Pseudo Blob Classification and Merging Techniques for Cellular Segmentation using Bayesian Epistemology

Sachin S Bhat
Dept of E&C
Shri Madhwa Vadiraja Institute of Technology
Udupi, Karnataka

Abstract - Segmentation of cellular images presents a challenging task for computer vision, especially when the cells without regular shape, boundary and intensity clump together to form complex microstructures like metallographic cells. In such cases it is hard to achieve a reasonable segmentation. In view of this, segmentation is performed by combining watershed transform with ridge detection technique. The watershed segmentation has been proved to be a powerful and fast technique for both contour detection and region-based segmentation. However, in an image of metallographical cells or cancer tissues will be having poor quality, disconnected boundaries with more than one seed, so objects in the image will be over-segmented into several parts so that it becomes even hard to identify the segmentation result.

In this, watershed segmentation depends on ridges to perform proper segmentation. The seeds are selected by a simple double-threshold approach, and the ridges are superimposed as the highest waterlines in watershed transform. To solve the over segmentation problem, we classify and merge the pseudo blobs iteratively using Bayesian epistemology which is famously known as Baye's probability theorem. The experimental results demonstrate the segmentation effectiveness of the proposed methodology.

Keywords
Image segmentation, Bayesian probability theorem, Watershed algorithm, Ridge detection, Pseudo Blob Classification and Merging

1. INTRODUCTION

Image segmentation is a process of partitioning a digital image into multiple regions. Each of the pixels in a region is similar with respect to some characteristics, such as intensity, gradient, or texture. Adjacent regions are significantly different with respect to the same characteristics. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Image segmentation is typically used to locate objects and boundaries in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. There exists several basic ways of approaching image segmentation. One is boundary based and detects the local changes in the boundaries or detects the edges. If there is contrast in intensity between the object and background, the detected edges can be used to trace the object boundaries. Another technique is region based and searches for pixel and region similarities. The image is partitioned into connected regions by grouping neighboring pixels of similar intensity levels. Adjacent regions are then merged under some criterion like sharpness of region boundaries. We shall see that the watershed transformation belongs to the latter case. With that, we use the threshold based methods which are very effective for images containing solid objects in a contrasting background.

Since there is a contrast in intensity between the object and the background, the detected edges can be used to trace the object boundaries. Here edges identified are often disconnected. But we need closed region boundaries to segment an object. The clustering methods choose cluster centres and assign each pixel in the image to the cluster that minimizes the variance between the pixel and the cluster centre. Region growing methods work under seed growing principle. Regions are expanded by comparing all unallocated neighbouring pixels to the regions.

Objects have severely irregular in shape and have inhomogeneous intensity in metallographic images and some of the medical images like mammogram or cancer tissues where it is necessary to track the interference and shape of the tissues. Then it is hard to achieve reasonable segmentation using conventional approaches. Hence we formulate the segmentation as watershed region growing with ridge information. To tackle the over-segmentation problem, blobs are classified and merged iteratively with the utilisation of Bayes classification rule.

2. PROBLEM DEFINATION

2.1 Importance of metallographic image analysis

Mechanical properties of a material are dependent on the nature and distribution of the phases and size of grains present in it. Metallography deals with the study of the structure of metals that provides information regarding the compositions and properties of the metal. Metallography is the science of preparing metal surface for analysis by grinding, polishing, etching to revel its microstructural constituents. Metallographic image analysis is essential for objectively assessing the material properties with its microstructure. Microstructure quantification is performed by examining the metallographic specimens under suitable electronic/optical instruments and the image obtained. The technique is valuable in the research and production of all metals and alloys and non-metallic or composite materials.

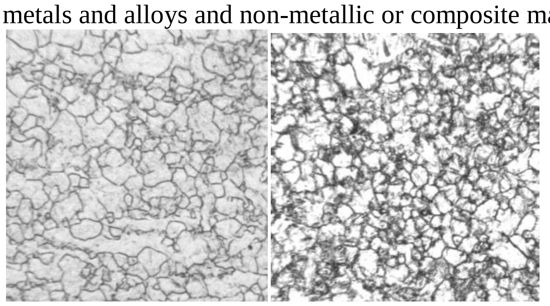

Fig. 1 Sample images

In the above figure, group of pixels making up a microstructure is called as blob. Blobs are the connected pixels with similar gray intensity and black boundary. Size, shape, curvature, orientation, and distribution of the blob are important in quality assessment. In order to perform the image analysis, first step is to segment each of the blobs. However, conventional segmentation techniques are unsuitable to achieve this goal because of some reasons.

