“SEGMENTATION OF ANTI NEUTROPHIL CYTOPLASMIC ANTIBODIES (ANCA) IMAGES BASED ON WATERSHED AND WAVELET”

BY

MUHAMAD HASYEME B NASIR

FINAL PROJECT REPORT

Submitted to the Department of Electrical & Electronic Engineering
in Partial Fulfillment of the Requirements
for the Degree
Bachelor of Engineering (Hons)
(Electrical & Electronic Engineering)

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

“SEGMENTATION OF ANTI NEUTROPHIL CYTOPLASMIC ANTIBODIES (ANCA) IMAGES BASED ON WATERSHED AND WAVELET”

by

Muhamad Hasyeme B Nasir

13186

A project dissertation submitted to the

Department of Electrical & Electronic Engineering

Universiti Teknologi PETRONAS

in partial fulfillment of the requirement for the

Bachelor of Engineering (Hons)

(Electrical & Electronic Engineering)

Approved by:

____

AP Dr. Ibrahima Faye

Project Supervisor

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

SEPTEMBER 2013

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

____

Muhamad Hasyeme B Nasir

Abstract

Autoimmune disease is a type of disease where immune system unable to tell between the good side and bad side which lead to the misguided attack on the healthy cells and tissues. Autoimmune disease can be classified to more than 80 types depending on the affected area. The test also varies according to the suspected type of disease. Some examples of the test are Enzyme-Linked Immunosorbent Assay (ELISA) test, Indirect Immunofluorescence (IIF) test of Antinuclear Antibody (ANA) by using HeP-2 Cells and IIF test for Anti Neutrophil Cytoplasmic Antibodies (ANCA). However in this project, author only focus on the ANCA images with two major staining patterns which are P-ANCA and C-ANCA. Currently the positivity of the images depends solely on the experience of the physician which led to variety of result and lack of reliability. Besides the time to get the result is time consuming. Thus an automatic classification system has been developed to overcome the manual process. The vital process inside the automatic system is the segmentation part. Many researchers suggest different techniques of segmentation to segment the ANCA images before being further processed. In this research, author focus on Watershed technique to segment the ANCA images by implementing the algorithm in Matlab. Author use Wavelet transform to suppress noise to avoid from over segmentation of the ANCA images. Using Rand Index method, the result of segmentations is verified. Combination of Watershed and Wavelet transform gives a very promising result. Recommendation for future work is to explore on automatic determination of noise variance inside images.

ACKNOWLEDGEMENT

Author strongly believe that the completion of this project almost impossible without the permission from Allah almighty. With His blessing and guidance, author manages to finish the research within the schedule framework.

In addition there are several parties and individuals who are directly involved in the completion of this project. Therefore author would like to take this opportunity to express deepest gratitude to them. First of all, the author would like to thank the supervisor for this project, AP Dr Ibrahima Faye. He is a very supportive person and has contributed a lot of ideas and suggestions to make sure this project is a success.

Special thanks go to co-supervisor, Ms Zazilah May for the technical support in obtaining the medical images. Not forgetting the help from Dr Mohamed Meselhy Elthoukhy for helping the author during the final stage of project.

The author would also like to thank Dr. Nazreen Bt Badruddin , the FYP coordinator, for organizing seminars, talk and preparation of guidelines for FYP 1 and 2. Last but not least, author would like to thanks moral support to friend and families especially Shafiq Nor, Amin Fahim, Liyana Mehfar and Thania Awatif.

Finally, the author would like to thank all parties involved once again for making this Final Year Project a success.

TABLE OF CONTENT

LIST OF TABLE

LIST of graph

LIST OF EQUATION

LIST OF FIGURE

LIST OF ABBREVIATION

HUSM	Hospital Universiti Sains Malaysia
SNR	Signal to Noise Ratio
ELISA	Enzyme-Linked Immunosorbent Assay
CTD	Connective Tissue Disease
ROIs	Region of Interests
ANCA	Anti- Neutrophil Cytoplasmic Antibodies
P-ANCA	Perinuclear Anti- Neutrophil Cytoplasmic Antibodies
C-ANCA	Cytoplasmic Anti- Neutrophil Cytoplasmic Antibodies
DWT	Discrete Waveletet Transform
RI	Rand Index
IIF	Indirect immunofluorescence
RGB	Red, Green, Blue channel
FCM	Fuzzy Clustering Method
SOM	Self Organizing Map
ANA	Antinuclear- Antibodies
GA	Genetic Algorithm

Chapter 1: Introduction

1.1 Background

Human being has 3 lines of defense mechanism which protects the bodies from foreign entities. The first layer is called as physical barriers which protect the body from pathogen. These barriers consist of skin and mucous membrane. When there is a breach of security in the first defense, our second layer of defense will take in action. Second layer of defense consist of nonspecific defensive mainly known as white blood cell (leucocyte). White blood cell will attack and destroy the pathogen that try to harm the body. If foreign agent manages to penetrate the nonspecific barriers, we have the last option of defense mechanism which is the specific immune response. Specific immune response will be activated when foreign protein entities which are recognized by the body diffuse into the body thorough the blood circulatory system or deposited in the tissue. In medical term we called this as antigens. Antigens will trigger the production of specific protein which known as antibodies. Antibodies will combine with the antigen to inactivate the foreign proteins.

The core ability of immune system is to differentiate between foreign invaders and self-structure. Thus it will only fight against the antigen and protect body from harm. However sometime there is mistake which makes the immune system unable to tell between the good side and bad side which lead to the misguided attack on the healthy cells and tissues. The attack is referred as autoimmune disease. Autoimmune disease can be classified to more than 80 types depending on the affected area. In America, 23.5 million of women are affected with this disease. Some examples of the disease are Connective Tissue Disease (CTD), diabetes mellitus (type I), multiple sclerosis and rheumatoid arthritis.[1]

One of the most reliable approaches to diagnose the presence of autoimmune disease is using indirect immunofluorescence (IIF) test to the Antineutrophil cytoplasmic antibodies (ANCA). The IIF test is applied to ANCA to detect 2 staining pattern which referred to cytoplasmic (C-ANCA) and perinuclear (P-ANCA) [2].IIF test on ANCA is mostly related to the several specific autoimmune diseases which are Wegener Wranulomatosis, Microscopic Polyangiitis and Churg-Strauss syndrome. All of these are the same classes with small-vessel vasculitis[3].

. Manually interpreting the pattern has several limitations. The accuracy of test result is hardly dependent on the experience physician and may vary subjectively. Other than that manual process require higher time expenditure. There is also poor standardization which will lead to incompatibility of result around the globe. An outcome from manual testing drawback is development of other alternative such as Enzyme-Linked Immunosorbent Assay (ELISA) test. Compared to the iFFT, ELISA is easier to used but it lack of reliability. A study has concluded automated diagnosis of digital image is more reliable compared to the manual process [4]

1.2 Problem Statement

Automatic segmentation of ANCA images play important role in the preliminary processing stage of Computer-Aided Diagnostic System for Auto-immune disease. A lot of research paper suggested that it is crucial to have proper segmentation of the fluorescence pattern before the images being classified into the according group. Segmentation will provide a clearer edge parameter for the cells which increase the rate accuracy at the classification stage. However despite the importance of the cell segmentation task in ANCA images, not many papers suggest proper segmentation techniques. In addition such paper used private data for testing which makes it difficult to objectively compare the reported results. According to the literature, improvement needs to be done to segmentation technique to provide accurate segmentation on fluorescence pattern to differentiate between C-ANCA and P-ANCA.

1.3 Objectives of Study

1.3.1 To develop a segmentation algorithm for segmenting the 2 major patterns of ANCA images in the Indirect Immunofluorescence (IIF) test using Watershed and Wavelets.

1.3.2 To assess the performance of this method using common and publicly available dataset of ANCA images.

1.4 Scope of study

1.4.1 The study will focus on to two staining pattern of the ANCA images which are C-ANCA and P-ANCA

1.4.3 The main technique for the segmentation of the ANCA images is Watershed approach with the preliminary image enhancement using Wavelet transform.

1.4.4 The image output is binary images with the input images in RGB.

1.4.5 MATLAB software version 2012b will be used for the algorithm implementation.

1.4 Relevancy of the project the

The main objective of this project is to improve the segmentation process of cells immunofluorescence pattern. Proper segmentation play vital role in term of image understanding and pattern recognition as segmentation is the preprocessing step in digital image application. Thus the outcome of this project will help to increase the accuracy of segmentation analysis which will create a very significant impact in medical image application. Proper segmentation will produce high accuracy of classification of the ANCA images to the respected group

1.5 Feasibility of Project

The scope of study is within the reach. The author will focus only one technique of segmentation which is the watershed technique. According to the literature review, this project can be completed within the time frame because if we avoid from using or involves any long data time collection. Furthermore the images for data collection purpose are ready to be analyzed. Time taken to analyze the images is not very long as author will get help from physician and expert on the field. Therefore this project can be run accordingly.

Chapter 2: Literature Review

2.1 Segmentation of Immune Fluorescence (IIF) Pattern

Nowadays the development of automated detection of Immune Fluorescence (IIF) Pattern has been increase vigorously due the many advantages compared to the manual method. The main component of the automated detection system is the image segmentation. Segmentation is defined as dividing the image into several regions of interests (ROIs) depends on the specific parameters that has been predetermined before. The parameters can be referred as color, intensity, or texture [5]. Image segmentation also can be defined as decoding image of several particular pixels of interest which mostly will enhance the shape and edges of particular object [2].

