M. Tech. (Computer Science) Dissertation Series

Detection and Classification of Psoriasis in Histopathology Images

a dissertation submitted in partial fulfilment of the requirements for the M.Tech. (Computer Science) degree of Indian Statistical Institute

By

Kushal Sen

under the supervision of

Dr. Utpal Garain

Computer Vision and Pattern Recognition Unit
Indian Statistical Institute
203, Barrackpore Trunk Road
Kolkata-700 108

July 14, 2015

Acknowledgments

First and foremost I would like to thank my mentor **Dr. Utpal Garain, Associate Professor**, CVPRU, ISI Kolkata. He has been a constant inspiration throughout the work . I thank him for appreciating and criticizing my work as and when it was necessary.

I would like to thank Dr. Raghunath Chatterjee, Assistant Professor, HGU, ISI Kolkata for providing the images, providing an insight, better understanding and real life significance of the problem, and the deterministic feature of the images.

I would also like to thank Prof. Dipti Prasad Mukherjee,Professor and Head, ECSU, ISI Kolkata, who evaluated my work during the mid term evaluation. It was his insight that had lead to a better understanding of the problem and the approaches to be taken. He was kind enough to share his experience for the benefit of my work.

I would also like to thank senior and colleague Anabik Pal, JRF, CVPRU, ISI Kolkata for bearing with me and guiding me through the work, providing important suggestions and helping me through the phases of development in this work.

I would also like to thank ISI Kolkata, for providing me the opportunity to be acquainted with so many great minds,my teachers and my friends and be inspired from them.I thank all of them for making me a part of my family.

I thank my parents, whose support and constant scolding made me the person I am now.

Thank You

Declaration

I, **Kushal Sen** (**CS1307**), registered as a student of **M. Tech** program in **Computer Science**, **Indian Statistical Institute**, **Kolkata** do hereby submit my Dissertation Report entitled "**Detection and Classification of Psoriasis in Histopathology Images**". I certify

1. The work contained in this Dissertation Report is original and has been done by me under the guidance of my supervisor.

2. The material contained in this Dissertation Report has not been submitted to any University or Institute for the award of any degree.

3. I followed by guidelines provided by the Institute in preparing the report.

4. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of report and giving their details in the bibliography.

.....

Place : ISI, Kolkata **Kushal Sen**

Date : July, 2015 **(CS1307)**

Ceritificate of Approval

This is to certify that this thesis titled **"Detection & Classification of Psoriasis in Histopathological Images"** submitted by Kushal Sen , embodies the work done under my supervision.

Dr. Utpal Garain

Comuter Vision and Pattern Recognition Unit,

Indian Statistical Institute,

Kolkata

Abstract

Recent advances in imaging techniques has lead better visual representation of the internals of our body for clinical analysis and medical intervention, however the task is tedious and subject to interpreter variability. An automated quantitative analysis of the images would not only relieve us of the human effort, but considerably reduce the inaccuracies involved. The current work explores the techniques of image processing and analysis to extract vital information out of psoriasis histopathology images, and discuss a method of classifying these into diseased or non diseased classes.

Contents

Chapter 1

Introduction

1.1 Background

1.1.1 Psoriasis

Psoriasis also known as psoriasis vulgaris is a immune triggering, chronic, relapsing, systemic disease characterized by skin lesions including red, scaly patches, papules, and plaques, which usually itch. The skin lesions seen in psoriasis may vary in severity from minor localized patches to complete body coverage.The disease affects 24% of the general population.Plaque psoriasis, the most common form, typically manifests as red and white scaly patches on the top layer of the skin. Skin cells rapidly accumulate at these plaque sites and create a silvery-white appearance. Plaques frequently occur on the skin of the elbows and knees, but can affect any area, including the scalp, palms of hands, and soles of feet, and genitals. Psoriasis is non-contagious , it cannot transmit from one to the other [9].

The causes of psoriasis are not fully understood. It is not purely

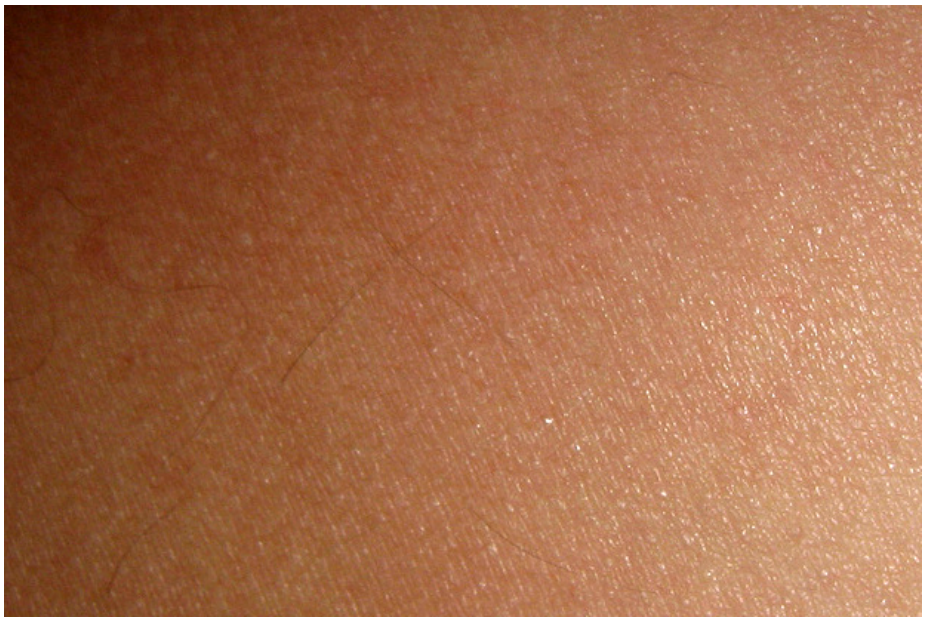
a skin disorder and can have a negative impact on many organ systems [3]. Psoriasis has been associated with an increased risk of certain cancers, cardiovascular disease, and other immune-mediated disorders such as Crohn's disease and ulcerative colitis. It is generally considered a **"genetic"** disease, thought to be triggered or influenced by environmental factors [3]. Psoriasis develops when the immune system mistakes a normal skin cell for a pathogen, and sends out faulty signals that cause overproduction of new skin cells.

Though many treatments are available, psoriasis can be difficult to treat due to the fact that a particular type of treatment may not be effective for all affected patient and because of their recurrent nature.

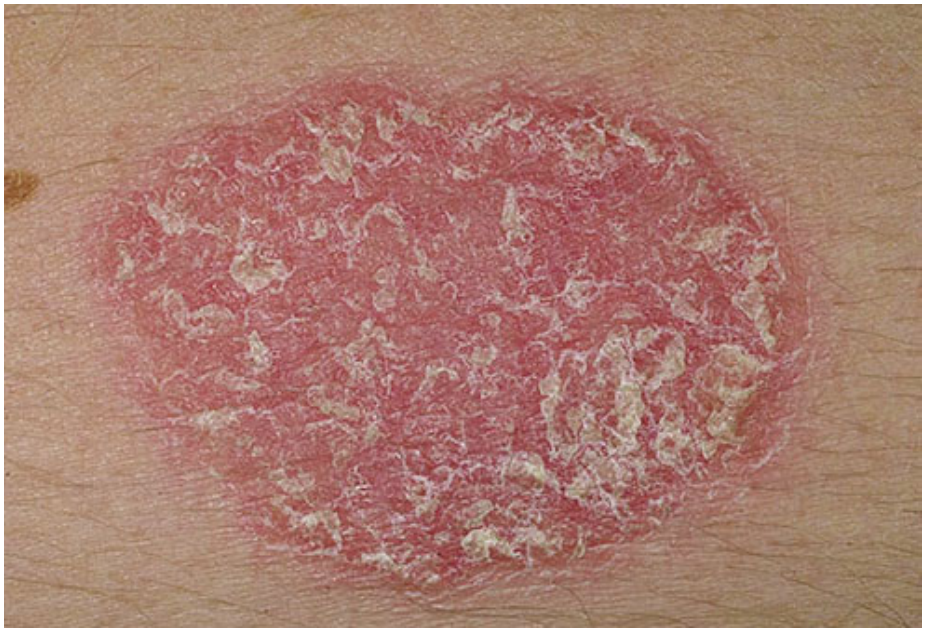
However there is no known cure for psoriasis but medication can ease the discomfort associated with the disease. People suffering from psoriasis generally exhibit strong emotional reactions like anger,sadness or embarrassment.

Discovering the genes that cause psoriasis will help identify the cause of the disease. Understanding the genetic component of psoriasis will go beyond showing that the immune system is activated, to telling us why it is activated and how that activation leads to defects in the skin. This will open the door to better control of psoriasis through more precise and effective treatments.

Once all the genetic factors causing the disease are found, researchers will be able to study why some people get psoriasis and others don't. By understanding how people who develop psoriasis are genetically different from people who do not, scientists may be



(a) Normal Skin



(b) Psoriasis Skin

Figure 1.1: Skin Images

able to create treatments that "correct" the behaviour of cells.