2.2 Why conventional techniques are unsuitable?

Firstly, each microstructure in the image has irregular shape. If it is the case of blood cells, it can be modeled as circle, and can be detected using Hough transform. But here, size of each microstructure is different from one another. Some of the cells may be very larger compared to another one. Secondly, pixel intensities inside the blob are inhomogeneous due to texture properties of the tissues. With that, image quality may be further degraded by uneven illumination and noise disturbance. So by using the threshold-approaches, it is hard to achieve a satisfactory partition. A lot of microstructures are connected without a regular boundary. So, boundary is not useful to separate the blobs. Since there is a sharp contrast in the intensities at the region boundaries, edge detection cannot be used alone to perform the segmentation.

Mathematical morphology is a nonlinear branch of the signal processing field and concerns the application of set theory concepts to image analysis. Morphology refers to the study of shapes and structures from a general scientific perspective. Morphological filters or operators are nonlinear transformations, which modify geometric features of images. The watershed transform is the traditional segmentation technique used in mathematical morphology. Beucher and Lantuejoel were the first to apply the concept of watershed and divide lines to segmentation problem. They used it to segment images of bubbles. The intuitive idea underlying this method comes from geography or landscape: imagine the landscape being immersed in a lake, with holes pierced in local minima. Basins (also called `catchment basins') will fill up with water starting at these local minima, and, at points where water coming from different basins would meet, dams are built. When the water level has reached the highest peak in the landscape, the process is stopped. As a result, the landscape is partitioned into regions or basins separated by dams, called watershed lines or simply watersheds.

We will define a set of markers to mark the regions that need segmentation and we extend the marker-controlled watershed transformation. The controlled markers consist of seeds and markers. These seeds are selected using double threshold approach and boundaries using ridge detection. There is no need that seed should correspond to one closed boundary. Over segmentation can be overcome by iterative classification and merging.

3. ALGORITHM IMPLEMENTATION

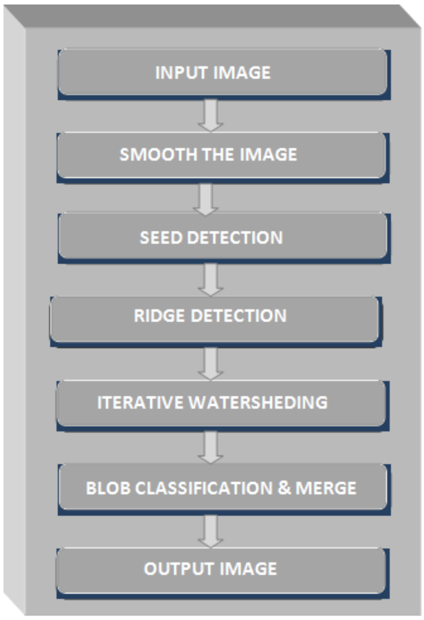

Figure: 2.1 Algorithms for segmentation

3.1 Double threshold seed detection

In the image, seeds work as the lowest basins and mark each of the initial objects to be segmented. Seeds are used for the region growing. Main body of a microstructure is intensity which is brighter than the dark boundary in nature. As the pixels in the blobs are much brighter than in the boundaries, seeds can be identified automatically by double-threshold approach. The morphological extended-maxima transform is used to find the regional joined components. Holes whose area is less than the area threshold in the blob will be eliminated.

Double threshold in the sense, we consider two different threshold parameters such as pixel intensity threshold(T_G) and blob area threshold(T_A). We can find regional connected components using morphological extended maxima transform with intensity threshold. From this, we will get a binary image. Next blob less than blob area threshold are removed. Here, we consider $T_G = 0.1$ for grey level[0,1] and $T_A = 10$ pixels. As a result, a binary image after seed detection will be got which is shown in Fig. 4.1.