2.2 Reliability of image segmentation

Image segmentation is used widely in medical image processing due to its constant reliability of result. As has been discussed earlier, image segmentation requires specific parameter to achieve the desired output which requires different segmentations techniques depend on the type of medical image. Every medical image possessed a different level of color intensity; region based contour, image dimensionality. As such, the main goal is to determine the features to identify object recognition [3]. Moreover the segmentation process helps in term of noise reduction and reduced the blurry image while preserving the natural condition of image shape and texture. [6]

2.3 Otsu Tresholding

The region of interest (ROIs) to be examined is the cell nucleus or a bunch of nuclei. One main nature of cell cytoplasmic structure is not visible independently throughout the segmentation. So the dispersion of background during the inhomogeneous luminance will result in low quality of segmentation. One method being proposed to overcome this obstacle is by applying an adaptive Otsu Tresholding. Image is divided to four pixels block averaging 250 x 250 with typical diameter of 46 pixels and the Otsu algorithm is applied independently to each block resulting a new bounding box shaped ROI around it. This process will continue iteratively until the number of segmented iterations dropped to 1% [7, 8]. However Tresholding method has severe limitation as it fail to recognize two antinuclear- antibodies (ANAs) pattern namely nucleolar and spekled. [9]. Beside that this method also consume significant amount of computing time which result in inefficiency of time management [10].

2.4 Fuzzy Clustering Method (FCM)

Other than Tresholding, clustering approach is also one of the most used method to differentiate object of interest between foreground and background. Clustering approach is an unsupervised method which can be divided into 2 main techniques. The first technique is known as crisp (hard) and a second technique is known as fuzzy clustering. Crisp provide efficient result because it allow image to be converted into binary version. It will separate data into a predetermined cluster but in real medical image often belong to more than one cluster. Therefore this limitation is overcome with Fuzzy-c means (FCM). It offer flexibility and able to tolerate with outlier[10]. FCM is straight forward and easy to implement. Despite the advantages, clustering techniques have limitation under the natural environment. It is very sensitive to color variation due to the existence of illumination. In addition it also time consuming like the tresholding method.

2.5 Neural Network Approach

Another segmentation method that provides promising result is segmentation with neural network approach. Neural network approach mostly referred as Artificial Neural Networks (ANNs). The Hopfield neural networks bring significant contribution in the image reconstruction. It converge the segmented digital image into stable state while minimizing the energy function. However the feed forward neural network and Kohonen self-organizing map (SOM) network provide better segmentation by providing a linear approximation directly from finite element simulations of problems [11]. Hence the disadvantage of this technique is it requires complex computation. For optimization of result Genetic Algorithm (GA) being applied to SOM output. In medical application, SOM being applied to the magnetic resonance image (MRI) of brain for early detection of tumor [5].

2.6 Watershed and Wavelet Approach

The last technique being reviewed is the watershed technique. The watershed transform has remarkable properties that make it useful for much different image segmentation [2][12]. Some of the fundamental properties are it is easy to handle with simple algorithm calculation, can be merged with other segmentation method and produces a very compromising result on segmentation. Despite the disadvantages such as over segmentation, very sensitive to noise and lower detection of noise structure, pre-processing using morphology method, Gaussian filter and edge base before and after the segmentation can overcome the drawbacks [13, 14]. One paper the usage of Wavelet transforms as the enhancement techniques before the images being segmented. Purpose of the enhancement is to eliminate noise and enhance the images edges[15]. The combination of these two techniques produces a very promising result. Hence this method can bring major advantages in the segmentation of medical images. Experimental result demonstrates combination of these method produce promising result [16-18]

2.7 Analysis on literature review

Based on the literature review above, there abundant of segmentation techniques to be applied on the medical images before further processing such as features extraction, classification and positivity test. Segmentation play vital role in the medical images analysis as it bring huge impact on the outcome of the test.

Different segmentation techniques have their own advantages and drawback. Otsu Tresholding is easy to handle and provide good result but a lot research has been done with this technique. FCM offer flexibility and straight forward but very sensitive to color variation. In addition Neural Network Approach require a lot of training data which author don’t have in the provided dataset

After reviewed all the techniques, Watershed approach is the most suitable method to segment the ANCA images because it compose of simple algorithm, promising result and the drawback can be overcome by combining with other segmentation approach. Author proposed to combine Watershed with Wavelet Transform as the image enhancement [19, 20]

Chapter 3: Methodology and Project Work

3.1 Methodology

3.1.1 Research

The very first step in this project is to have a brief insight on the topic. By reviewing the research on the topic, it will help to get basic idea on the techniques and method of segmentation. Segmentation is a very fundamental approach in digital image processing. There are a lot of approaches to segment an image with different their own advantages and point to improve. Besides, segmentation also being applied in many other applications such as satellite imaging and military based activity. Thus it is crucial to review the most suitable techniques for segmentation of medical image

3.1.2 Critical analysis on segmentation approach

In the literature review, there are abundant of outstanding segmentation approach with promising result such as Edge Based Method, Otsu Tresholding, Fuzzy Clustering Method, Watershed approach and Discrete Wavelet Transform (DWT). Critical analyses on the literature review suggest to use DWT in the image pre-processing before the image undergoes segmentation. Watershed will over segmented region inside the cell due to high sensitivity to random noise and low image quality. On the theoretical part, DWT will remove the noise and provide better image for watershed segmentation

3.1.3 Preparing the segmentation algorithm

After the segmentation method has been finalized, the project will continue with the development of the algorithm for segmentation. This project will apply the uses of Matlab with help of image processing and Wavelet and Watershed toolbox which is available freely in the web.

Segmentation algorithm is the most crucial phase because it affects the output of segmentation. Poor algorithm will result in unreliability of automated image segmentation.

3.1.4 Evaluation of segmentation algorithm

After the segmentation method has been finalized, the project will continue with the evaluation of the algorithm. The research is empirical. Hence results are obtained from the implementation of programming to make justification on the objectives. This project will use Matlab image processing toolbox (Watershed and Wavelet) which is available freely in the web. The segmentation algorithm will be evaluated using ground truth data[10].

3.1.5 Optimization of algorithm

After we have evaluated the performances of proposed algorithm, changes will be done according to the specific requirement so that the algorithm will produce the best segmentation result. In this phase it is necessary to review thoroughly the body of coding so that it compatible with the desired output

3.2 Key milestone

3.2.1 Extensive research

The main purpose of literature review is to study the early development of segmentation techniques and it reliability. Besides it also aimed to find point of improvement in the segmentation techniques. Overall, literature analysis will increase understanding on the project by supplying general information on staining pattern of ANCA and segmentation approach. Duration of reviewing literature is about 3 to 4 weeks.

3.2.2 Segmentation method and algorithm computing

Based on the literature analysis, the best method for segmentation based on the neural network will be chosen. Algorithm base on the specific method will be developed.

3.2.3 Implementation of algorithm

Evaluation of algorithm will be done in Matlab. The raw medical images of P-ANCA and C-ANCA will be taken from publicly available dataset. This is to make sure the algorithm can be applied accordingly in the Matlab

3.2.3.1 Compiling coding to M-file

Once the coding has been completely developed, it will then be compiled to M-file to prepare it for use.

3.2.3.2 Testing DWT coding

To verify and debug the coding in segmentation phase

3.2.3.3 Testing pre-processing coding

To verify and debug coding in pre-processing phase

3.2.3.4 Testing Watershed coding

To verify and debug coding in Watershed phase

3.2.3.5 Testing Rand Index coding

To verify and debug coding for Rand Index.

3.2.4 Improving algorithm being developed

There will be always room to increase the accuracy of segmentation.

3.2.5 Documentation and reporting

All the research material and methodology will be properly documented in different type of reports as reference for the future

3.3 Gantt chart

Refer to the Appendix

3.4 Flow Chart

Figure 1 below show the general work flow of overall process in conducting the research

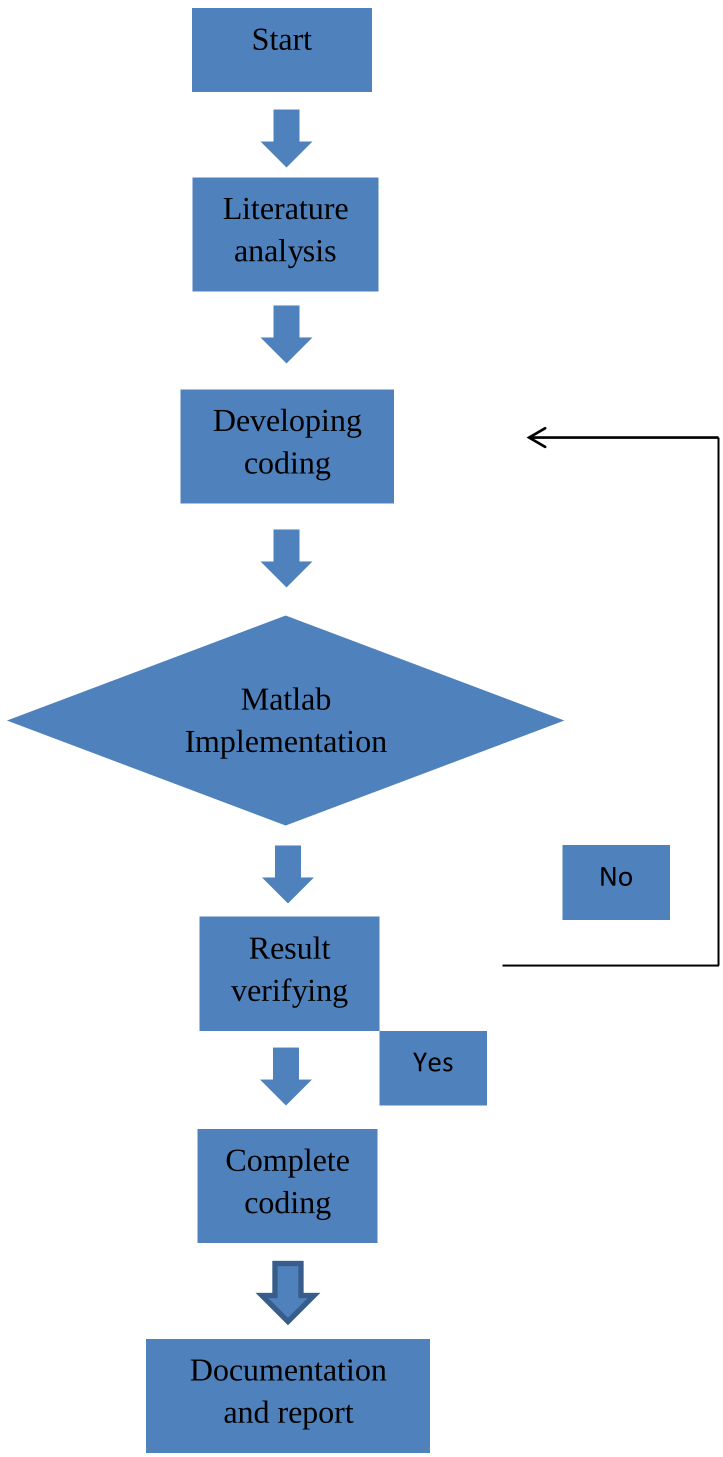


Figure 1: General research work flow

Figure 2 below show the workflow of algorithm develop for this project. The algorithm can be divided into 4 main parts which are DWT for image de-noising, image pre-processing (gray scale, median filter, erosion, etc.), Watershed segmentation and evaluation of segmentation using Rand Index.