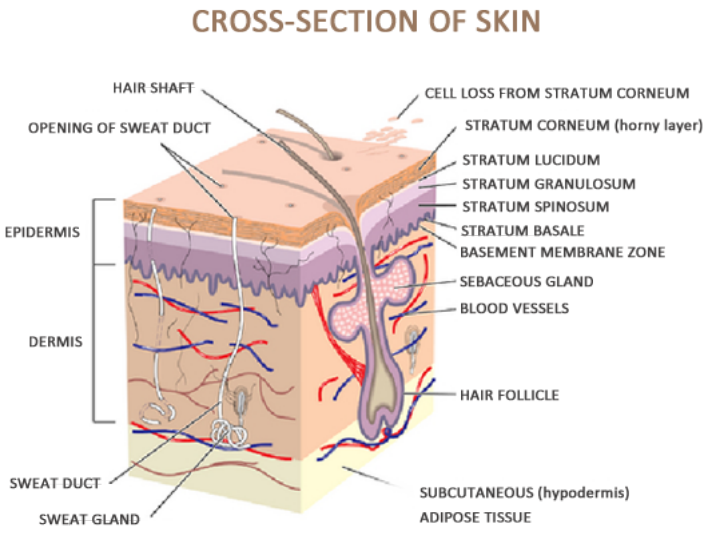


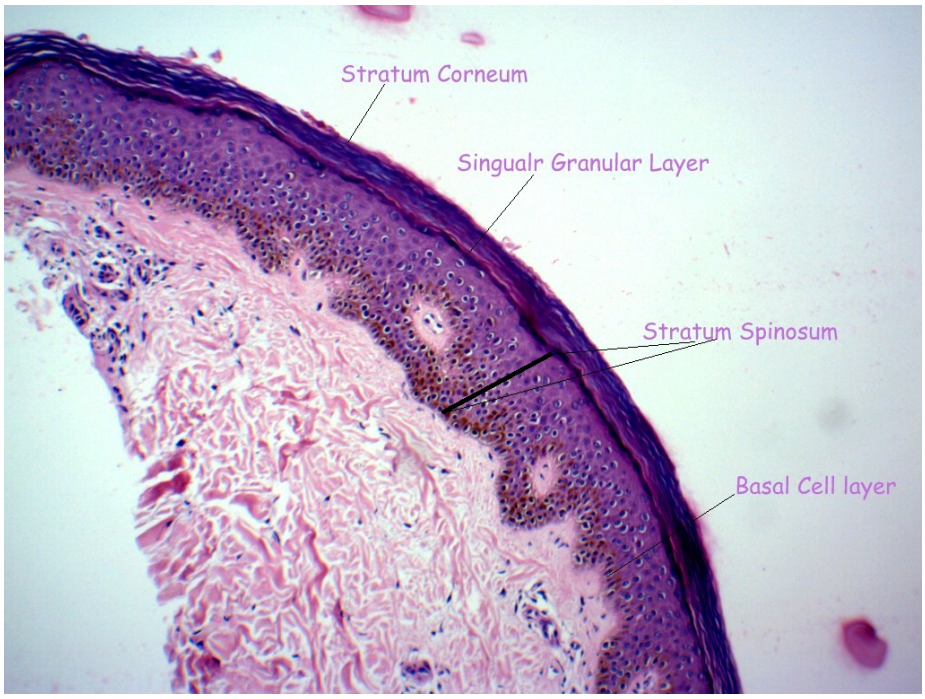
Figure 1.2: Skin cross section

1.1.2 Histopathology Images

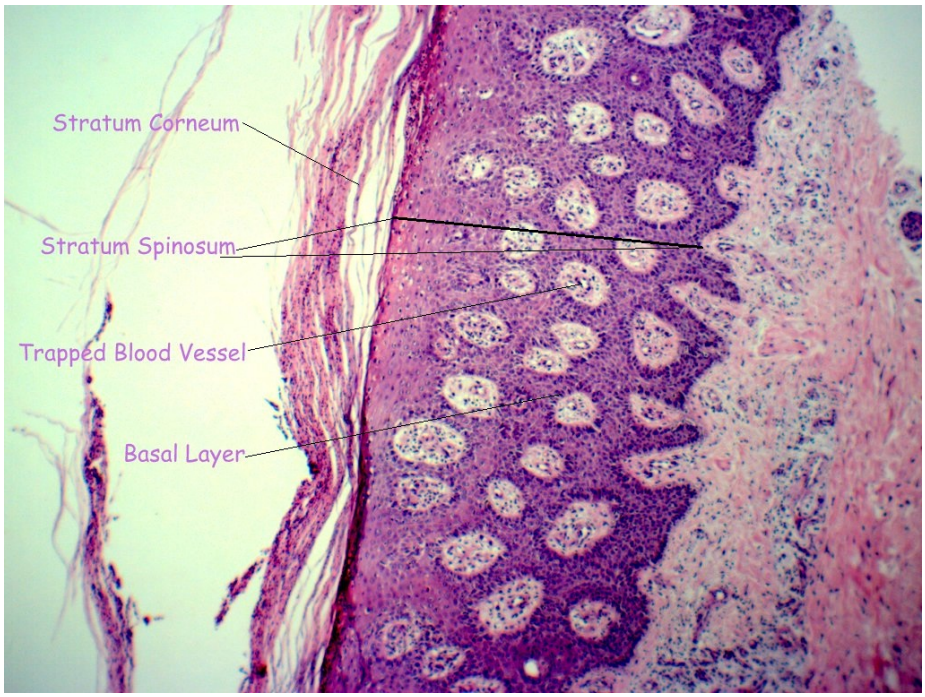
Histopathology is the study of the microscopic examination of the tissue often performed to study the manifestation of the disease. The most commonly used stain in histopathology is Hematoxylin and Eosin (abbr: H& E). Hematoxylin is used to stain nuclei blue, while eosin stains cytoplasm and the extracellular connective tissue matrix pink.

Normal case

Here are list of distinctive feature of Skin Histopathology section in case of Non Psoriasis skin Stratum Corneum : Thin and Continuous Even layer in normal issue Granular layer: only a single layer present Stratum Spinosum: Even width, continuous, smooth Basal Cell layer:Monolayer cell on the inner edge of stratum spinosum As



(a) Normal Skin



(b) Psoriasis Skin

Figure 1.3: Skin Histopathology Images

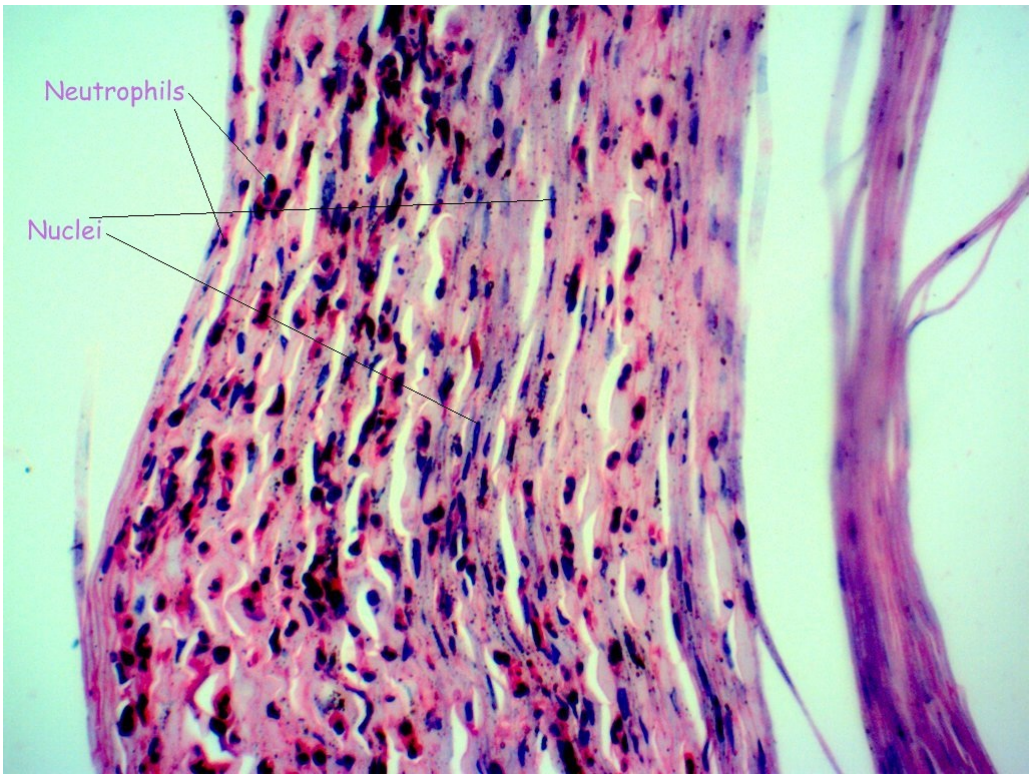


Figure 1.4: Stratum Corneum

referred to by 1.3a on page 11.

Diseased case

Here are list of distinctive feature of Skin Histopathology section in case of Non Psoriasis skin Stratum Corneum: Tearing apart, Thickening. Stratum Spinosum : Thickening, Non Uniform Elongation. Basal Layer: Trapper inside Stratum Spinosum. Blood Vessel : Entrapped within the Stratum Spinosum. Granular Layer:Absent. As referred to by 1.3b on page 11.

Parakeratosis

Parakeratosis is a class of psoriasis which is characterised by the retention of nuclei in the stratum corneum As referred to by 1.4 on page 12.

Munro's microabscess

Munro's microabscess is an abscess (collection of neutrophils) in the stratum corneum of the epidermis due to the infiltration of neutrophils from papillary dermis into the epidermal stratum corneum. As referred to by 1.4 on page 12.

1.2 Problem Statement

Psoriasis possesses a major threat, the recurrent nature of the disease, the non-uniform treatment and the unsuccessful attempt to understand the cause behind the disease lead to a bigger problem.The key may lie within the genetic components of an individual, how they are affected by environmental factors, the treatment they respond to. A study of histopathology images1.3a in relation to psoriasis may hint towards a genetic linkage and a better treatment for psoriatic patient. A major challenge in the detection and classification of psoriasis from histopathology images require an accurate segmentation of the cytological features in the microscopic tissue images, An automated approach is required as more and more imaging techniques are developed there is a requirement for quantizing the approach of detection of abnormality using these images, as

currently the examination mainly relies on manual interpretation which is subject to observers variability[19]. Separate identification of the stratum corneum, stratum spinosum, rete pegs, blood vessel entrapment would allow us to extract determining feature for the determination of psoriasis and on further investigation of the stratum corneum for infiltration of neutrophil and presence of nuclei in otherwise dead protein stack would lead to classification of the disease to Perakeratosis or Munro's Microabcess

So, essentially the problem in concern is reduced to automatic identification of different types of psoriasis, which in turn is segmented as:

- background removal
- lighting correction
- identifying the stratum corneum
- identifying the stratum spinosum
- identifying the boundaries between them
- severity of rete peg elongation
- presence of nuclei in stratum corneum
- presence of neutrophil in stratum corneum

1.3 Previous works

1.3.1 MSSC using Fuzzy Texture & 2D Fuzzy Colour

Earlier approaches to detect psoriasis from skin images include trying to separate out the psoriasis region from the unaffected parts of the skin using image segmentation approaches, a proposed method uses Multiresolution based orthogonal subspace classifier.

The aforementioned approach uses a 2D Fuzzy Color Histogram and a Fuzzy Texture Spectrum to obtain a feature vector of length 486. Split the image into overlapping blocks of a size r , r is input parameter, and detect homogeneous regions within the image, used as training regions, classified into normal skin, and psoriasis affected skin. The rest of the non-homogeneous portions of the image are then compared with [17] the signatures developed and assigned to either the diseased section or the non-affected part, the gold standard of comparison is determined by Psoriasis Area Severity Index (**PASI**) score.

Fan et al [11] showed that most image segementation can be categorized into 4 classes, i.e.

- Threshold technique
- Boundary detection approaches
- Region growing methods and
- Hybrid techniques

Thresholding techniques are often inefficient and make bad decisions[17]. Boundary detection techniques like Sobel, Robert and Canny use

spatial information and contour detection and are sensitive to noisy images giving rise to unnecessary edges. Orthogonal Subspace Projection [OSP] has been applied to hyperspectral images.[13] The OSP classifier annihilate all undesired signature within a pixel and uses an optimal detector to extract desired signature. It is difficult generally to obtain such prior knowledge a posteriori. Orthogonal Subspace classifier [21] [OSC] works on unconstrained least square estimation. It first decomposes an observed pixel into a signature space and a noise space to eliminate the noise successfully. Taur et al. [17]shows an automatic method to segment psoriasis image using a multiresolution based signature subspace classifier [MSSC].

The proposed method deals the image in two parts feature extraction and segmentation. The feature vector consists of a fuzzy teture spectrum[18] and two dimensional fuzzy HS color histogram in HSV space, where H stands for hue, S for saturation both representing the chroma part, and V - value represents the luma. The color information is hidden in the H and S component, the V only contributes to the light and shadow affects[1], and hence is not used. The feature vector is designed to locate homogeneous region within the image and group them into two class, to be later used as the training region, ie normal skin and psoriasis affected region.Once only the non-homogeneous blocks are left, to achieve computational efficiency nine small portions are chosen1.5, if all nine belong to the same class, the bigger block is assumed to belong to that class, otherwise it is broken into four equal parts and the process repeated.

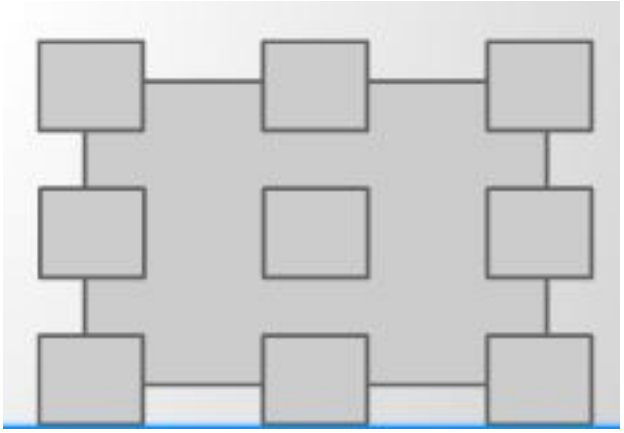


Figure 1.5: Multi-resolution Classifier

Obtaining the feature vectors are an interesting part of this study, the Fuzzy Texture Spectrum is described by taking a 3 x 3 block neighbourhood and marked as p, z, n depending on the difference value from the center pixel. The differences are plotted on a membership function1.6 to obtain the membership value marked as s, e, l respectively.