3.2 Ridge detection

Different definitions of ridges have been studied by mathematicians and by image processing researchers over the last decades. In one of the oldest studies of this kind, Saint-Venant identified ridges as loci of minimum gradient magnitude along level curve of a relief. Saint-Venant's condition was later reformulated by Haralick for applications in vision. A ridge is detected at loci of extrema of the image function in the direction along which the second order directional derivative has the greatest magnitude. In topography, ridge is defined as the long,

ISSN: 2278-0181

narrow region of the ground which is slightly above the ground.

Ridge is superimposed as the highest waterline in the image at the first time watershed transform to assist the segmentation and play an essential role in pseudo-blobs classification. As the boundaries are relatively evident and robust with constant intensity (dark) and width among the cluttered crystal structures in metallographic images or irregular cells in cancer tissues it is easy and efficient to utilize the ridge detection instead of edge detection for the irregular shapes and discontinuous boundaries as stated previously.

In favour of segmentation, we can define the ridge as the local extreme point in the direction of the largest surface curvature and can be detected by computing the Eigen values of the Hessian matrix, a matrix of second order derivatives of intensity image L(x, y).

Eigen values (the large eigenvalue, λ+, and the small eigenvalue, λ−, where λ+ > λ−) of the Hessian matrix of the intensity image $L(x, y)$ are calculated as

$$\lambda+ = \frac{(\mathrm{L_{xx}} + \mathrm{L_{yy}} + \alpha)}{2}$$

$$\lambda- = \frac{(\mathrm{L_{xx}} + \mathrm{L_{yy}} - \alpha)}{2}$$

where L_{xx}, L_{yy} are the second derivatives of the intensity image in x and y directions,

and $\alpha = \sqrt{[(\mathrm{L_{xx}} - \mathrm{L_{yy}})^2 + 4L_{xy}^2]}$

Then, the local maximum of the large eigenvalue λ_{max} is calculated as :

$\lambda max = max\,[\lambda+(s)]$

Eigen vectors can be calculated using eigen matrix as:

$$\Theta_+ = \tan^{-1}\frac{(\mathrm{L_{yy}} - \mathrm{L_{xx}} + \alpha)}{2\mathrm{L_{xy}}}$$

$$\Theta_- = \tan^{-1}\frac{2\mathrm{L_{xy}}}{(\mathrm{L_{yy}} - \mathrm{L_{xx}} + \alpha)}$$

Where $\boldsymbol{\theta}_+$ and $\boldsymbol{\theta}_-$ corresponds to eigenvectors.

It is noted that the fixed-scale ridge definition can be very sensitive to the different object width. Hence Scale space ridges are used which allows the scale parameter to be automatically tuned to the width of ridge structure in the image domain. Although the ridge descriptor works well at indicating boundaries with different widths and orientations, the detector still yields false/missing detections. In this work, we choose to minimize the false detection error (i.e. Type II error). Because watershed transform always have an over-segmentation result, an under-detection result for ridge detection is preferred.

3.3 Iterative scheme

Watershed algorithm is an image processing algorithm that achieves segmentation by splitting an image into areas, on the basis of image topology.

First, ridges are superimposed as the highest waterline and seeds are marked as the lowest water basin on the original image. Then watershed transform is performed. This is not enough as there still exists over segmentation problem.

3.4 Bayesian epistemological probability theorem for blob classification and merge

In order to improve the accuracy, iterative watershed is proposed with blob classification and blob merging using Bayes' epistemology theorem.

Bayes' epistemology theorem shows the relation between two conditional probabilities which are the reverse of each other. Bayes' theorem relates the conditional and marginal probabilities of events *A* and *B*, provided that the probability of *B* does not equal zero:

$$P(A\,|B) = \frac{P(B\,|A)\,P(A)}{P(B)}$$

Using Bayes rule, first step is to decide which blob is over-segmented one, named as pseudo-blob. Next step is to reallocate the pixels of a pseudo-blob into other real blobs.