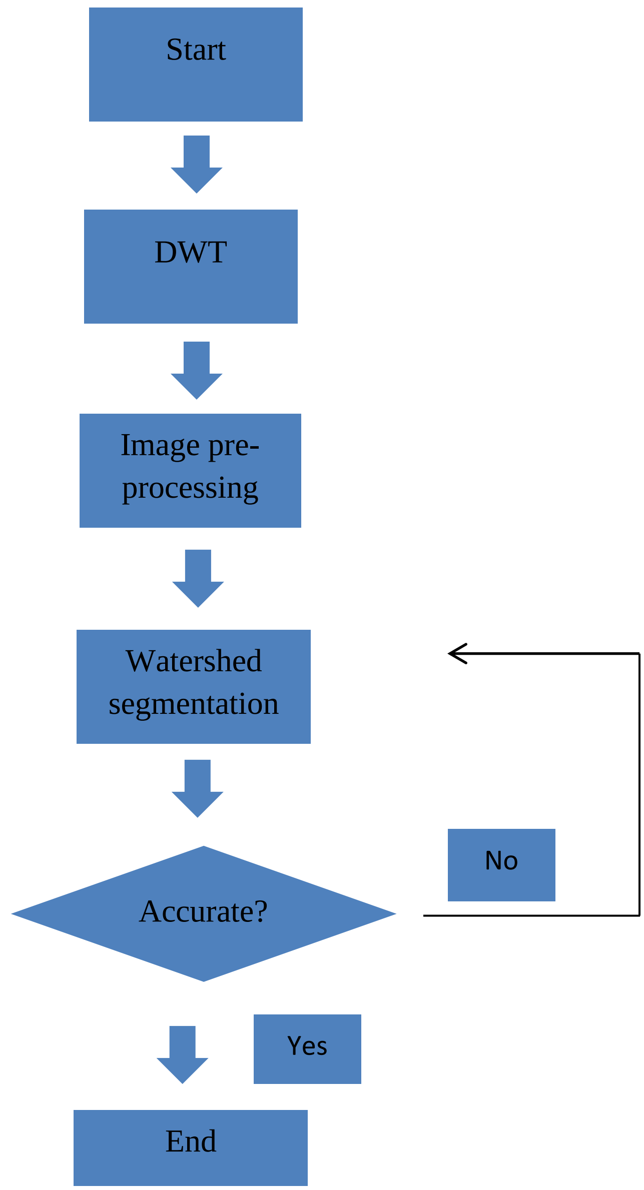


Figure 2: Overall algorithm work flow chart

Figure 3 below show detailed pre-processing method. Since Watershed approach will result in over segmentation, pre-processing the de-noised image play significant role to increase accuracy of segmentation.

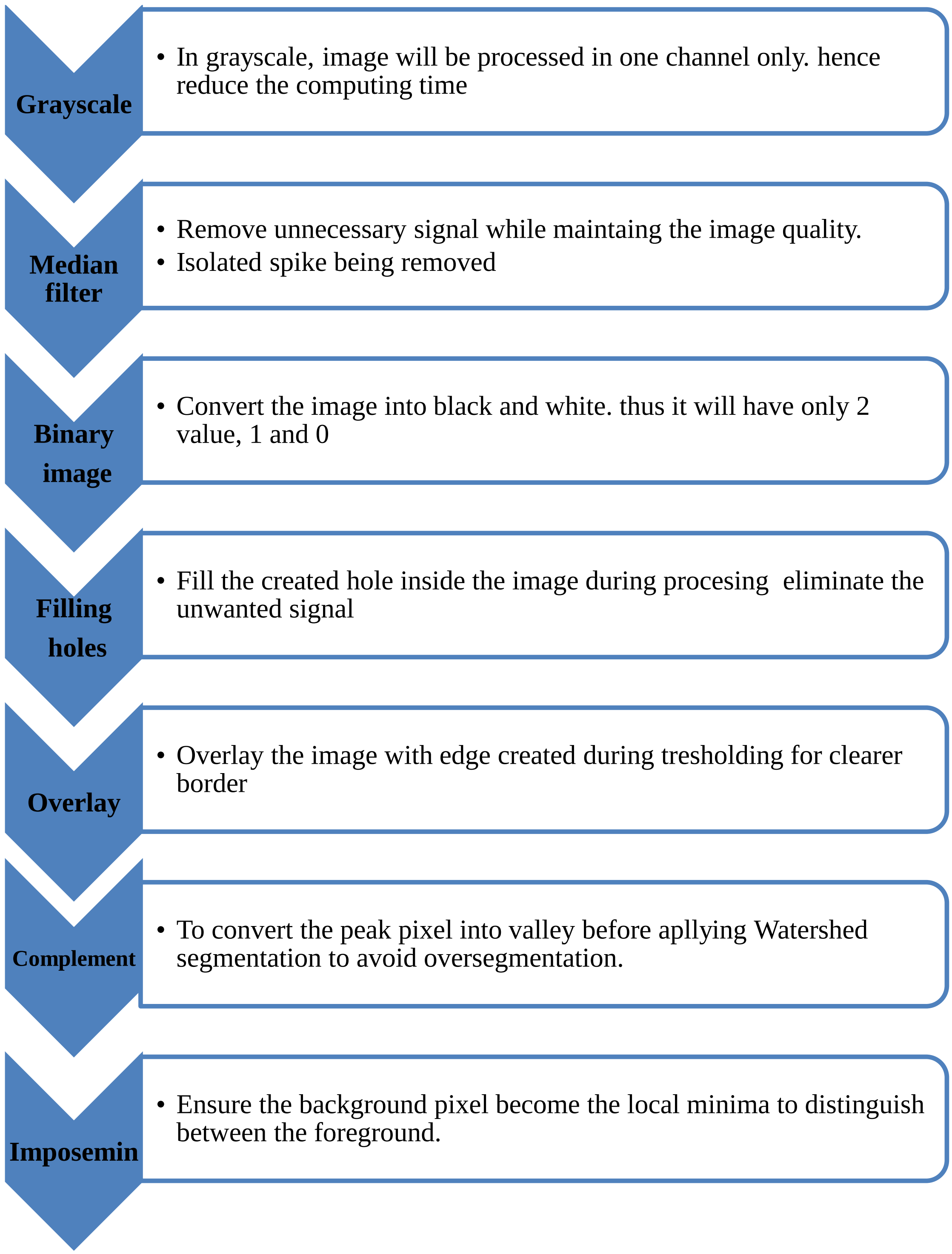


Figure 3: Overall pre-processing work flow chart

3.5 Tools

This project involves several imaging appliances such as microscope for the data collection and visualization but the image will be retrieved from lab at Hospital Universiti Sains Malaysia (HUSM) Kubang Kerian by expert. For software, Matlab will be used to segment the image using specific algorithm. Below are the summaries:-

i. Hardware
   a. Computer

ii. Software
   a. MATLAB R2012

3.6 Project Schedule

The planned scheduled for this research are displayed in table 1 and table 2 below:

Table 1: Project schedule for Final Year Project I

Title selection	23 May 2013
Extended proposal submission	5 June 2013
Proposal defense	15 July 2013
Draft interim report	8 August 2013
Final Interim report	25 August 2013

Table 2: Project schedule for Final Year Project II

Progress Report	12 November 2013
ELEXTREX	4 December 2013
Draft Report	16 December 2013
Final Report	23 December 2013
Technical Paper	30 December 2013
VIVA	30 December 2013
Dissertation (Hard bound Report)	1 January 2014

Chapter 4: Result and Discussion

A total of 60 images of 2 major staining pattern of ANCA; P-ANCA and C-ANCA have been segmented using the developed algorithm. The image passes through several pre-processing to enhance the image quality. Pre-processing include DWT for noise removal. Signal to Noise Ratio (SNR) is calculated to verify the noise removal. The de-noise image is then undergoes Watershed segmentation. The performance of algorithm is being measured by comparing the segmented image with ground truth data. Comparison is done using the Rand Index value.

4.1 ANCA image pattern

Type of autoimmune disease can be approximately identified by identifying the staining pattern of ANCA. There are only 2 major staining pattern of ANCA which cytoplasmic ANCA or C-ANCA and perinuclear ANCA or P-ANCA. ANCA is normally associated with Wegener's Granulomatosis, Microscopic Polyangiitis and Churg–Strauss.