$s + e + l = 1$

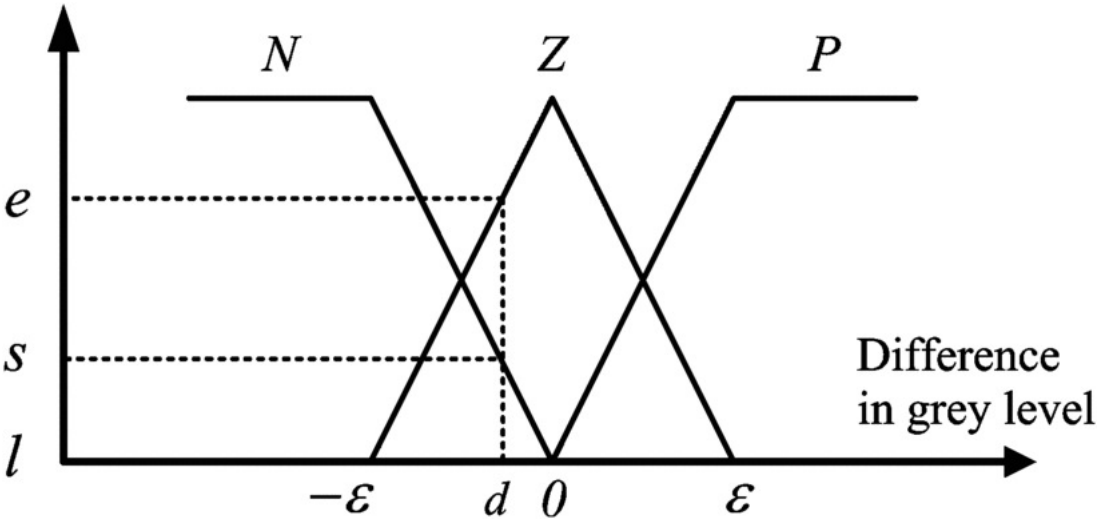


Figure 1.6: Fuzzy Membership function for e, s & l

so for all 8 neighbours

$s_a + e_a + l_a = 8$

which gives 45 integer solutions

each of the values of e_a, s_a, l_a will be fuzzified according to membership fucction 1.7 the minimum of the S,E,L is chosen to represent the fuzzy texture vector

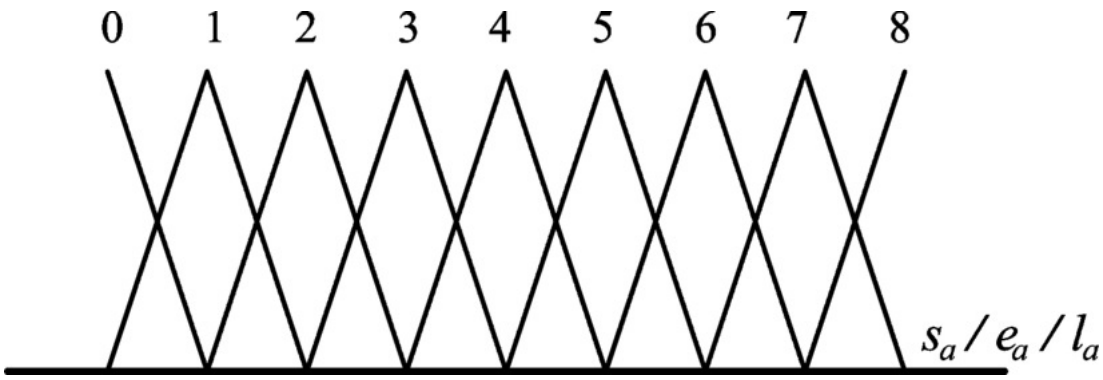


Figure 1.7: Fuzzy Membership function for $e_a, s_a \& l_a$

The two dimensional fuzzy color histogram is based on the hue-saturation value derived from the RBG value using equation

$$\begin{bmatrix} h \\ s \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & +\frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{R}{255} \\ \frac{G}{255} \\ \frac{B}{255} \end{bmatrix} \tag{1.1}$$

Discarding the top 0.05% and the bottom 0.05% of the hue/saturation values, and quantizing the H and S values to 21 discrete bins, we obtain a combination of 21 X 21 = 441 unique (H,S) pair. The membership of each such HS pair is again determined by the Fuzzy Membership Functions1.8

This two feature combined forms the feature vector of length 481.

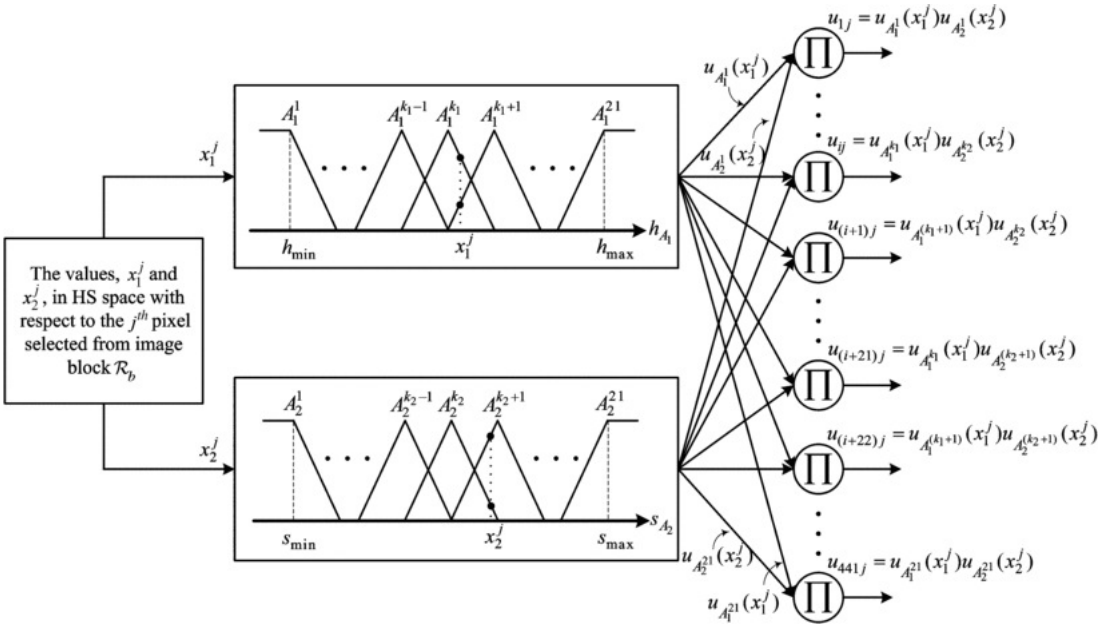


Figure 1.8: Fuzzy Membership function for H and S

Homogeneous Region Detection: A fixed size window is split into twelve sub-window, as shown in 1.9, then the if max(distance(sb_i, sb_j)) ¡= threshold , it is determined as a homogenous candidate, feature is again the 486 Dimensional vector, and distance is Manhattan. After

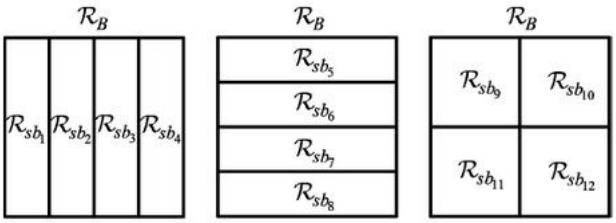


Figure 1.9: Homogeneous Region Detection Window

all the homogeneous regions have been detected, the furthest apart homogeneous region are considered to be either from the psoriasis region or from the normal skin region. The homogeneous region which shows higher amount of texture is considered to be from the

psoriasis region. Then all the homogeneous regions are classified as either as a psoriasis region or as normal skin region depending on the Manhattan distance from the initial sequence. Once the signature psoriasis region and normal skin regions have been identified, the rest of the nonhomogeneous part of the image is classified between these two using MSSC.

Implementing this algorithm as a part of the literature review, it failed to give promising result for the histopathology images,the texture of the histopathology images even for the same class were not as uniform as that of the skin images these algorithm was designed for.

1.3.2 Histopathology image segmentation

Histopathology refers to the study of tissue under the microscope for the purpose of understanding the manifestation of a disease.A biopsy is generally recommended when the physician suspects an abnormality within the tissue. A series of pre-examination steps are performed before obtaining the image we have seen earlier 1.3a

- **Fixing** is the process of reserving the biopsy samples to represent a living tissue both structurally and chemically as close as possible. It is a necessary step to determine the exact cause and extent of the abnormality we are seeking.

- **Processing** or 'dehydrating' constitute of replacing the tissue water particles by a medium which would solidify, thus enabling slicing the tissue into thin layers required to examine it under

the microscope.

- **Embedding** is done before the slicing ensuring hardening of the medium used to replace water and assure its transparency to the light of the background illumination.

- **Sectioning** is performed to have sufficiently thin samples of the tissue so as to able to visualise the intricate details of the tissue under the microscope.

- **Staining** is the most important part related to our study of any tissue under the microscope , there are a lot of staining methods in practice , such as Alician Blue,Cajal Stain,Giemsa Stain [4] and a few others. However our principle interest lies with Hematoxylin and Eosin **H & E**. H & E is the most common staining among histopathology images, Hematoxylin is used to stain nuclei blue and eosin stains cytoplasm and the extracellular connective tissue matrix pink.[4]

segmentation approaches related to histopathology images are inspired on feature as noted by a physician. We as humans can process a lot of texture information and hence detect region based on such homogeneity which are apparently invisible to the machine at a pixel level. Hence we take approaches which can find homogeneity in apparently haphazardly textured environment keeping in mind the domain specifics to segment out important component feature out of the image.

1.3.3 Graph Run Length Matrix

Identification of tissue components

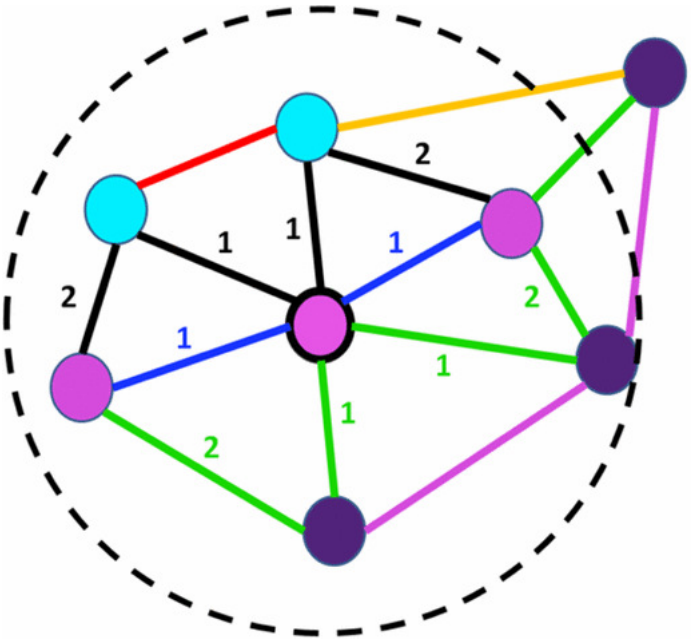
There is a scarcity of algorithm that work heterogeneous tissue image. H & E stained histopathology images mainly region which can be divided into the following colors **purple**,**pink** and **white** representing **nuclei**, **stromal** and **luminal** regions respectively [20]. The tissue image is not a random distribution of components, it has a pattern which alters with the existence of disease in the tissue component. The discussed approach uses Graph-Run-Length Matrix[19] Segmentation technique to segment out distinguishing feature of the psoriasis histopathology image. Following are the steps to perform the segmentation

Object decomposition

The first step in the process is to use a four-means clustering approach to approximately represent the primitives using black , pink, purple and white for the background , stromal , luminal and nuclei , they are then represented by circular primitives [6]. The black region is exempted as the background. A delaunay triangulation is then constructed on the centroids of this primitives. There are a total of six different types of edges formed on the construction of the delaunay triangle depending on the three different type primitives at its end point.

