automated Process for Extracting White Blood Cellular Data from Microscopic Digital Injured Skeletal Muscle Images

Thesis Examination Committee:

Chairperson, Dr. Mei Yang, Ph.D.
Committee Member, Dr. Yingtao Jiang, Ph.D.
Committee Member, Dr. Venkatesan Muthukumar, Ph.D.
Graduate Faculty Representative, Dr. Barbara St. Pierre Schneider, Ph.D.