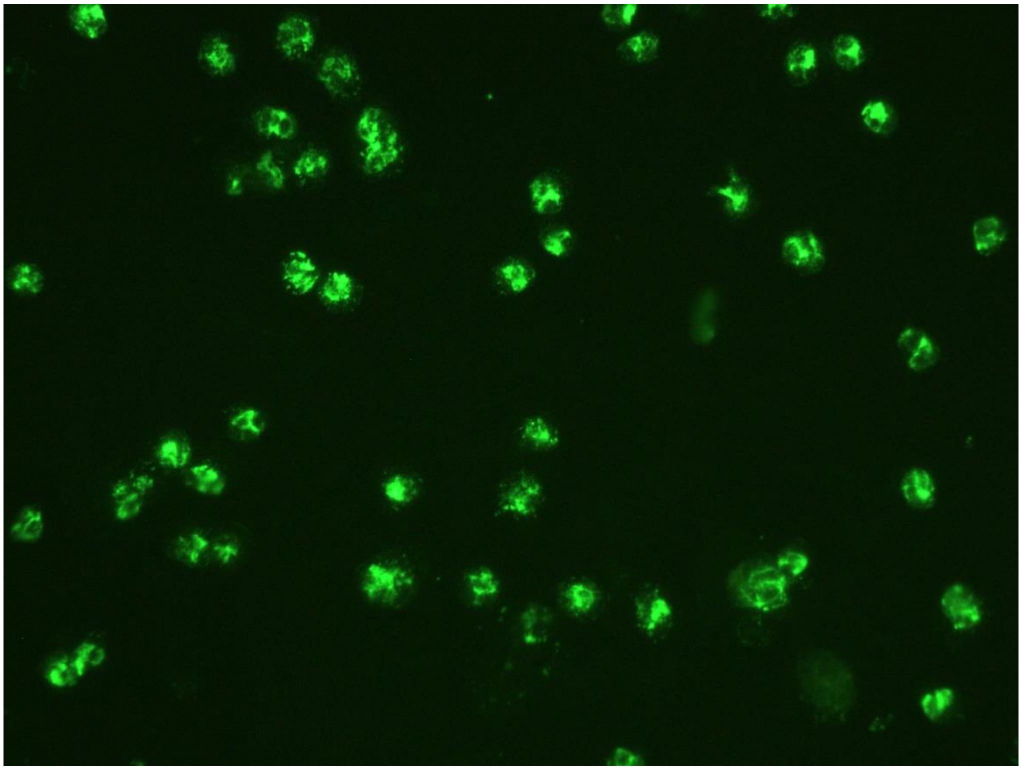


Figure 4: Sample of positive C-ANCA

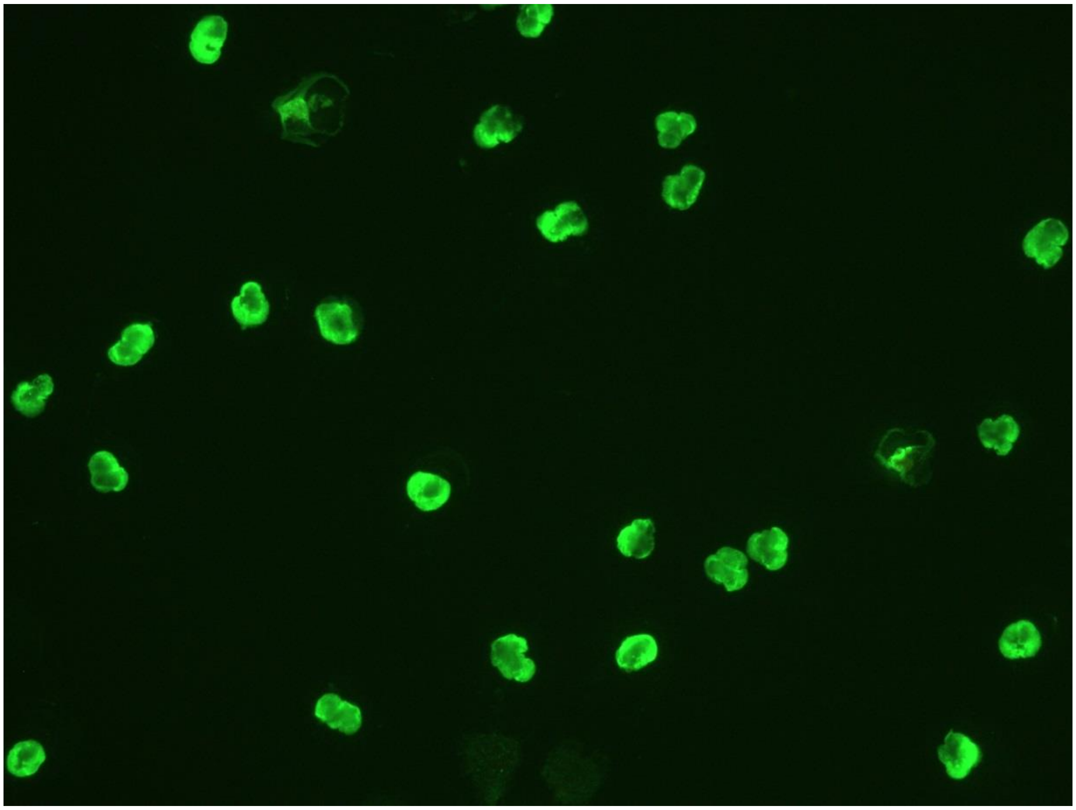


Figure 5: Sample of positive P-ANCA

Figure 4 and Figure 5 above show clear images of C-ANCA and P-ANCA. The C-ANCA pattern is characteristically confined to the cytoplasm with somewhat central accentuation meanwhile the P-ANCA pattern of staining is confined around nucleus. For P-ANCA nucleus and cytoplasm remain devoid from any staining.

4.2 DWT image de-noising

Watershed approach is very sensitive to noise. Presence of noise will lead to over segmentation. Thus image de-noising form very important initial step to suppress noise within the image. Medical images from HUSM contain and noise and lack of quality. Thus by applying signal wavelet decomposition using DWT, noise manage to be reduced. This can be observed from the increment of SNR. SNR is calculated from the formula below:-

*SNR = 10*log10 (Signal Value/Noise Value)…………………………………….. (1)*

30 ANCA images undergo DWT for noise removal. SNR for each image is calculated and the result is tabulated in table below:

Table 3: SNR of C-ANCA images after DWT

No	Image Label	SNRO	SNRT
1	C-ANCA1	7.1684	18.6048
2	C-ANCA2	8.4980	19.7915
3	C-ANCA3	6.6319	18.0129
4	C-ANCA4	11.3606	19.0266
5	C-ANCA5	10.5488	17.6149
6	C-ANCA6	12.5720	22.0421
7	C-ANCA7	12.5801	22.0609
8	C-ANCA8	11.0020	16.8583
9	C-ANCA9	11.3369	17.0576
10	C-ANCA10	11.1207	17.0014
11	C-ANCA11	14.5828	21.5482
12	C-ANCA12	15.7147	22.7569
13	C-ANCA13	17.1118	24.207
14	C-ANCA14	16.5164	23.5182
15	C-ANCA15	18.7238	25.7411
16	C-ANCA16	14.5547	21.5233
17	C-ANCA17	15.6618	22.6295
18	C-ANCA18	15.7063	22.6499
19	C-ANCA19	17.3306	24.2332
20	C-ANCA20	17.4495	24.3357
21	C-ANCA21	10.8316	20.4028
22	C-ANCA22	8.8609	19.1193
23	C-ANCA23	8.9819	18.0646
24	C-ANCA24	11.1112	18.0014
25	C-ANCA25	9.8002	18.7397
26	C-ANCA26	11.2068	20.4554
27	C-ANCA27	11.3721	19.3198
28	C-ANCA28	9.8063	15.6731
29	C-ANCA29	9.8783	19.5724
30	C-ANCA30	10.2575	20.3273