The image is now split into regions each of a window size of **W**, next the run length feature is calculated for each node within that

particular region. For each node a circular bounded region of radius **R** is defined, and the count of each of the run originating from the selected node R(c,l) with a run-length of l and an edge of type c is tabulated.



type\run	1	2	3
red	0	0	0
pink	0	0	0
black	0	2	0
blue	2	0	0
green	0	2	0
orange	0	0	0

Figure 1.10: Graph Run Length Table

Feature Extraction

Gray run length matrix proposed the following measure of descriptors [12]

$$Short\ Run\ Emphasis = \frac{1}{n_r}\sum_i\sum_j I(i,j)/j^2$$

$$Long\ Run\ Emphasis = \frac{1}{n_r} \sum_i \sum_j I(i,j) \times j^2$$

$$Graph\ Level\ Nonuniformity = \frac{1}{n_r} \sum_i \left(\sum_j I(i,j) \right)^2$$

$$Run\ Length\ Nonuniformity = \frac{1}{n_r} \sum_j \left(\sum_i I(i,j) \right)^2$$

$$Run\ Percentage = \frac{\sum_i \sum_j I(i,j)}{p}$$

Region growing

Region growing is achieved by merging of the primitives. For any primitive node a matrix containing all the run-length-matrices of all the primitives located within a radius R is taken into account and forms the feature descriptor of the Original centred node.

Euclidean distance between these descriptors are used to calculate the similarity.

The distance between every pair of adjacent primitives are calcu-

lated and merged if they are below a certain threshold. Component having a size larger than the given threshold are taken into account to form the seed regions. Now for the remaining primitives they are assigned to their adjacent seed points if they are closer than a certain threshold,iterative relaxation of the threshold is done till there are no primitives left. In the next step, two regions are merged only if their distance is less than a merge threshold.

After the process is done, a voronoi space is created to capture all the pixels in the image to their respective primitives to finally segment the image.

1.3.4 Detection of Nuclei

Detection of Nuclei is an essential part in determining the condition of the skin in consideration, presence of nuclei is the stratum corneum layer suggeest a condition of psoriasis known as parakeratosis. Histopathology images are routinely stained by Hematoxylin and Eosin, Hematoxylin stains the nuclei 'Blue'. The difficulties arising in segmenting a nuclei arises from , a 2D sectioning of a 3D structure, partially damaged nuclei , blurred pictures [15]. Extracting the foreground , the initial seed points are detected , and the segmenting them constitutes the basic procedures for the segmentation of the nuclei. A basic fault with thresholding technique lies with the under-segmentation of fused nuclei.[22] Which is mainly resolved by using the watershed model explained later2.6. Hongming et al[22] proposed an approach using an adaptive thresholding to detect the potential nuclei markers, isolates the independent nuclei via

an elliptical nuclei marker, nuclei which failed to be approximated to an ellipse were treated as clubbed nuclei, then a voting meachanism used to detect seeds of irregular shaped nuclei, and watershed modelling used to segment them further.

1.4 Contribution and Organization of this Thesis

This work aims to develop a novel methodology towards detection of psoriasis in case of histopathological images using state of the art techniques. The data i.e. the microscopic tissue images of human skins, containing psoriasis were obtained from **Dr. Raghunath Chatterjee** Human Genetics Unit of Indian Statistical Institute , Kolkata.

Chapter 2 contains a general discussion on image segmentation techniques. Chapter 3 talks about the approaches we have taken for background removal and extraction of region of interest, the detection of stratum layer. Chapter 4 Discusses about the classification parameter, and proposed future approach.

Chapter 2

Image Segmentation

Digital images are made up of discrete packets of uniform color information known as pixel information,pixels are the basic building blocks of any digital image.When we visualise an image we recognise this color information along with its spatial information which is later rendered in our mind enabling us to grasp the information within the image. Computers do not naturally have this ability to understand concepts or grasp information from an image only by looking at it, the individual pixel intensities do not carry enough information as to represent an object. The goal of image segmentation is to cluster pixels of similar type, or pixels from one object together. The absolute measure of the correctness of similarity is of course depended on the human judge. Generally image segmentation algorithm take into account color information,intensity,texture,contour.

We provide a brief discussion of some of the algorithm in literature which we would later refer to for segmentation purposes in our work.

2.1 Thresholding

Thresholding is the simplest of segmentation technique where an image is converted to a binary image with a 0 or 1 value, or may be 0 or 255 depending on a threshold

$$f(x,y) = \begin{cases} 0 & iff(x,y) < T \\ 1 & iff(x,y) \geq T \end{cases}$$

where T is the threshold value.

2.2 K-Means:

K- means[14] algorithm is a supervised clustering algorithm, where K needs to be determined beforehand , taking K as an input , the algorithm would select K random seed points and then classify each point as one of the K classes based on its distance from the seed points, after all the points are classified it would recompute the mean withing each class and find a new seed point. It is an iterative algorithm which will keep on computing new seed points until it converges ie the intra class variance is minimized, where all the points in a particular class stays in the class itself.

The algorithm:

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \leq \left\| x_p - m_j^{(t)} \right\|^2 \ \forall j, 1 \leq j \leq k \right\}$$

where each x_p is assigned to one class.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

The mean is computed in this way.

An improved version of K-Means is K-Means++ [7] where it significantly improves over the worst case running complexity by initially choosing the seed points intelligently to guarantee a O(log k)

Kmeans is a very effective color based clustering algorithm for images, it easily computes the similarity between two color to effectively segment them.

2.3 Edge detection

Edge detectors are an integral part of any image segmentation study, the goal is to construct the a solid boundary between separate regions of interest. Edge detector detect the sudden change in the intensity values. Image noise are a big problem they give rise to false edges. Our target using this algorithm is to identify the edges separating the stratum corneum and the stratum spinosum from the rest of the images.

2.3.1 Canny Edge Detection

Canny edge detector [8] developed by John F. Canny is one of the most used edge detection methods realized in 1986 works in the order of the following steps

- Noise Reduction: To suppress noise affecting the edge detection

a gaussian filter is used to smoothen the image.

- Gradient Deduction

  G_x and G_y are the x and the y gradient

$$Edge\_Gradient\ (G) = \sqrt{G_x^2 + G_y^2}$$

$$Angle\ (\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

Non maximum suppression It basically finds thin sharp edges. Hsterisis Thresholding This stage decides which edges to keep and which to discard based on two threshold values, ub and lb i.e. the upper boundary and the lower boundary respectively. Any edge value above the ub is preserved, any edge below the lb is immediately discarded. For edges between the ub and lb, a edge is checked for connectivity with an edge above the ub threshold, if found it is preserved or else discarded.

2.4 Meanshift

Meanshift is a non-parametric clustering technique which does not require the prior information of the number of clusters. The algorithm is works on a very simple principle, picking up a seed point it would try to optimally find the mean of all the points currently in its radius of interest, and shift the centre towards it, keep on repeating unless the current seed is the mean of the density of points within its radius of interest, prune these point, repeat the whole procedure

until all points have found their respective clusters.

2.5 L0 Smoothing

L0 smoothing [23] algorithm is an edge preserving smoothing algorithm based on L0 minimization technique, a very novel approach, it flattens the low amplitude non-zero gradients, preserving the high amplitude ones, thus even for a very high smoothing , the prominent edges would still be preserved.

2.6 Watershed

A segmentation technique primarily would divide the region into part, which exhibit a certain degree of homogeneity , the Watershed [16] is a region based approach , the idea behind the algorithm is that of a catchment basin , it starts flooding the local minima to the point where water from two such regions merge, and thus forms the line separating between the two catchment basins, called as watershed markers. Such models may be used to separate overlapping nuclei in our images, the seed points would serve as the catchment basin and the watershed markers would help separate the nucleus.

2.7 Discussion on existing Segmentation algorithm and challenges

The existing edge detection or segmentation techniques discussed above do not produce satisfactory result in segmenting the skin

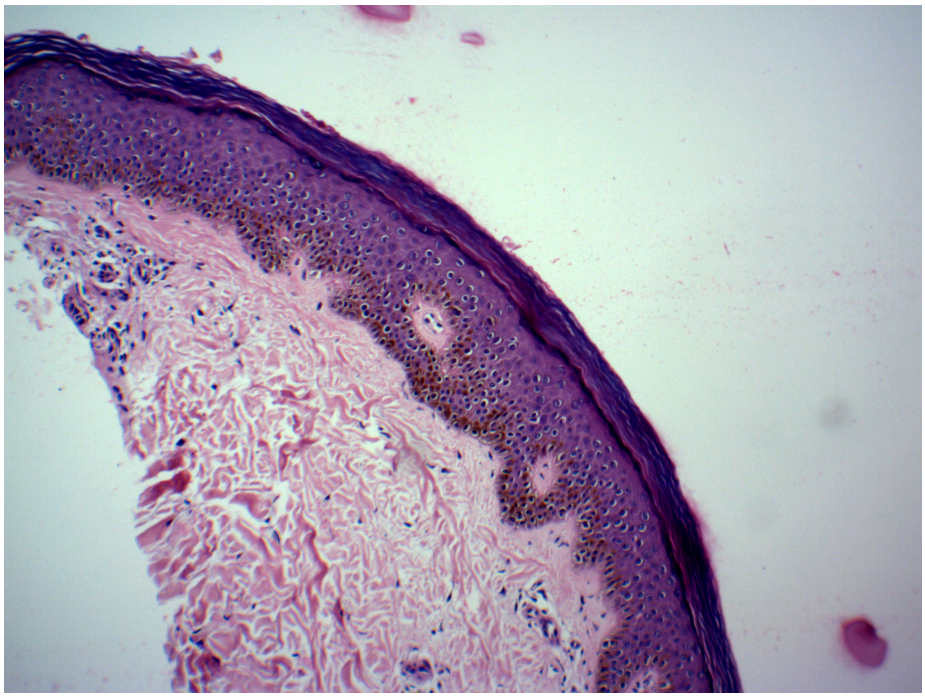
histopathological images of our interest , the primary failure of these approach is the failure to capture the cytological components of the image, mere texture or color information alone is not enough to capture the components together.

Chapter 3

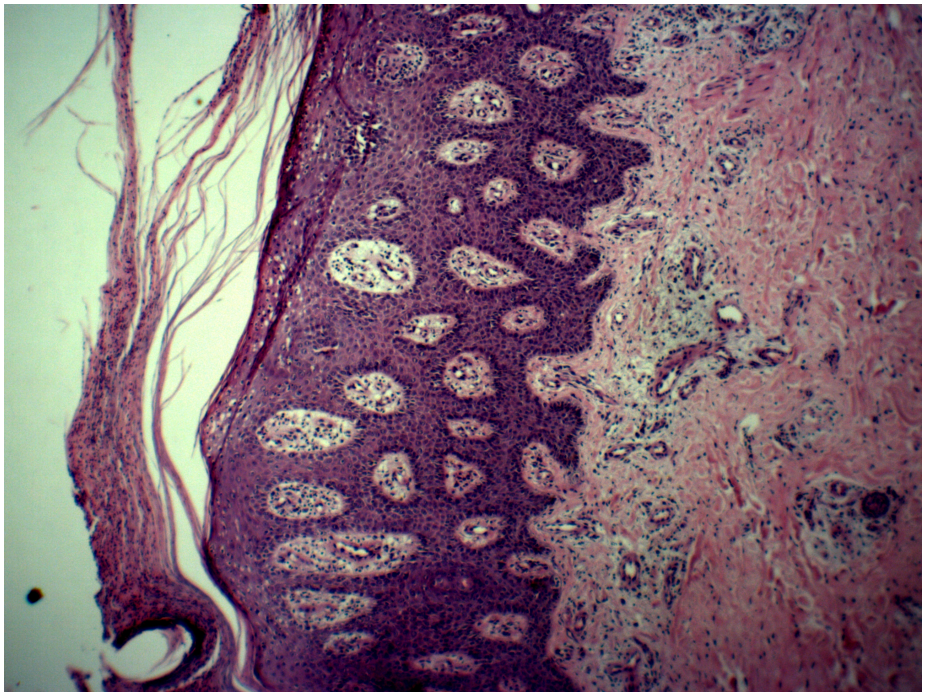
Segmentation of Psoriasis Histopathology images

3.1 Preprocessing

We apply few preprocessing techniques discussed below to segment out the region of interest from our raw images, the procedure is explained stepwise below