3.4.1 Blob classification

Let V_s be a feature vector extracted from image pixel at location *s* = (*x*, *y*) and iteration time *t*. For a blob, the posterior probability of V_s from the pseudo-blob 'p' or real blob 'r' is given by

$$P(\mathrm{p}\,|\,V_s\,,t) = \frac{[\,P(\,V_s\;|\mathrm{p}\,,t)\,P(\mathrm{p}|\;t)]}{P(V_s\,,t)} \quad \ldots (1)$$

And

$$P(r|\,V_s\,,t) = \frac{[\,P(\,V_s\;|\,r\,,t)P(r\;|\;t)]}{P(V_s\,,t)} \quad \ldots (2)$$

Using the Bayes probability theorem, the pixel is classified as inside a pseudo-blob if the feature vector satisfies

$$P(\mathrm{p}\,|V_s,\;t) > \;P(r\,|V_s,\;t) \quad \ldots (3)$$

The feature vector V_s associated the pixel *s* = (*x*, *y*) are either from a real-blob or a pseudo-blob, it follows the

$$P(V_s\,,t) = \;P(\,V_s\;\;|\,\mathrm{p},t)P(\mathrm{p},t) + \;P(\,V_s\;\;|\,r,t)P(r,t) \quad \ldots (4)$$

Taking all the pixels in a blob into account, and substituting (1) and (4) to (3), it becomes

$$2\;P(\mathrm{p}\,|t) > \frac{\sum P(V_s|t)}{\sum P(V_s\;|\mathrm{p},t)} \quad \ldots (5)$$

The prior probability of pseudo-blob at iteration t is updated recursively using

$$P(\mathrm{p}\,|\,t) = \;\alpha\;P(\mathrm{p}\,|\,t-1) \quad \ldots (6)$$

$$\alpha = \;\left[\frac{\alpha_0}{P(\mathrm{p}|0)}\right]^{1/N} \quad \ldots (7)$$

Where, *N* - maximum number of iterations

At the final iteration,

$$P(p \mid N) = \alpha_0 \qquad \ldots(8)$$

For simplicity purpose, feature vector v_s is chosen as the binary ridge descriptor,

i.e $P(V_s|t) = 1$ for s = ridge and
$P(V_s|t) = 0$ for s ≠ ridge.

3.4.2 Pseudo blob merge and real blob classification

There are basically two ideas to merge the pixels inside a pseudo-blob. Either pixel-by-pixel or all pixels as a whole. Here we will first label the pixels inside a pseudo-blob as blank. Then Re-watershed the images to let the surrounding blobs encroach the pseudo-blob pixels.
Then count the maximum occurrence label of a pseudo blob pixels and assign the label to all pseudo blobs.

Using the Bayes decision rule, the pixel will be classified as inside a real-blob if the *Vs* satisfy

$$P(B_2 \mid V_s\, t) > P(\mathrm{B1} \mid V_s\, t) \qquad \ldots(9)$$

Substituting

$$2\, P(B_2\, |t) > \frac{\sum P(V_s|t)}{\sum P(V_s\, |B_2,t)} \qquad \ldots(10)$$

The prior probability of real-blob at iteration t is updated recursively using

$$P(B_2 \mid t) = \beta\, P(\mathrm{B1} \mid t - 1)$$

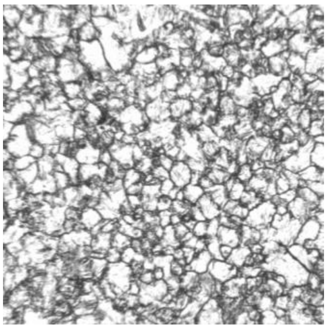

Fig 4.1 Input image

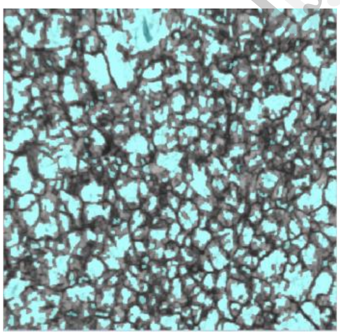

Fig 4.2 Seed Detection

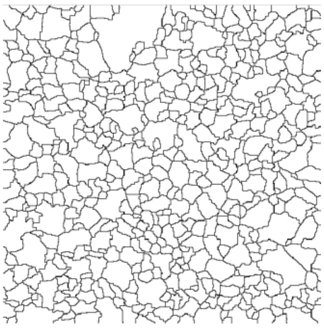

Fig 4.3 Over segmentation after Watershed

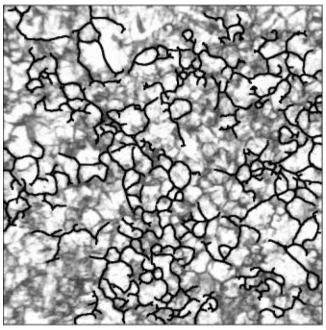

Fig 4.4 Ridge detection

$$\beta = \left[\frac{\beta_0}{P(r|0)}\right]^{1/N}$$

4. RESULTS

Segmentation of the images are implemented using MATLAB VERSION 7.1 and following results are obtained. Result images are shown below. Few parameters are defined for the ease of purpose e.g T_G = 0.1, T_A = 10 pixels, number of iterations or N = 3- 5, P(p|0) = 0.1, α_0 = 0.5, β = 0.68.