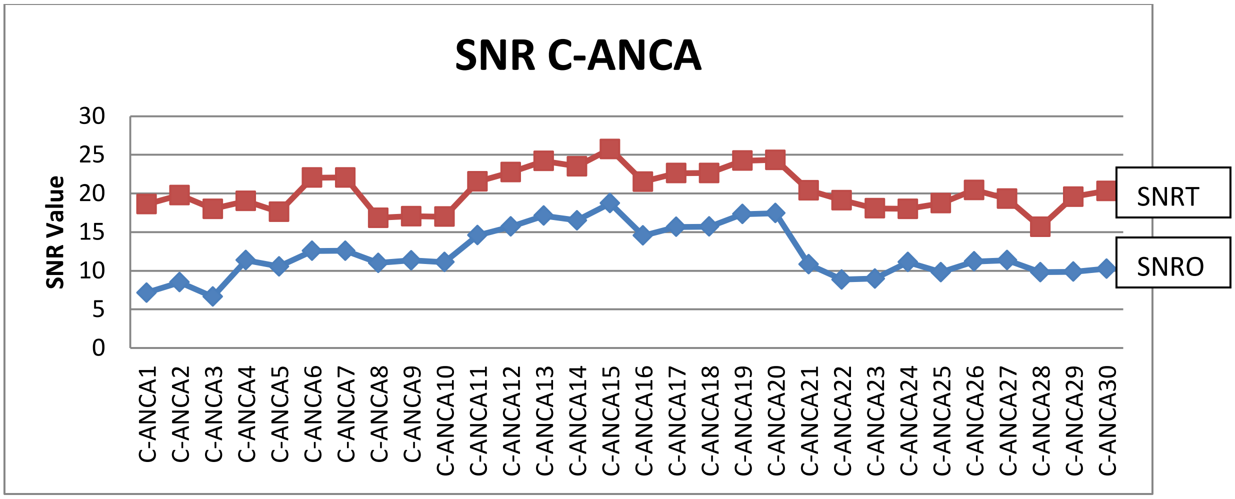
*SNRO – SNR for original image

*SNRT – SNR for DWT image

Table 4: SNR of P-ANCA images after DWT

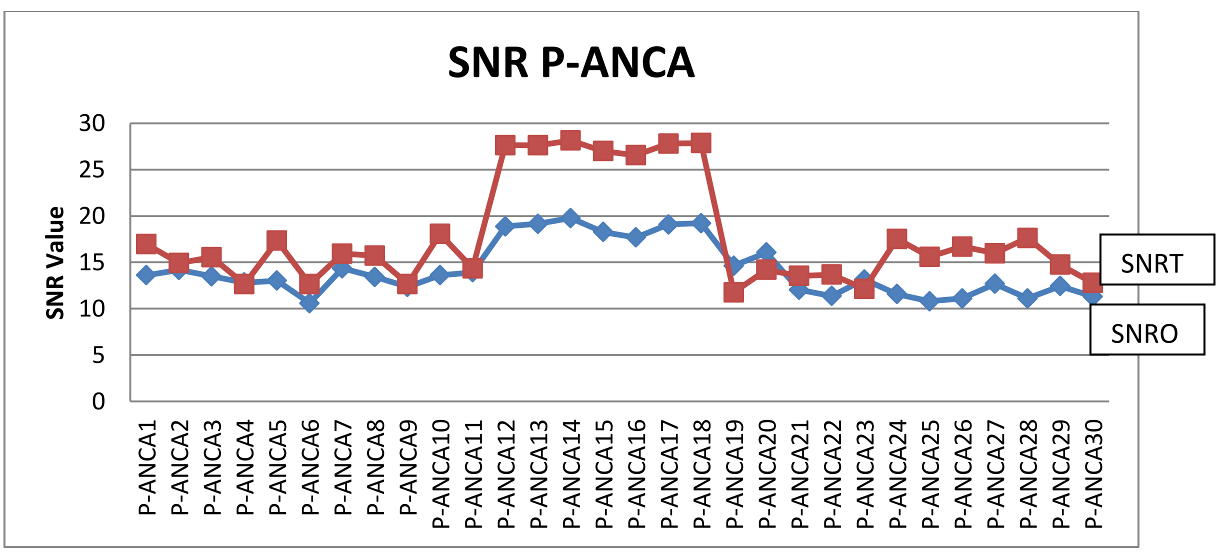
No	Image Label	SNRO	SNRT
1	P-ANCA1	13.6111	16.9549
2	P-ANCA2	14.1741	14.9188
3	P-ANCA3	13.4918	15.5641
4	P-ANCA4	12.8155	12.6432
5	P-ANCA5	13.0296	17.3397
6	P-ANCA6	10.5921	12.649
7	P-ANCA7	14.3479	15.9324
8	P-ANCA8	13.4135	15.7096
9	P-ANCA9	12.3457	12.6654
10	P-ANCA10	13.6138	18.0591
11	P-ANCA11	13.9291	14.3587
12	P-ANCA12	18.8628	27.6333
13	P-ANCA13	19.1394	27.6136
14	P-ANCA14	19.7694	28.1353
15	P-ANCA15	18.2605	27.0021
16	P-ANCA16	17.6845	26.5442
17	P-ANCA17	19.0754	27.8031
18	P-ANCA18	19.2091	27.8574
19	P-ANCA19	14.6241	11.762
20	P-ANCA20	16.0547	14.1977
21	P-ANCA21	12.0469	13.5434
22	P-ANCA22	11.3754	13.6741
23	P-ANCA23	13.1419	12.1473
24	P-ANCA24	11.5762	17.5305
25	P-ANCA25	10.8016	15.5879
26	P-ANCA26	11.1011	16.6958
27	P-ANCA27	12.6849	15.9676
28	P-ANCA28	11.1025	17.6245
29	P-ANCA29	12.4437	14.7426
30	P-ANCA30	11.3223	12.7849

Table 1 and table 2 above compare SNR value before and after DWT on the medical images. Total of 30 P-ANCA images and 30 C-ANCA images be studied and the result tabulated in the table. From both table, SNR value after DWT show increment. This translated to increasing of signal strength with suppression of noise within the image. The result is further interpreted by graph below:



Graph 1: SNR for C-ANCA Pattern

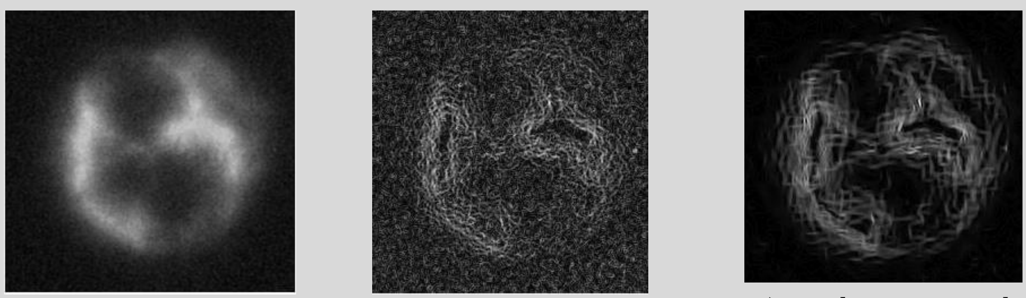
Graph above show the comparison of original SNR of image before and after the DWT for C-ANCA. Blue line represents the SNR of original image meanwhile the red line is SNR of DWT. From the graph, 100% of image show increment in SNR. This translates into better image quality with less noise. Average SNR for original image is 12.2759 per image and for DWT are 20.3600 per image. Further analyses on the SNR give us each image have averagely increase their SNR of 8. 0841.



Graph 2: SNR for P-ANCA Pattern

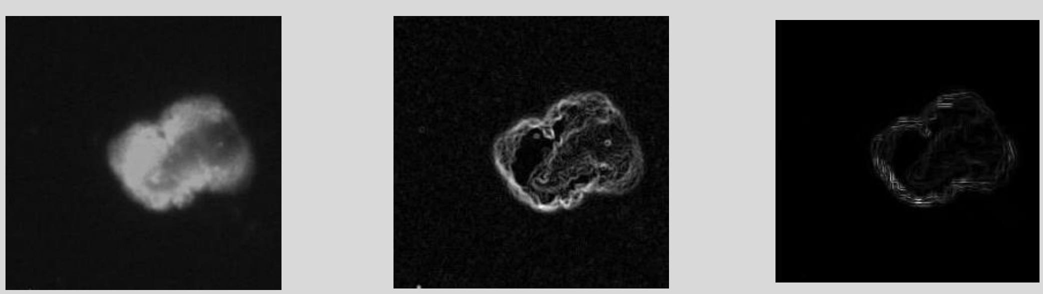
Graph 2 above show the comparison of original SNR of image before and after the DWT for P-ANCA. Blue line represents the SNR of original image meanwhile the red line is SNR of DWT. From the graph, approximately 90% of image show increment in SNR. 10% of image has lower SNR after DWT. This is because during the wavelet transform, some coefficients of image has been misinterpreted as noised and being remove. Average SNR for original image is 14.1900 and for DWT are 17.85. Percentage of increment is 20.5%.

To conclude the SNR result analysis, DWT work effectively to de-noise image. Although there are image that show lower SNR value after DWT, the percentage is small and can be neglected.



A) Original gray image *B) Gradient magnitude of original image* *C) Gradient magnitude of transformed image*

Figure 6: comparison of C-ANCA original image with transform image



A) Original gray image *B) Gradient magnitude of original image* *C) Gradient magnitude of DWT image*

Figure 7: comparison of P-ANCA original image with transform image

Figure 1 and Figure 2 above show the original gray image of 2 staining ANCA pattern, gradient magnitude of original image and gradient magnitude of DWT image. For qualitative evaluation, gradient magnitude is used to analyses the result of DWT on noisy image. To reduce computation time, image is processed in gray scale. Gray scale run on one channel while preserving the nature and information within the image.

Gradient magnitude of original image show there are unwanted signal in the image which considered as noise. After applying DWT, the noise is successfully removed. By observing the gradient magnitude of DWT image, the unwanted signal has been removed. However, the florescent intensity on DWT image reduces.

4.3 Watershed segmentation

Watershed segmentation take place after the image has been de-noising. Watershed literally refers to catchment basin. In image processing, a gray scale image is regard as topographic relief. When the topographic relief being flooded, water will raise up until the water coming from different minima meet up and form a dam. Due to different topographic structure, many dams are form and line between the dams is regard as segmentation[16]. Sometimes when there is too many regional minima, over segmentation happen. This is a well-known phenomenon in Watershed.

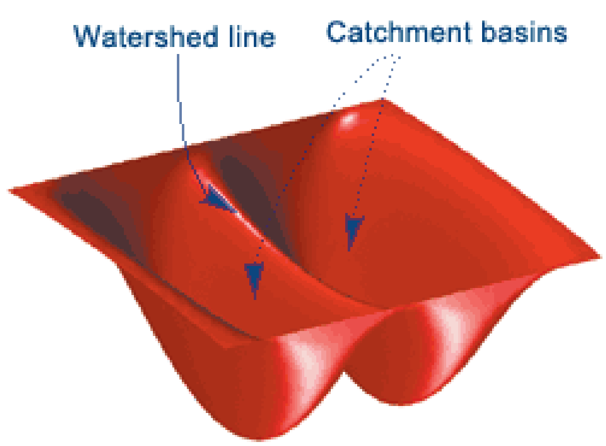
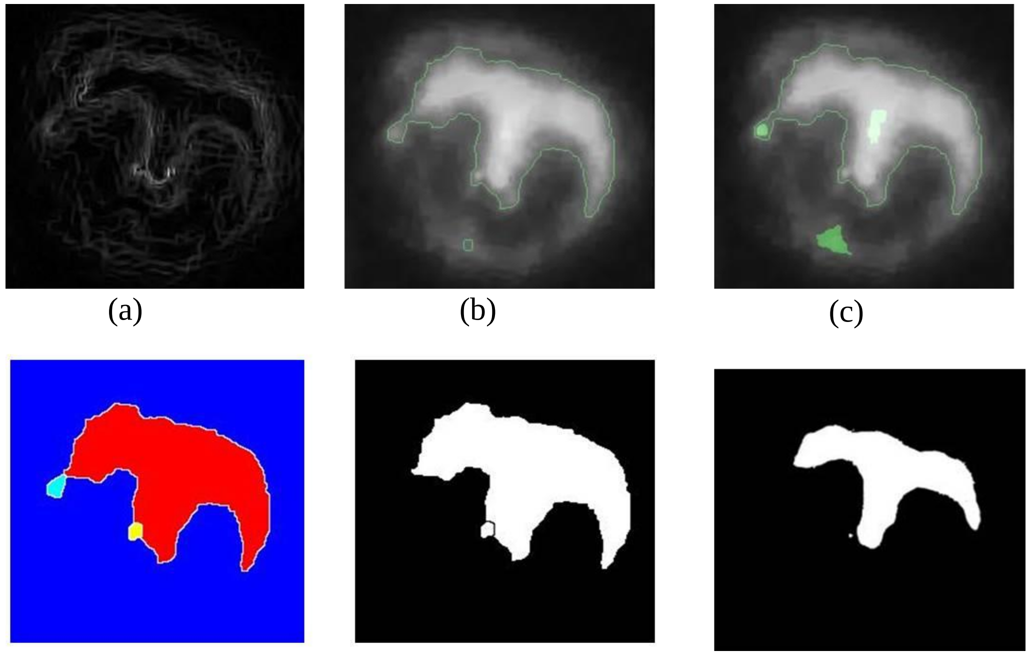
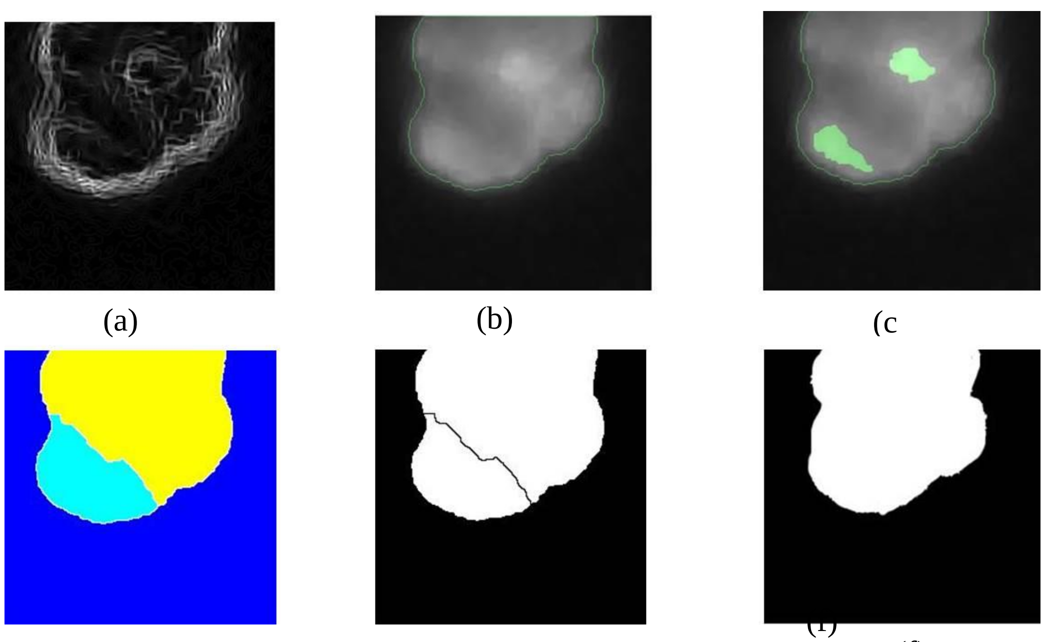