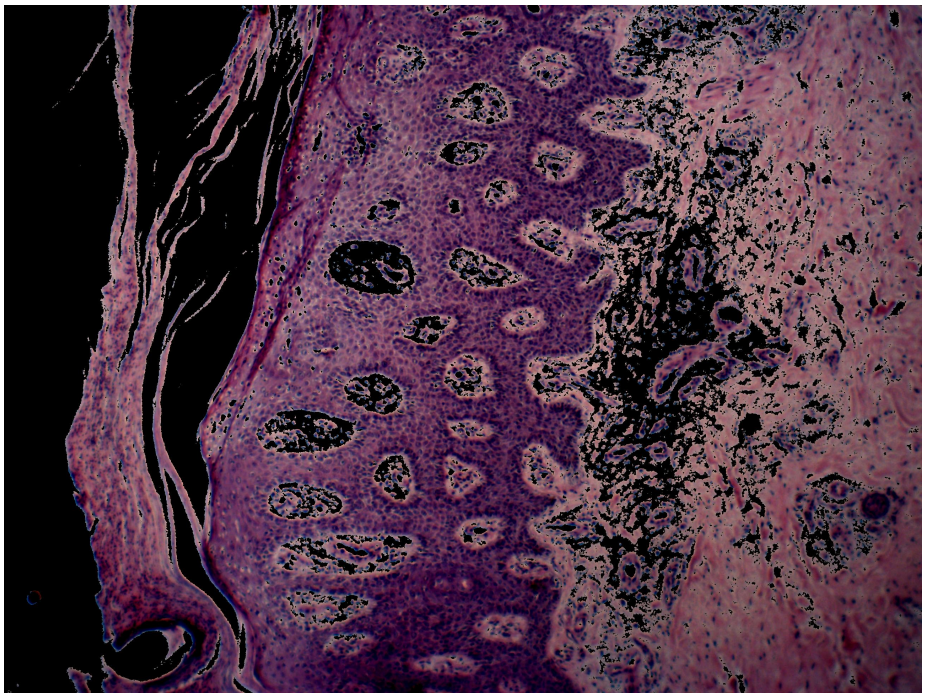
(a) Normal Skin



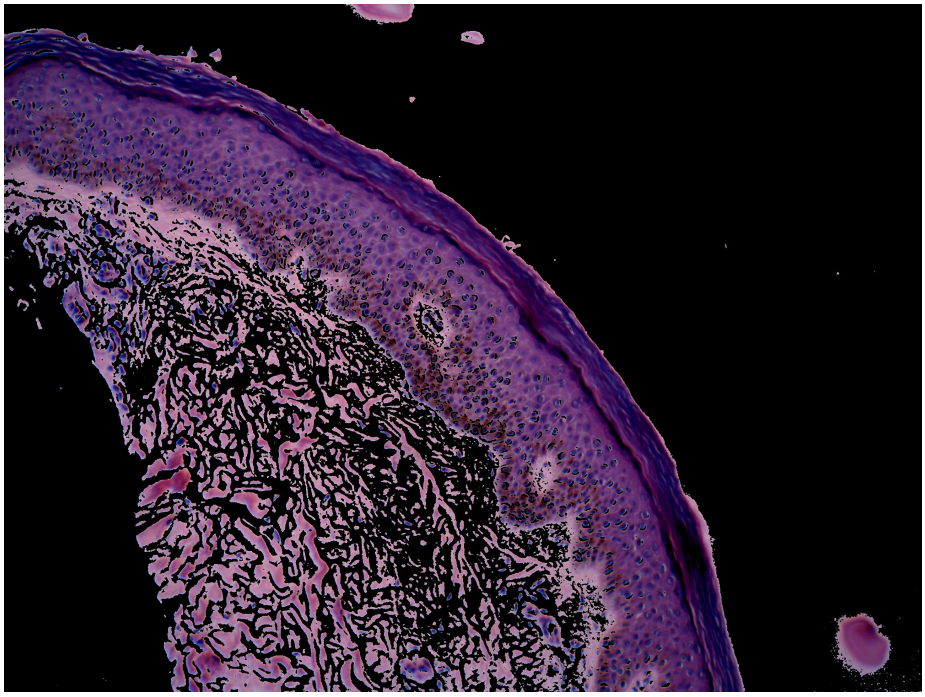
(b) Psoriasis Skin

Figure 3.1: Skin Histopathology Images

3.1.1 Capturing the Background



(a) Normal Skin

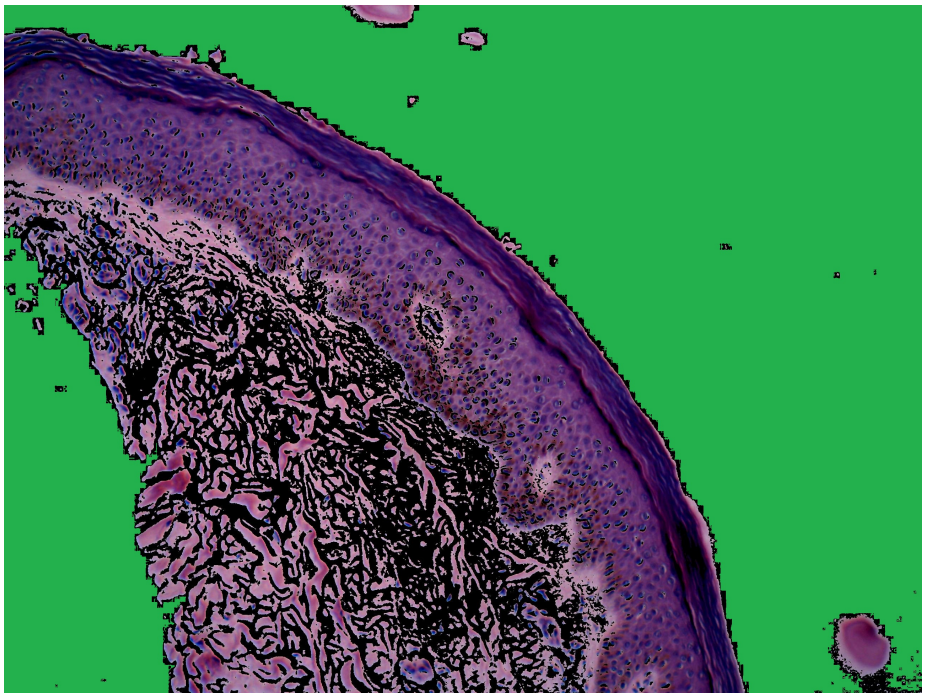


(b) Psoriasis Skin

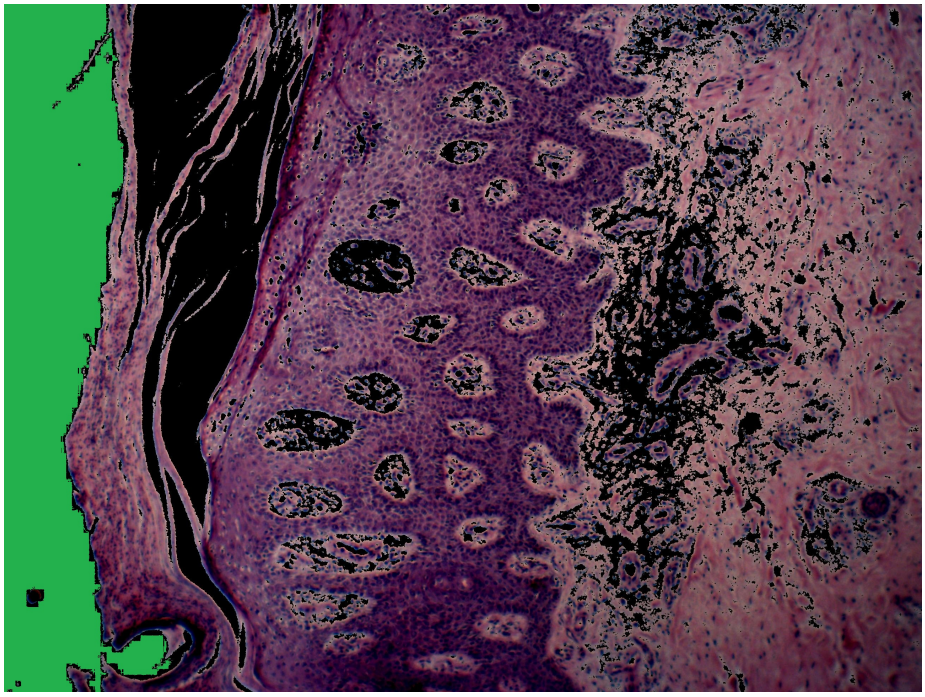
Figure 3.2: 2-means clustering result

To effectively capture the area of interest in our image, next I have used k-means clustering which was earlier described. Using k = 2 in Lab color space [2], it could effectively separate the whites which is a major part of the background from the rest of the colors3.2a 3.2b, now that we have white and the pink-blue color separated we device a method to automatically choose the segment containing the important tissue feature, and discard the ones containing the white portions. The trick was to compare the two images grey-value pixel by pixel and count the ones above a threshold T, experimentally found that a $0.8*255$ grey value threshold would give optimal result. The one having a lower count of white pixels were chosen to be the images of interest.

To clearly separate the background we performed a flood-fill initiated at black pixels on the boundary of the image, which clearly separated the background from the objects in the foreground 3.3a 3.3b. Flood-fill is a recursive algorithm which determines the extent of connectivity, it starts from a seed point, and fills any pixel lying on its edges to the new selected value if it was originally of the same value as the seed point. This step continues until either all pixels have been coloured or no new pixel having the same value as the seed can be reached from any of the currently selected pixels.



(a) Normal Skin



(b) Psoriasis Skin

Figure 3.3: Background Floodfill

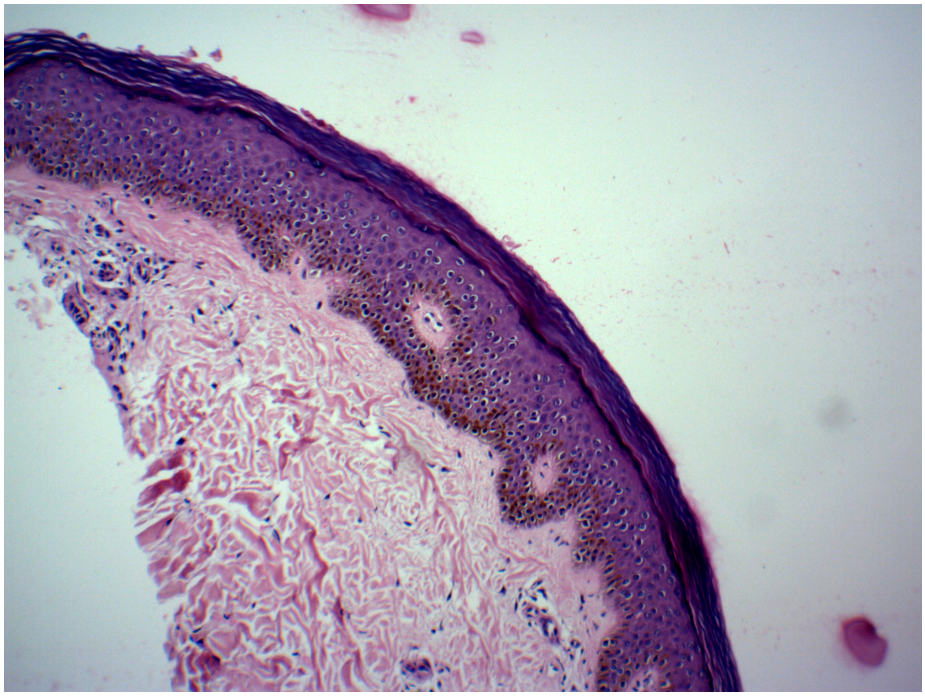
3.1.2 Anti-Vignetting

The original images 3.1a 3.1a, as obtained using bright-field microscopy, which contains random noise, banding noise and a major problem with background illumination intensity. Background illumination intensity problem arises due to the fact that in bright-field microscopic technique, a source of light is below the slide, and observed from the top, the resulting images are obtained due to absorbance of light in the dense areas. The source of light is usually not uniform throughout the span of the slide, and hence results in a stronger brightness in the centre ,gradually fading out as we move towards the edges commonly known as vignetting.