Resulted images of seed detection, first level of watershed resulting over segmentation, detecting and superimposing ridges, iterative watershed segmentation, marking pseudo blobs and merging with real blobs using evidential probability etc. are shown below. For the morphological analysis of each of the cells in the image, each blob is modeled as ellipse with parameters such as centroid, major and minor axis, orientation. With that area and average diameter of each blob is calculated.

5. CONCLUSION AND FUTURE WORK

The aim of this paper is to formulate a segmentation method for the complex images without clear regular boundary and homogenous intensity using iterative watershed region growing with the information of the ridges by adding evidential blob classification and merge rules. The proposed method is not only efficient for metallographic images, but also in other areas, such as medical applications. It was fairly simple and the results for the image segmentation are impressive. As can be seen by the results ridge detection provides more accurate information about cell structure compared to other edge detection techniques. Similarly, evidential probability

ISSN: 2278-0181

Figure .4.5 Pseudo blobs for 1st round of watershed

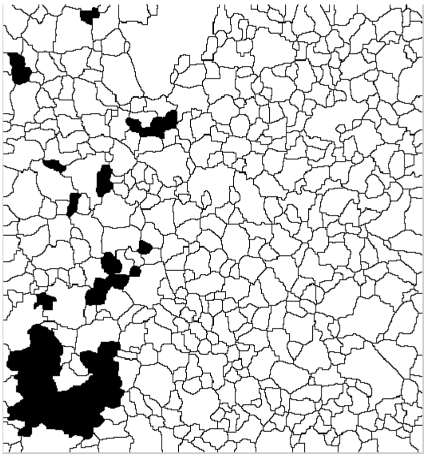

Figure 4.6 Mark pseudo blobs

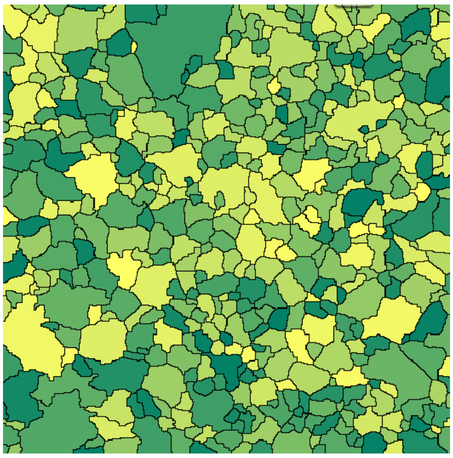

Fig 4. 7 1st round of watershed to merge with real blobs

Figure .4.8 Pseudo blobs for Nth round of watershed

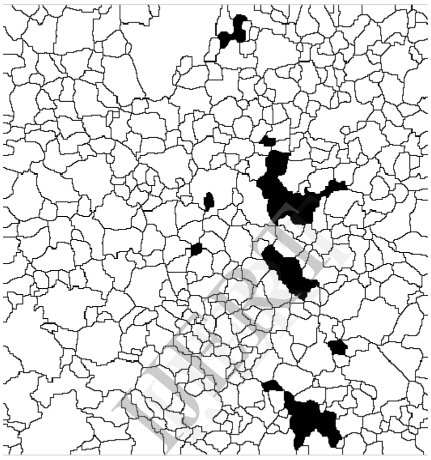

Figure 4.9 Mark pseudo blob

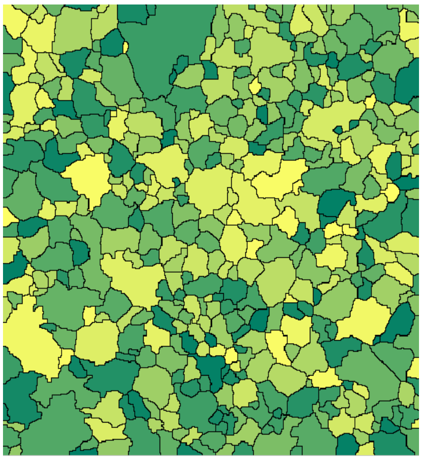

Fig 4. 10 Nth round of watershed to merge with real blobs