Figure 9: Watershed with catchment basin



(d) (e) (f)

Figure 10 above show the pre-processing and segmentation process of C-ANCA images. Image labeled (a) to (c) are the processing step involving DWT, detection of region of interest, and gradient magnitude. Meanwhile image (f) is the ground truth data and image (d) is the segmentation result of P-ANCA. (e) Is binary image for (d)



(d) (e) (f)

Figure 11 above show the pre-processing and segmentation process of P-ANCA images. Image labeled (a) to (c) are the processing step involving DWT, detection of region of interest, and gradient magnitude. Meanwhile image (f) is the ground truth data and image (d) is the segmentation result of P-ANCA. (e) Is binary image for (d)

4.4 Quantitative evaluation of segmentation using Rand Index

Reliability of the segmentation is evaluated in this section. As benchmark to segmented medical image, 40 ground truth dataset was created manually which correspond to the medical images that will be segmented. In order to make sure the ground truth dataset has high accuracy, the ground truth are created using hand. Ground truth refers to a binary image with value '1' represent foreground image and value '0' is the background. One precaution during preparing the ground the size must same with the segmentation image. Sample of original and corresponding ground truth image is presented below:

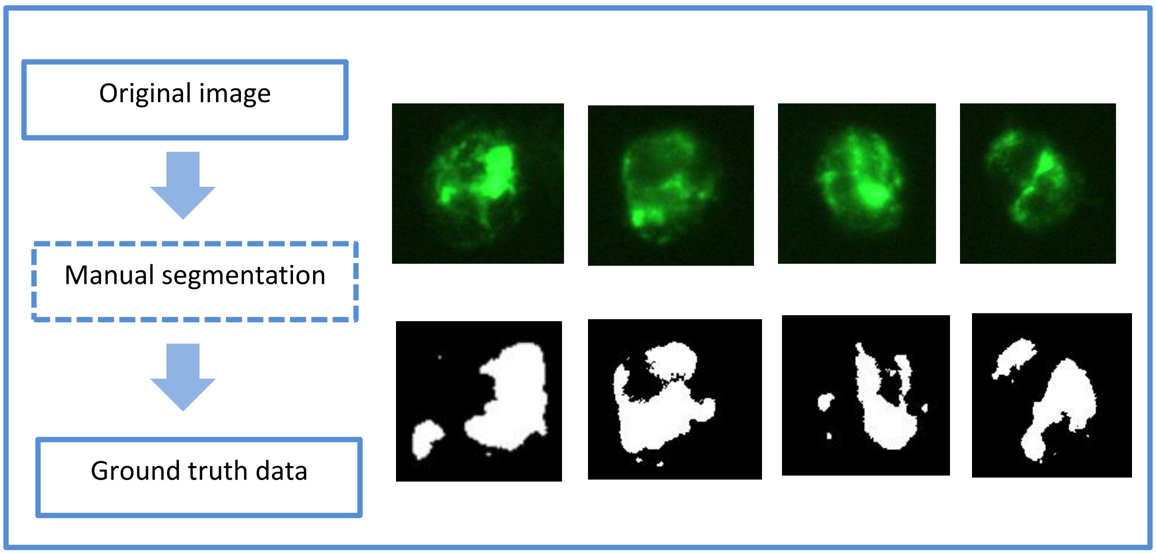


Figure 12: Sample of original image with corresponding ground truth

Qualitative evaluation is not reliable because the interpretation of segmentation result varies from one person to other person. Thus Rand Index is used to evaluate the performance of segmentation quantitatively. This method work by calculating the number of matched pixel on ground truth with the segmentation result and divided by total pixel pairs. The output value is ranging between 0 and 1 where the higher the RI value, the greater is the performance of the segmentation technique. The formula is given as below:

RI = correct pixel pair / total pixels pair…………………………….…. *(2)*

Each segmented image for 2 staining pattern, C-ANCA and P-ANCA has been evaluated using RI. The result is tabulated as below:

No	Image	Rand Index (RI)
1	C-ANCA1	0.9758
2	C-ANCA2	0.9689
3	C-ANCA3	0.9746
4	C-ANCA4	0.8908
5	C-ANCA5	0.9439
6	C-ANCA6	0.8631
7	C-ANCA7	0.8212
8	C-ANCA8	0.934
9	C-ANCA9	0.8933
10	C-ANCA10	0.8731
11	C-ANCA11	0.7965
12	C-ANCA12	0.7132
13	C-ANCA13	0.8717
14	C-ANCA14	0.7994
15	C-ANCA15	0.8484
16	C-ANCA16	0.7628
17	C-ANCA17	0.7355
18	C-ANCA18	0.7495
19	C-ANCA19	0.736
20	C-ANCA20	0.6829
21	C-ANCA21	0.8915
22	C-ANCA22	0.9423
23	C-ANCA23	0.9238
24	C-ANCA24	0.9451
25	C-ANCA25	0.9679
26	C-ANCA26	0.9071
27	C-ANCA27	0.9458
28	C-ANCA28	0.9414
29	C-ANCA29	0.9571
30	C-ANCA30	0.9287

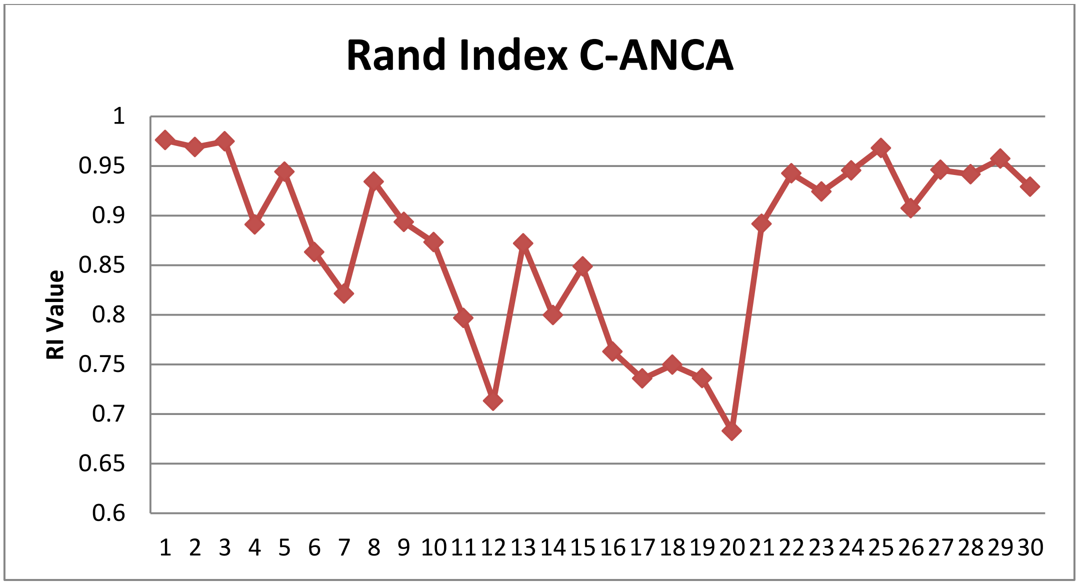
Table 6: Rand Index for P-ANCA pattern

Graph below are show the RI for P-ANCA images. Total of 30 segmented images be evaluated using Rand Index.