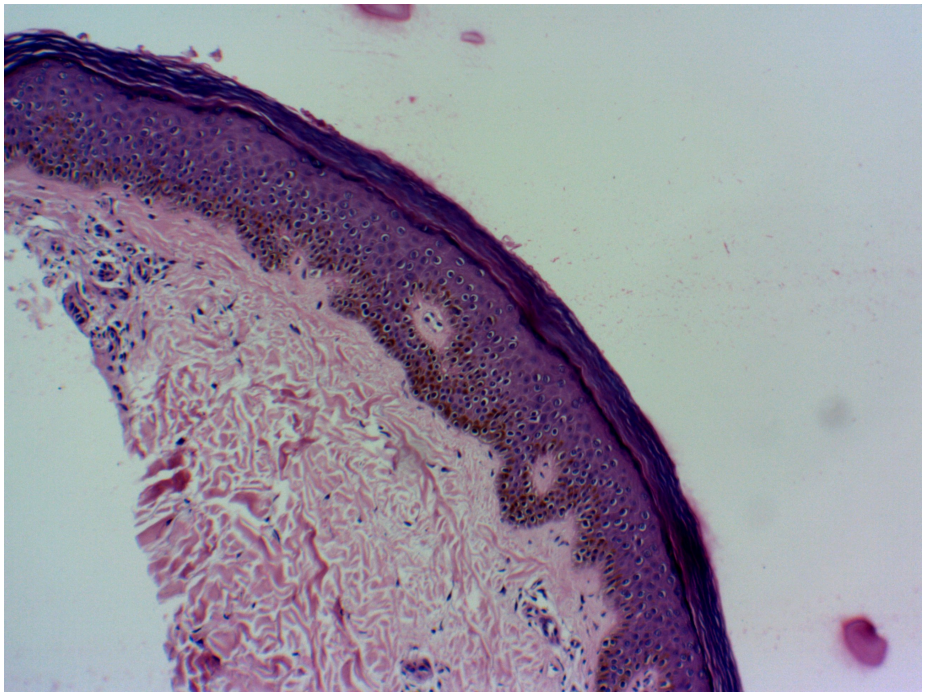
A correction technique [10] state that convolving the image with a gaussian kernel

$$g(x, y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where *sigma* defines the spread of the function. The optimal size of the filter is obviously dependant on the image in consideration, this smoothed version is then applied as an inverted filter on the original image, thus negating the effect of the non uniform background illumination to some extent, but the estimation of the *sigma* is background dependent, so detecting the background is an important pre-requisite for estimating the correction parameter.



(a) Original Image



(b) Corrected Illumination Image

Figure 3.4: Bright-Field Compensation

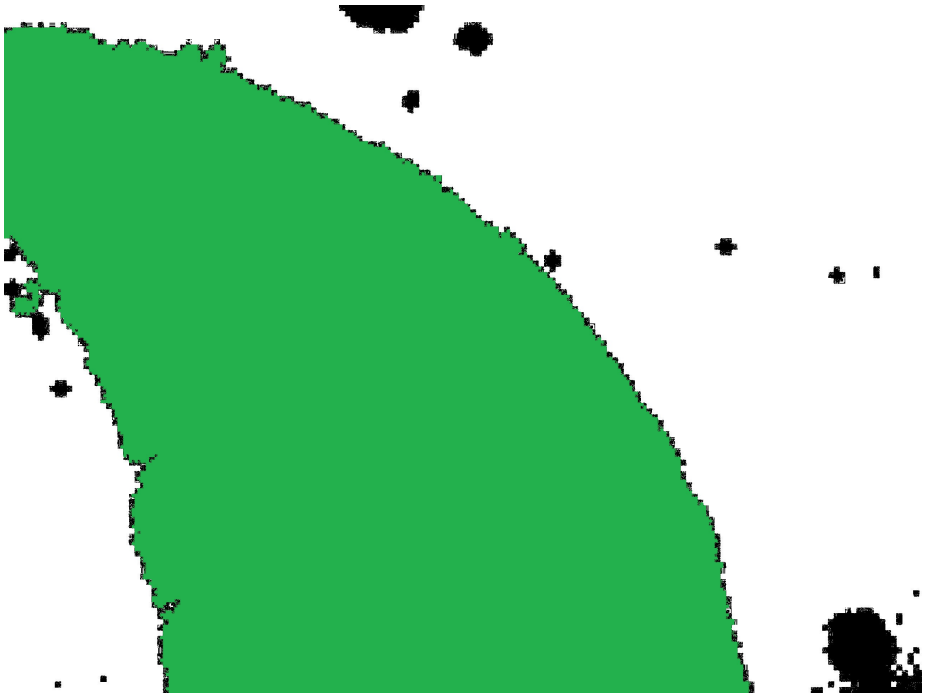
Figure 3.5: Predicted Shading

3.1.3 Shading Model identification, and background illumination correction

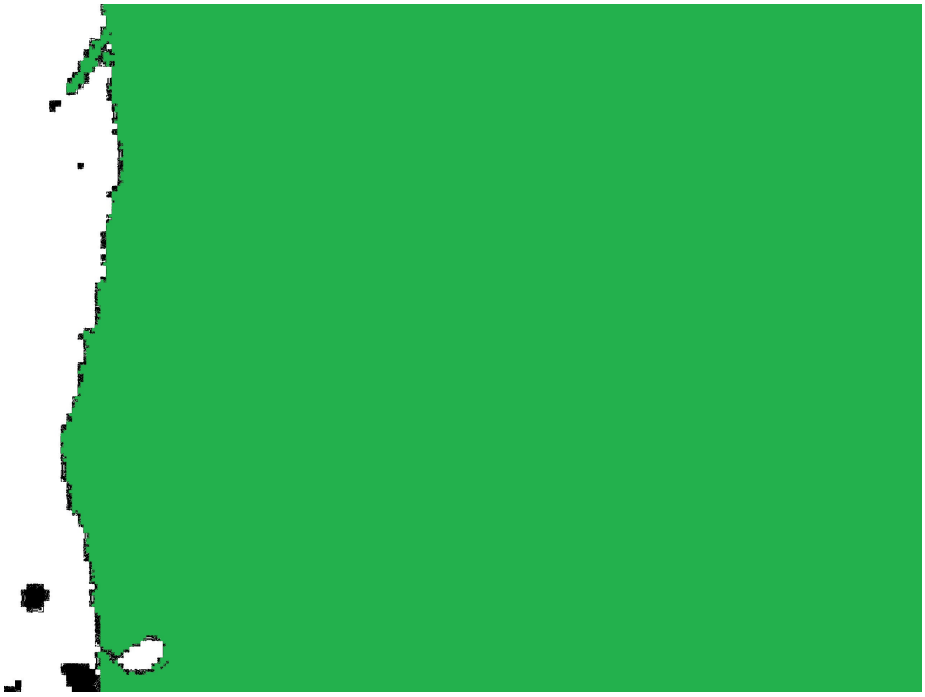
After the background have been identified it is used to perform an estimate of the shading as stated earlier in 3.1.2.

3.1.4 Foreground Extraction

Once the background has been separated the next step involve separating the main tissue component from other fragments of tissue present within the slide. We use a boundary fill [24] algorithm to fill up the individual components, and keep a note of the seed and the number of pixel in the family of each component. Later the component with the largest size is chosen as is determined to be our actual region of interest in foreground. Boundary fill algorithm is similar to that of the flood fill algorithm, the difference lies in their stopping criterion, flood-fill would fill in all components reachable from the seed point, having the value of the seed point, boundary fill on the other hand would reach out from a seed point to all connected pixel, until it reaches pixel containing the boundary value. We obtain pixel value and mark them manually using paints floodfill algorithm.



(a) Normal Skin

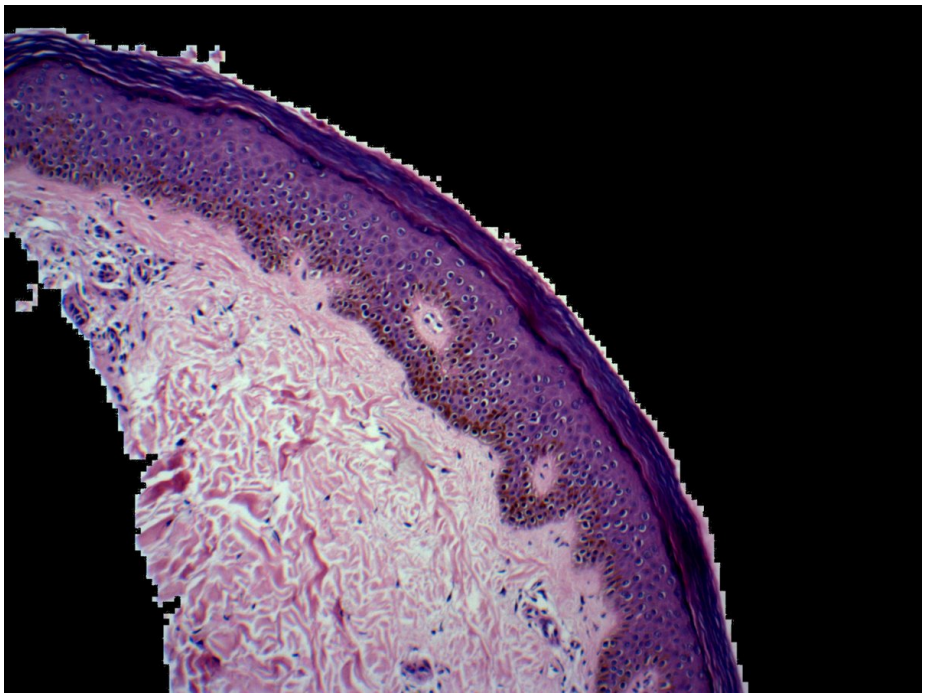


(b) Psoriasis Skin

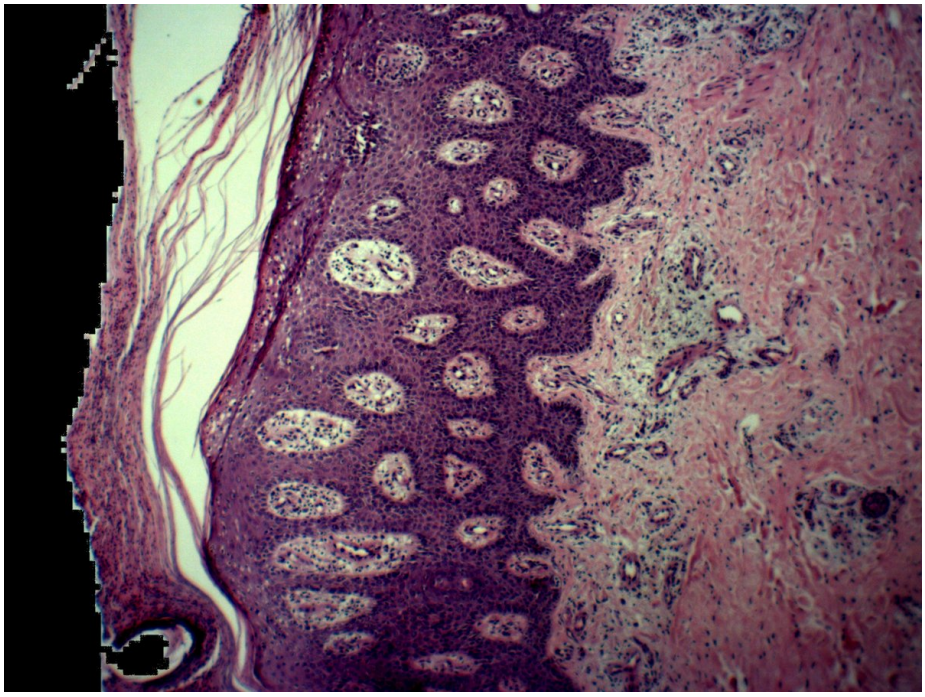
Figure 3.6: Biggest Component Mask

The image mask showing the largest foreground component is shown in 3.6a and 3.6b.