No	**Image**	**Rand Index(RI)**
1	P-ANCA1	0.8222
2	P-ANCA2	0.8892
3	P-ANCA3	0.8802
4	P-ANCA4	0.8656
5	P-ANCA5	0.6732
6	P-ANCA6	0.8952
7	P-ANCA7	0.7987
8	P-ANCA8	0.8684
9	P-ANCA9	0.9208
10	P-ANCA10	0.8940
11	P-ANCA11	0.9540
12	P-ANCA12	0.8919
13	P-ANCA13	0.6632
14	P-ANCA14	0.7163
15	P-ANCA15	0.5708
16	P-ANCA16	0.9295
17	P-ANCA17	0.7546
18	P-ANCA18	0.7808
19	P-ANCA19	0.8198
20	P-ANCA20	0.7065
21	P-ANCA21	0.8705
22	P-ANCA22	0.8492
23	P-ANCA23	0.8776
24	P-ANCA24	0.8634
25	P-ANCA25	0.8540
26	P-ANCA26	0.8236
27	P-ANCA27	0.9020
28	P-ANCA28	0.8305
29	P-ANCA29	0.8875
30	P-ANCA30	0.9140

Table 7: Rand Index for C-ANCA pattern

Further analysis on the result is by the graph below:

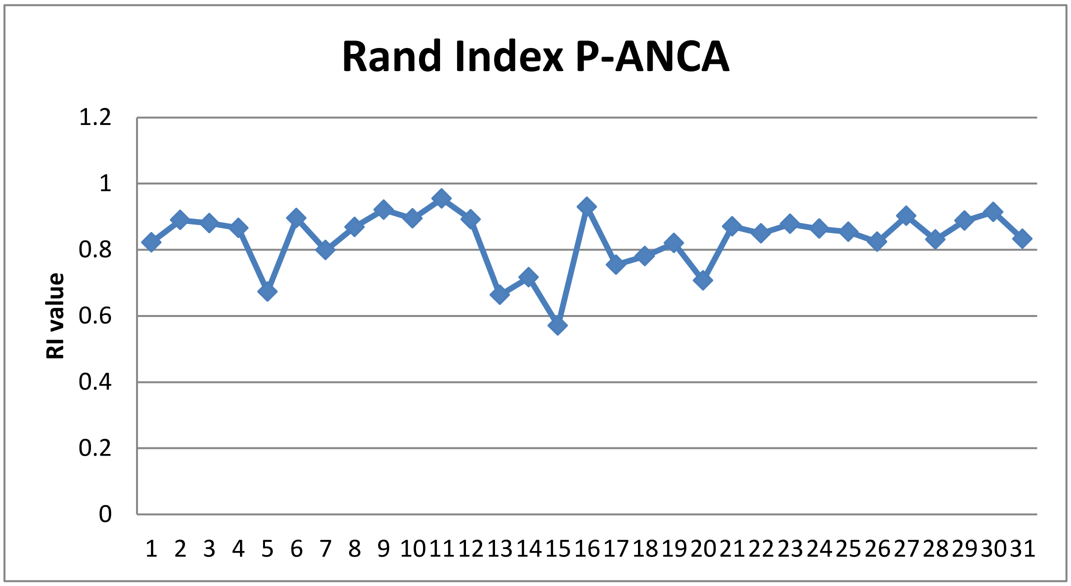


Graph 3: Rand Index for C-ANCA pattern

Graph 3 shows the Rand Index for watershed segmentation of C-ANCA. Y-axis shows the RI value while X-axis represent image. In the graph, there are 30 value of RI corresponding to 30 segmented images. The best segmentation result is RI equal to 1. In the graph below, highest value of RI is 0.97 meanwhile the lowest RI value is 0.682. Using visual inspection, segmented image with high RI value has high SNR value. This translate to image with low noise can avoid over segmentation and increase the segmentation accuracy.

Further interpretation of the data show each image have average of RI approximately 0.873. Using the dedicated RI value of each image, the accuracy of segmentation result can be calculated. The accuracy for C-ANCA segmentation is 87%. Formula for calculating accuracy is given as below:-

$$\textit{Segmentation Accuracy:} \ \sum \textit{Actual RI} \,/ \sum \textit{Desired RI} \times 100 \ldots\ldots\ldots\ldots \ (3)$$



Graph 4: Rand Index for P-ANCA pattern

Graph 4 shows the Rand Index for watershed segmentation of P-ANCA images.Y-axis shows the RI value while X-axis represent image. In the graph, there are 30 value of RI corresponding to 30 segmented images. The best segmentation result is RI equal to 1. In the graph below, highest value of RI is 0.954 meanwhile the lowest RI value is 0.578. Using visual inspection, segmented image with high RI value has high SNR value. This translate to image with low noise can avoid over segmentation and increase the segmentation accuracy.

Further interpretation of the data show each image have average of RI approximately 0.8322. Comparing average RI for P-ANCA with C-ANCA, RI for P-PANCA is lower. However with average RI per image is 0.8322, the segmentation is still reliable. Using the dedicated RI value of each image, the accuracy of segmentation result can be calculated. The accuracy for C-ANCA segmentation is 83%.

Chapter 5: Conclusion

Segmentation is an important step in image processing. Good segmentation will result in significant effect on texture analysis, feature extraction and classification. It able to enhance the information within an image to be further processed. Literally, it provide mean to differentiate between foreground and background.

Based on the result analysis above, wavelet transform manage to de-noise with increasing of SNR value within the image. Wavelet transform is a very powerful tool to convert image into signal and remove certain coefficient with respect to the predefined noise variance and unwanted signal. For C-ANCA, each image has increased the SNR with average of 17.85 meanwhile the increment rate of SNR for P-ANCA is 8.0841.

One of the reasons using DWT on medical images is to suppress noise and prevent over segmentation when watershed is applied on the image. This theory is proved when watershed algorithm manages to segment the cell without major over segmentation. The evaluation of segmentation is done using Rand Index. 30 images of positive C-ANCA are segmented and result in average rand index of 0.873 which translate to 87% accuracy. In addition, 30 images of P-ANCA are segmented and result in average Rand Index of 0.832 which translate to 83% accuracy. Therefore this research manages to obtain good segmentation result.

There are some recommendations for the project to be extended. To increase the reliability of this algorithm, ground truth data should be prepared by the professional and not by amateur. In this research, author prepare the ground truth data with supervision from literature review

In addition, it is almost impossible to completely reduce noise inside the image without knowing actual variance of the noise. Therefore to further increase the reliability of DWT for noise suppression, future DWT algorithm must be able to automatically estimate noise variance within an image

REFRENCES

[1] P. Soda, "Early Experiences in the Staining Pattern Classification of HEp-2 Slides," in *Computer-Based Medical Systems, 2007. CBMS '07. Twentieth IEEE International Symposium on*, 2007, pp. 219-224.

[2] W. Pollock, S. Jovanovich, and J. Savige, "Antineutrophil cytoplasmic antibody (ANCA) testing of routine sera varies in different laboratories but concordance is greater for cytoplasmic fluorescence (C-ANCA) and myeloperoxidase specificity (MPO-ANCA)," *Journal of Immunological Methods,* vol. 347, pp. 19-23, 8/15/ 2009.

[3] J. Savige, W. Pollock, and M. Trevisin, "What do antineutrophil cytoplasmic antibodies (ANCA) tell us?," *Best Practice & Research Clinical Rheumatology,* vol. 19, pp. 263-276, 4// 2005.

[4] A. Rigon, P. Soda, D. Zennaro, G. Iannello, and A. Afeltra, *Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose* vol. 72b, 2007.

[5] S. U. Indira and A. C. Ramesh, "Image Segmentation Using Artificial Neural Network and Genetic Algorithm: A Comparative Analysis," in *Process Automation, Control and Computing (PACC), 2011 International Conference on*, 2011, pp. 1-6.

[6] C. Ai-Shan, L. Po-Ting, L. Yi-Ping, H. Yu-Len, and H. Tsu-Yi, "Automatic detection of antinuclear autoantibodies cells in indirect immunofluorescence images," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, 2010, pp. 137-140.

[7] P. Elbischger, S. Geerts, K. Sander, G. Ziervogel-Lukas, and P. Sinah, "Algorithmic framework for HEp-2 fluorescence pattern classification to aid auto-immune diseases diagnosis," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, 2009, pp. 562-565.

[8] Y. Wei, C. Chang, T. Jia, and X. Xu, "Segmentation of regions of interest in lung CT images based on 2-D OTSU optimized by genetic algorithm," in *Control and Decision Conference, 2009. CCDC '09. Chinese*, 2009, pp. 5185-5189.

[9] H. Yu-Len, J. Yu-Lang, H. Tsu-Yi, and C. Chia-Wei, "Adaptive Automatic Segmentation of HEp-2 Cells in Indirect Immunofluorescence Images," in *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC '08. IEEE International Conference on*, 2008, pp. 418-422.

[10] S. L. S. Abdullah, H. A. Hambali, and N. Jamil, "Segmentation of Natural Images Using an Improved Thresholding-Based Technique," *Procedia Engineering,* vol. 41, pp. 938-944, // 2012.

[11] W. S. Gan, "Application of neural networks to the processing of medical images," in *Neural Networks, 1991. 1991 IEEE International Joint Conference on*, 1991, pp. 300-306 vol.1.

[12] L. Ran, "Medical Image Segmentation Based on Watershed Transformation and Rough Sets," in *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, 2010, pp. 1-5.

[13] J.-x. Chen and S. Liu, "A medical image segmentation method based on watershed transform," in *Computer and Information Technology, 2005. CIT 2005. The Fifth International Conference on*, 2005, pp. 634-638.

[14] H. Wei, Z. Sheng, and Y. Shuzhi, "One improved watershed transform for medical image segmentation," in *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, 2010, pp. V3-594-V3-598.

[15] C. R. Jung and J. Scharcanski, "Robust watershed segmentation using wavelets," *Image and Vision Computing,* vol. 23, pp. 661-669, 7/1/ 2005.

[16] A. Prochazka, O. Vysata, and E. Jerhotova, "Wavelet use for reduction of watershed transform over-segmentation in biomedical images processing," in *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, 2010, pp. 1-4.

[17] L. Haihua, C. Zhouhui, C. Xinhao, and C. Yaguang, "Multiresolution Medical Image Segmentation Based on Wavelet Transform," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2005, pp. 3418-3421.

[18] C. R. Jung and J. Scharcanski, "Robust watershed segmentation using the wavelet transform," in *Computer Graphics and Image Processing, 2002. Proceedings. XV Brazilian Symposium on*, 2002, pp. 131-137.

[19] C. Yu-hua, G. Li-Qun, L. Shun, and T. Lei, "Wavelet-based Watershed for Image Segmentation Algorithm," in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, 2006, pp. 9595-9599.

[20] U. Snekhalatha and M. Anburajan, "Dual tree wavelet transform based watershed algorithm for image segmentation in hand radiographs of arthritis patients and classification using BPN neural network," in *Information and Communication Technologies (WICT), 2012 World Congress on*, 2012, pp. 448-452.

APPENDIX A

CODING FOR DWT

```
clear;
close all;

%input image
I = imread('panca11.jpg');
figure, imshow(I),title('original');


%ground truth image
manual = imread('pbw11.jpg');
figure, imshow(manual),title('ground truth');


%converting image to gray
ImagetoGray = rgb2gray(I);
Igray=ImagetoGray(:,:,1);
figure, imshow(Igray),title('gray image');

ima1=double(Igray);
v=10;
pima=mean(mean(ima1.^2));
snro=10*log10(pima/(v^2))
[ld,hd,lr,hr]=wfilters('bior1.3');
wbase=2;
J=3;
[S,HW,WH,WW,etl] = ocwt2d(ima1,ld,hd,J);
s=[3 6 9 ];
rWW=denss(WW,v,wbase,1,s);
rWH=denss(WH,v,wbase,2,s);
rHW=denss(HW,v,wbase,3,s);
rima=iocwt2d(S,rHW,rWH,rWW,etl,lr,hr);
err=ima1-rima;
perr=mean(mean(err.^2));

snrss=10*log10(pima/perr)
figure ;clf;
imshow(rima,[0 255]); title('DWT image')


%cnvert to normal image
new = uint8(rima);

hy = fspecial('sobel');
hx = hy';
Iy = imfilter(double(new), hy, 'replicate');
Ix = imfilter(double(new), hx, 'replicate');
gradmag = sqrt(Ix.^2 + Iy.^2);
figure, imshow(gradmag,[]), title('gradmag')
```

APPENDIX B

CODING FOR IMAGE PRE PROCESSING

```
%median filtering
im3=medfilt2(new,[3 3]);
I_eq=im3;

%converting the image to binary (1,0)
bw = im2bw(I_eq, graythresh(I_eq));
figure, imshow(bw), title('bw image');

%filling holes iteratively
bw2 = imfill(bw,'holes');
figure, imshow(bw2), title('bw2 image');
bw3 = imopen(bw2, ones(5,5));
figure, imshow(bw3), title('bw3 image');
bw4 = bwareaopen(bw3, 40);
figure, imshow(bw4), title('bw4 image');
bw4_perim = bwperim(bw4);
figure, imshow(bw4_perim), title('b4 perim image');

%overlay images
overlay1 = imoverlay(I_eq, bw4_perim, [.3 1 .3]);
figure, imshow(overlay1), title('Image overlay');

%construct image masking
mask_em = imextendedmax(I_eq, 15);
figure, imshow(mask_em), title('Image mask');
mask_em1 = imclose(mask_em, ones(5,5));
figure, imshow(mask_em1), title('Image mask 2');

%open close hole, image masking and overlay
mask_em2 = imfill(mask_em1, 'holes');
figure, imshow(mask_em2), title('hole close');
mask_em3 = bwareaopen(mask_em2, 40);
figure, imshow(mask_em3), title('Image mask 3');
overlay2 = imoverlay(I_eq, bw4_perim | mask_em3, [.3 1 .3]);
```

APPENDIX C

CODING FOR WATERSHED AND RI

```
figure, imshow(overlay2), title('overlay 2');
I_eq_c = imcomplement(I_eq);
I_mod = imimposemin(I_eq_c, ~bw4 | mask_em);

%apply watershed to the preprocessed image
L = watershed(I_mod);
figure, imshow(label2rgb(L)), title('watershed image');


S = watershed(I_mod);
figure, imshow(label2rgb(S)), title('watershed image2');


binary = im2bw(L, graythresh(L));
figure, imshow(binary); title('binary 1,0');


compare_segmentations(binary,manual)

figure, imshow(I_eq), hold on
himage = imshow(label2rgb(L));
set(himage, 'AlphaData', 0.3);
title('Lrgb superimposed transparently on original image')

binary2 = im2bw(I_mod, graythresh(I_mod));
figure, imshow(binary2); title('binary2 1,0');

RI = compare_segmentations(binary,manual)


[cA1,cH1, cV1, cD1] = dwt2(binary, 'dB1');

[cA,cH, cV, cD] = dwt2(manual, 'dB1');


l=mean(cA1)
m=mean(cA);

figure, plot(m, 'yellow');
figure, plot(l, 'red');

%Especially important is the mask line:
mask_em = imextendedmax(I_eq, 15);
```

APPENDIX C

CODING FOR RI

```
function
[ri,gce,vi]=compare_segmentations(sampleLabels1,sampleLabels2)

% compare_segmentations
%
%   Computes several simple segmentation benchmarks. Written for
use with
%   images, but works for generic segmentation as well (i.e. if
the
%   sampleLabels inputs are just lists of labels, rather than
rectangular
%   arrays).
%
%   The measures:
%       Rand Index
%       Global Consistency Error
%       Variation of Information
%
%   The Rand Index can be easily extended to the Probabilistic
Rand Index
%   by averaging the result across all human segmentations of a
given
%   image:
%       PRI = 1/K sum_1^K RI( seg, humanSeg_K ).
%   With a little more work, this can also be extended to the
Normalized
%   PRI.
%
%   Inputs:
%       sampleLabels1 - n x m array whose entries are integers
between 1
%                       and K1
%       sampleLabels2 - n x m (sample size as sampleLabels1) array
whose
%                       entries are integers between 1 and K2 (not
%                       necessarily the same as K1).
%   Outputs:
%       ri  - Rand Index
%       gce - Global Consistency Error
%       vi  - Variation of Information
%
%   NOTE:
%       There are a few formulas here that look less-
straightforward (e.g.
%       the log2_quotient function). These exist to handle corner
cases
%       where some of the groups are empty, and quantities like 0
*
%       log(0/0) arise...
%
%   Oct. 2006
%       Questions? John Wright - jnwright@uiuc.edu
```

```
[imWidth,imHeight]=size(sampleLabels1);
[imWidth2,imHeight2]=size(sampleLabels2);
N=imWidth*imHeight;
if (imWidth~=imWidth2)||(imHeight~=imHeight2)
    disp( 'Input sizes: ' );
    disp( size(sampleLabels1) );
    disp( size(sampleLabels2) );
    error('Input sizes do not match in compare_segmentations.m');
end;

% make the group indices start at 1
if min(min(sampleLabels1)) < 1
    sampleLabels1 = sampleLabels1 - min(min(sampleLabels1)) + 1;
end
if min(min(sampleLabels2)) < 1
    sampleLabels2 = sampleLabels2 - min(min(sampleLabels2)) + 1;
end

segmentcount1=max(max(sampleLabels1));
segmentcount2=max(max(sampleLabels2));

% compute the count matrix
%  from this we can quickly compute rand index, GCE, VOI, ect...
n=zeros(segmentcount1,segmentcount2);

for i=1:imWidth
    for j=1:imHeight
        u=sampleLabels1(i,j);
        v=sampleLabels2(i,j);
        n(u,v)=n(u,v)+1;
    end;
end;

ri = rand_index(n);
gce = global_consistancy_error(n);
vi = variation_of_information(n);

return;


% the performance measures

% the rand index, in [0,1] ... higher => better
%  fast computation is based on equation (2.2) of Rand's paper.
function ri = rand_index(n)
N = sum(sum(n));
n_u=sum(n,2);
n_v=sum(n,1);
N_choose_2=N*(N-1)/2;
ri = 1 - ( sum(n_u .* n_u)/2 + sum(n_v .* n_v)/2 - sum(sum(n.*n))
)/N_choose_2;

% global consistancy error (from BSDS ICCV 01 paper) ... lower =>
better
```

```
function gce = global_consistancy_error(n)
N = sum(sum(n));
marginal_1 = sum(n,2);
marginal_2 = sum(n,1);
% the hackery is to protect against cases where some of the
marginals are
% zero (should never happen, but seems to...)
E1 = 1 - sum( sum(n.*n,2) ./ (marginal_1 + (marginal_1 == 0)) ) /
N;
E2 = 1 - sum( sum(n.*n,1) ./ (marginal_2 + (marginal_2 == 0)) ) /
N;
gce = min( E1, E2 );


% variation of information a "distance", in (0,vi_max] ... lower
=> better
function vi = variation_of_information(n)
N = sum(sum(n));
joint = n / N; % the joint pmf of the two labels
marginal_2 = sum(joint,1);  % row vector
marginal_1 = sum(joint,2);  % column vector
H1 = - sum( marginal_1 .* log2(marginal_1 + (marginal_1 == 0) ) );
% entropy of the first label
H2 = - sum( marginal_2 .* log2(marginal_2 + (marginal_2 == 0) ) );
% entropy of the second label
MI = sum(sum( joint .* log2_quotient( joint, marginal_1*marginal_2
)  )); % mutual information
vi = H1 + H2 - 2 * MI;

% log2_quotient
%   helper function for computing the mutual information
%   returns a matrix whose ij entry is
%     log2( a_ij / b_ij )   if a_ij, b_ij > 0
%     0                     if a_ij is 0
%     log2( a_ij + 1 )      if a_ij > 0 but b_ij = 0 (this
behavior should
%                                                 not be encountered in
practice!
function lq = log2_quotient( A, B )
lq = log2( (A + ((A==0).*B) + (B==0)) ./ (B + (B==0)) );
```

APPENDIX D

GANT CHART FYP 2

	Week No													
Activity/week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
developing preprocessing algorithm	■	■												
develop filter			■	■										
develop wavelet algorithm					■	■								
cropping the image by cell							■							
develop watershed algorithm					■		■							
applying algorithm to cell images							■	■						
develop algorithm for verifying result								■	■					
analyze the performance of algorithm										■	■			
compute final working algorithm											■	■		
progress report								■						
final presentation													■	