After the above steps are over we are left with the mask, identifying the key component in our image, we simply do a pixel by pixel comparison and choose those pixels from the original image to finally obtain a background removed, image of the tissue. 3.7a 3.7b



(a) Normal Skin



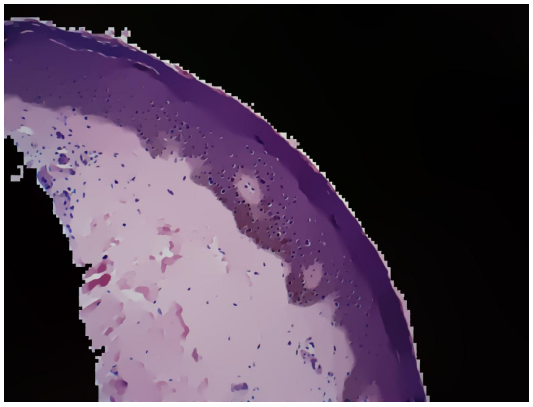
(b) Psoriasis Skin

Figure 3.7: Background removed

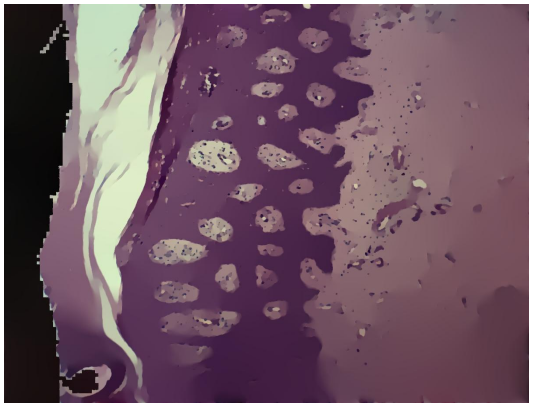
3.2 Segmentation

3.2.1 Smoothing

After the pre-processing step, a background separated and illumination corrected image is obtained. We next proceed with the smoothing, we choose L0 edge preserving smoothing [23] to even out the noises present in the image and yet preserve the edge boundaries, with smoothing parameter 0.05 experimentally found.Jiaya Jia's version of implementation of the algorithm is used to achieve the result.



(a) Normal Skin



(b) Psoriasis Skin

Figure 3.8: L0 smoothed Histopathology Images

3.2.2 Segmentation of Stratum

Once we have the smoothed illumination corrected image of the tissue component , we proceed with the segmentation, two key feature to extract our region of interest, or quantitatively define homogeneity is to understand the discontinuity involved in the image. We perform a mean-shift based color, spatial clustering of the image. Using euclidean distance measure. Over Segmenting the image into homogeneous clusters. Edison is an efficient implementation of the mean shift algorithm we discussed previously, we use a wrapper of this tool in Matlab for our purpose.

(a) Normal stratum corneum-stratum spinosum Skin

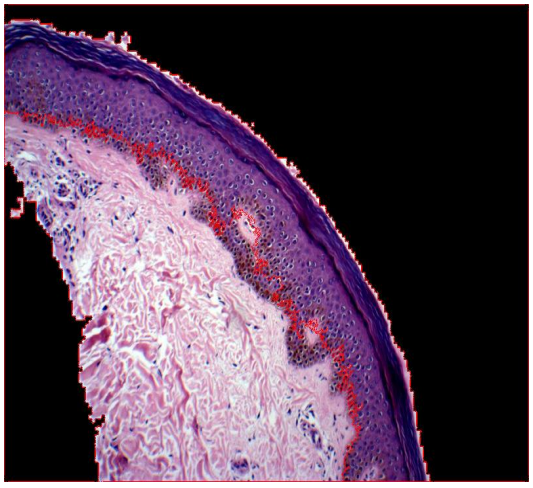
(b) Psoriasis stratum corneum-stratum spinosum Skin

Figure 3.9: Over segmentation

3.2.3 Region Merging

Over segmented regions are merged using a component-size threshold, component size below the min size threshold are merged into surrounding clusters, if there is a tie a color distance measure is obtained and the region assigned to the closest neighbour.

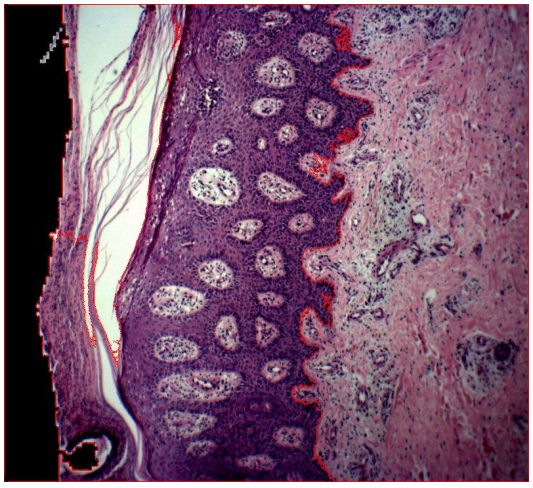
(a) Segmented Image



(b) Segmented Image with edge outlined

Figure 3.10: Normal Skin

(a) Segmented Image



(b) Segmented Image with edge outlined

Figure 3.11: Psoriasis

Chapter 4

Results

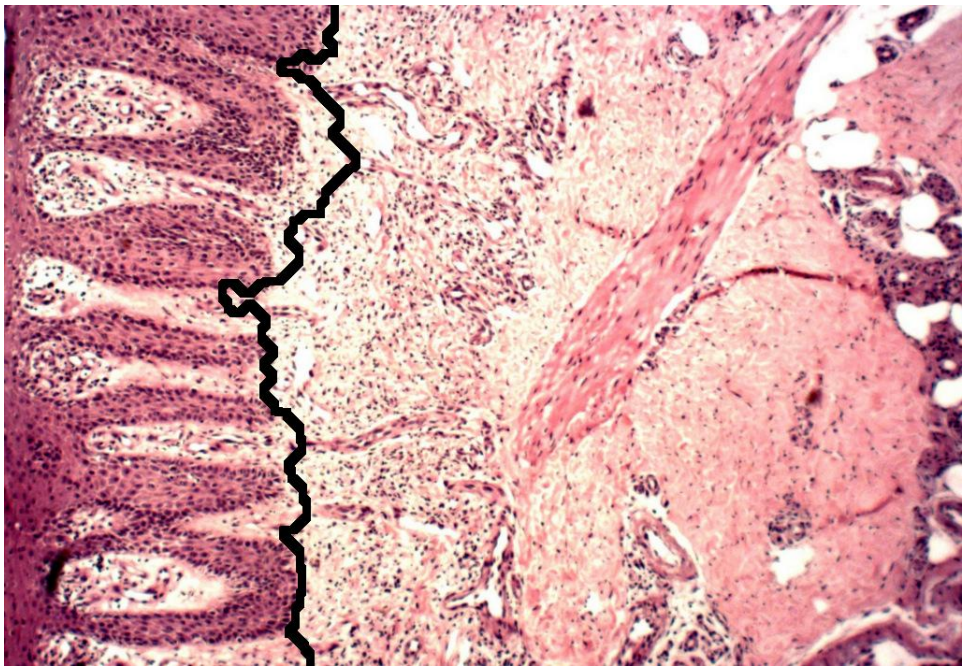


Figure 4.1: Graph Run Length Matrices Segmentation

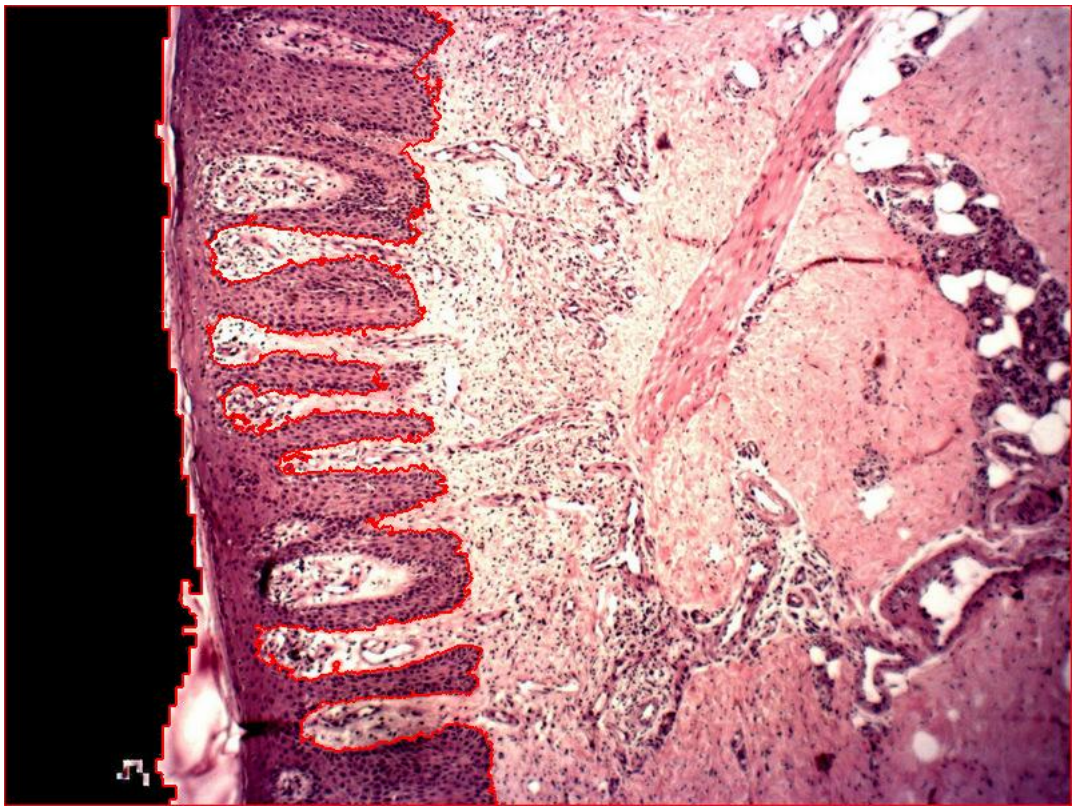


Figure 4.2: Proposed Method Segmentation

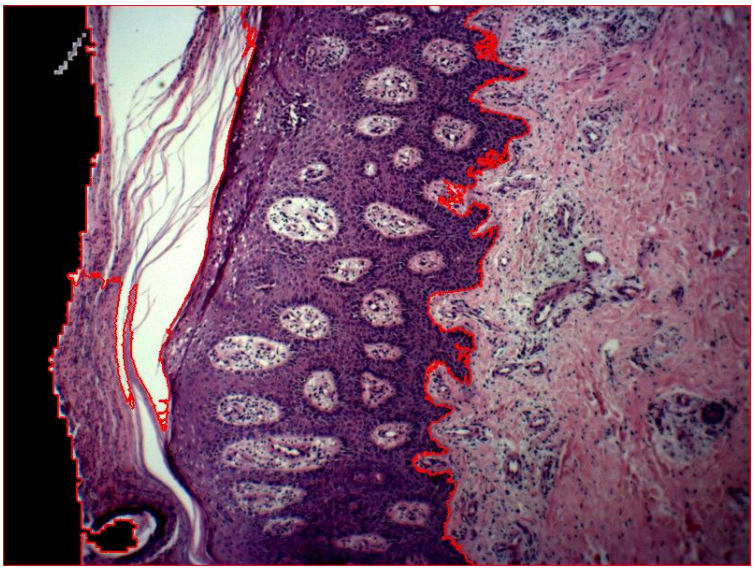


Figure 4.3: Proposed Method Segmentation

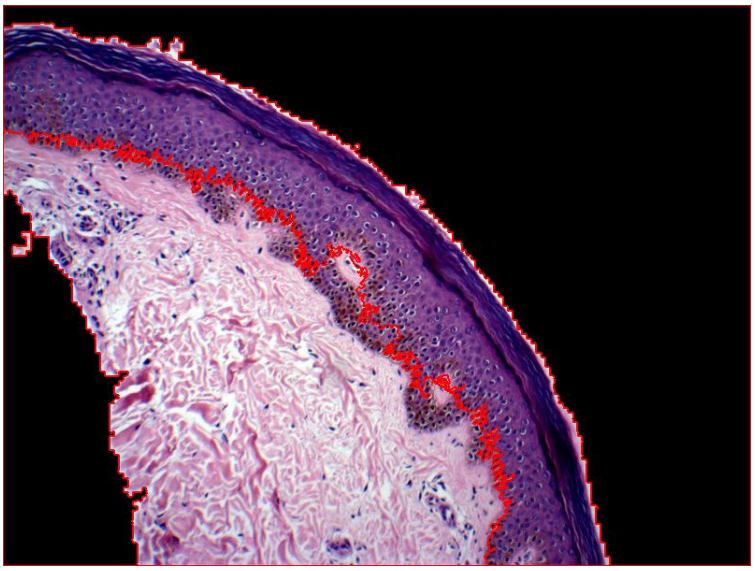


Figure 4.4: Proposed Method Segmentation

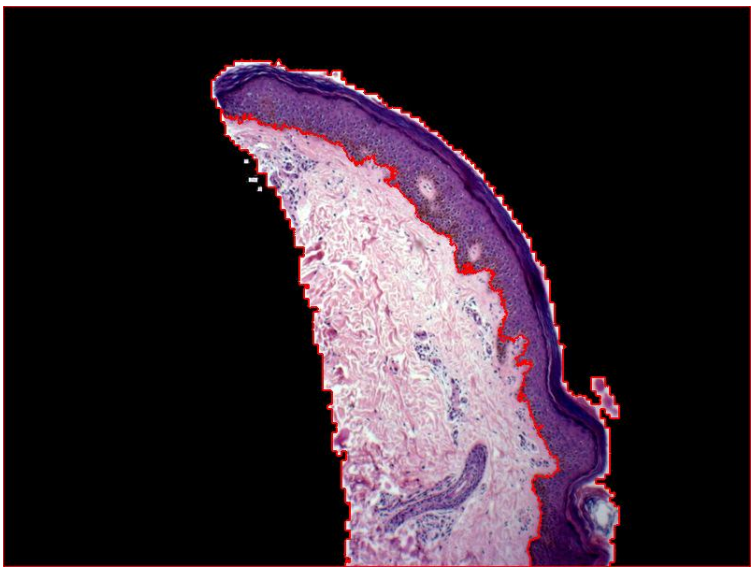


Figure 4.5: Proposed Method Segmentation

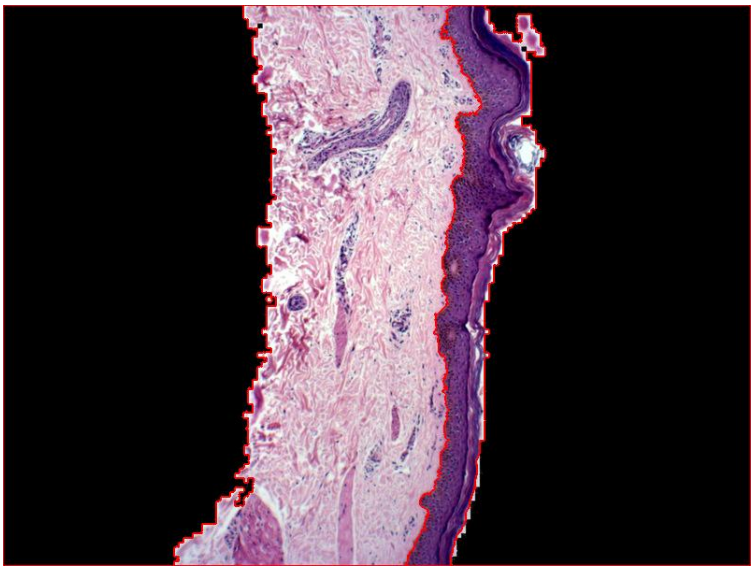


Figure 4.6: Proposed Method Segmentation

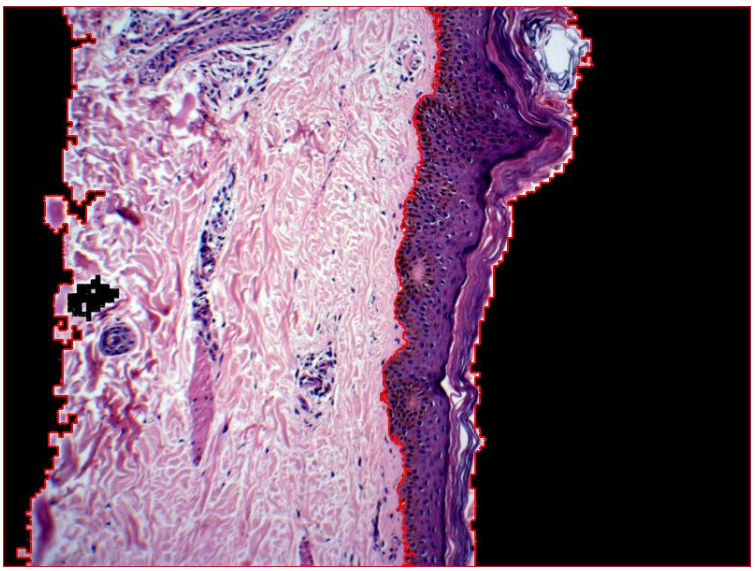


Figure 4.7: Proposed Method Segmentation

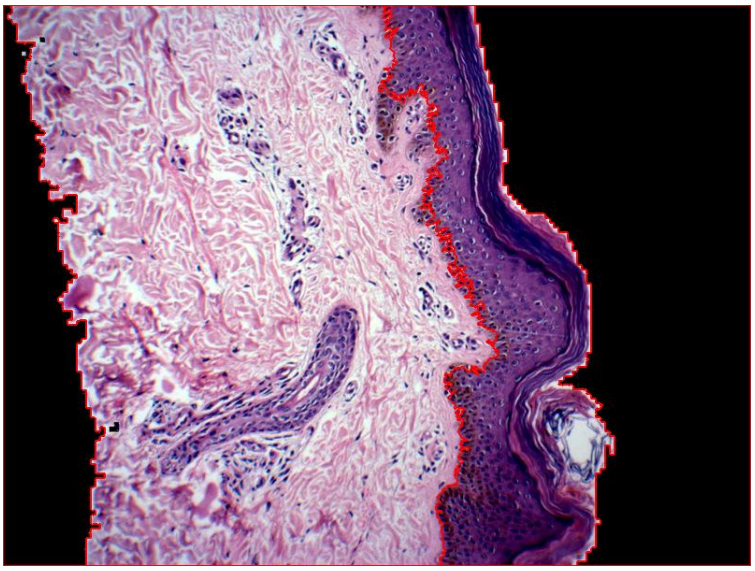
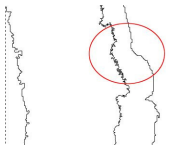
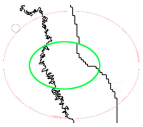


Figure 4.8: Proposed Method Segmentation

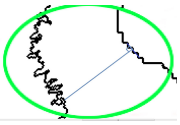


(a) The edge boundary



(b) A subfigure

Figure 4.9: a closer look



4.1 Descriptor Extraction

Once the edges have been determined we need an estimate of the rete peg elongation in order to determine psoriasis, estimating the exact width is a difficult task, because of the non uniform boundary. We propose an approach to estimate the slope of the boundary, and estimate the distance from this point to the edge of the stratum spinosum. For every pixel on the edge , we consider a neighbourhood of two pixel on each side, find the best-fit line and estimate the slope. Draw a perpendicular from the point and mark the intersection with the boundary. We take note of such lengths. Taking note of the longest and the shortest length we find the difference between them. Above a certain threshold it suggests psoriasis. This distribution of width forms our descriptor for detection of psoriasis.

4.2 Conclusion

The current study has discussed about the different histopathology segmentation algorithm in literature , medical importance , and

edge detection techniques , showed an effective way of estimating and removal of vignetting. Used mean-shift to segment the image, and removal of smaller component to remove unnecessary over segmentation. The nuclei segmentation techniques we tried failed to provide satisfactory result, hence discarded from the experiment[5].

4.3 Proposed Future work

4.3.1 Different Classes of psoriasis

There are different categorization of psoriasis , each of them exhibiting different characteristics and is supposed to be linked to the patient's gene. A key correlation can be estimated if these different forms can be identified. Segmenting the stratum corneum from the epidermis layer and to find methods to detect nuclei in the stratum corneum, and classify the images as psoriasis perakeratosis, or munro's microabcess.

4.3.2 Dataset

Dataset is a very important necessity in further growing of analysis in the field, annotated dataset along with patient genetic information might help is establishing the genetic linkage and offer better treatment.

Bibliography

[1] HSL and HSV. https://en.wikipedia.org/wiki/HSL_and_HSV/.

[2] Lab color space. http://ebooklibrary.org/articles/lab_color_space/.

[3] National Psoriasis Foundation. https://www.psoriasis.org/.

[4] National Psoriasis Foundation. http://www.ivyroses.com/HumanBody/Histology/Histology-Stains.php.

[5] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *Biomedical Engineering, IEEE Transactions on*, 57(4):841–852, April 2010.

[6] Dogan Altunbay, Celal Cigir, Cenk Sokmensuer, and Cigdem Gunduz-Demir. Color graphs for automated cancer diagnosis and grading. *Biomedical Engineering, IEEE Transactions on*, 57(3):665–674, 2010.

[7] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual*

ACM-SIAM symposium on Discrete algorithms, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[8] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, Nov 1986.

[9] Vinod Chandran and Siba P Raychaudhuri. Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *Journal of autoimmunity*, 34(3):J314–J321, 2010.

[10] M Brady F J W-M Leong and J OD McGee. Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of Clinical Pathology*, 2003.

[11] Jianping Fan, David KY Yau, Ahmed K Elmagarmid, and Walid G Aref. Automatic image segmentation by integrating color-edge extraction and seeded region growing. *Image Processing, IEEE Transactions on*, 10(10):1454–1466, 2001.

[12] Mary M Galloway. Texture analysis using gray level run lengths. *Computer graphics and image processing*, 4(2):172–179, 1975.

[13] Joseph C Harsanyi and Chein-I Chang. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4):779–785, 1994.

[14] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):pp. 100–108, 1979.

[15] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu. Methods for nuclei detection, segmentation, and classification in digital histopathology: A reviewcurrent status and future potential. *Biomedical Engineering, IEEE Reviews in*, 7:97–114, 2014.

[16] Jos BTM Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundam. Inform.*, 41(1-2):187–228, 2000.

[17] Jinshiuh S Taur, Gwo-Her Lee, Chin-Wang Tao, Chien-Chou Chen, and Ching-Wen Yang. Segmentation of psoriasis vulgaris images using multiresolution-based orthogonal subspace techniques. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(2):390–402, 2006.

[18] Jinshiuh S Taur and Chin-Wang Tao. Texture classification using a fuzzy texture spectrum and neural networks. *Journal of Electronic Imaging*, 7(1):29–35, 1998.

[19] A.B. Tosun and C. Gunduz-Demir. Graph run-length matrices for histopathological image segmentation. *Medical Imaging, IEEE Transactions on*, 30(3):721–732, March 2011.

[20] Akif Burak Tosun, Cenk Sokmensuer, and Cigdem Gunduz-Demir. Unsupervised tissue image segmentation through object-oriented texture. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2516–2519. IEEE, 2010.

[21] Te-Ming Tu, Chin-Hsing Chen, and Chein-I Chang. A posteriori least squares orthogonal subspace projection approach to desired signature extraction and detection. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(1):127–139, 1997.

[22] Hongming Xu, Cheng Lu, and M. Mandal. An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm. *Biomedical and Health Informatics, IEEE Journal of*, 18(5):1729–1741, Sept 2014.

[23] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l 0 gradient minimization. In *ACM Transactions on Graphics (TOG)*, volume 30, page 174. ACM, 2011.

[24] Vladimir M. Yanovsky, Israel A. Wagner, and Alfred M. Bruckstein. A linear-time constant-space algorithm for the boundary fill problem. *The Computer Journal*, 50(4):473–477, 2007.