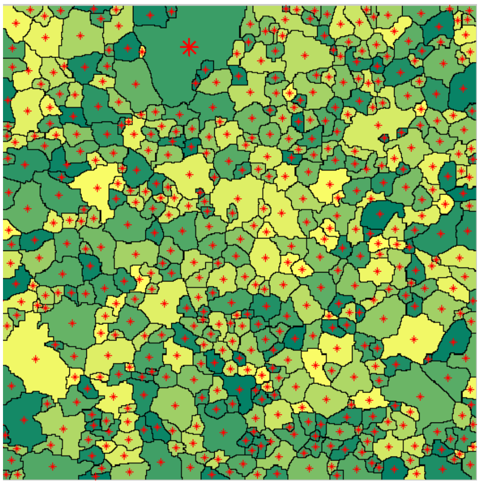

Fig 4.11 Detection of blob centres

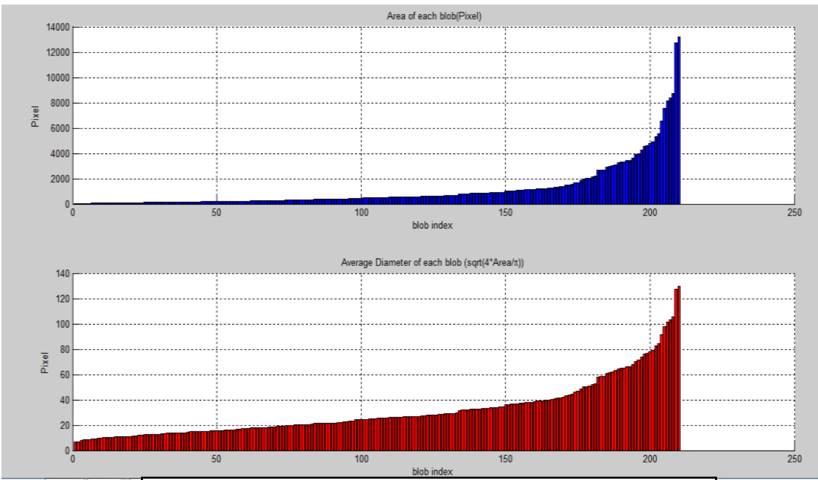

Fig 4.12 Area and average diameter of each blob

[1]. Jiangbo Liu and Jianxun Chen, " An Improved Iterative Watershed According to Ridge Detection for Segmentation", American Journal of Science and Engineering, Vol. 1, No. 1, 2011

[2]. A.Sri Nagesh, Dr.G.P.S.Varma, Dr A Govardhan, "An Improved Morphological Transformation Techniques for Segmentation of Microarray Images", *IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Application, CASCT, 2011.*

[3]. Chen Wei-bin, Zhang Xin, A New Watershed Algorithm for Cellular Image Segmentation Based on Mathematical Morphology, IEEE International Conference on Machine Vision and Human-machine Interface. 2011

[4]. DR.S.V.KASMIR RAJA, A.SHAIK ABDUL KHADIR, DR.S.S.RIAZ AHAMED, "Moving toward region-based image segmentation techniques: A study", Journal of Theoretical and Applied Information Technology, 2009

[5]. K. Kraus, M. Uiberacker, O. Martikainen, and R. Reda, Hot-Spot Blob Merging for Real-Time Image Segmentation, World Academy of Science, Engineering and Technology 46, 2008

[6]. K. Parvati, B. S. P. Rao, and M. M. Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation", *Discrete Dynamics in Nature and Society*, 2009, doi:10.1155/2008/384346.

[7]. Malik Sikandar Hayat Khiyal, Aihab Khan, and Amna Bibi, "Modified Watershed Algorithm for Segmentation of 2D Images", Issues in Informing Science and Information Technology, Volume 6, 2009

[8]. *Nancy M. Salem, Sameh A. Salem, and Asoke K. Nandi, "*Segmentation of retinal blood vesselsbased on analysis of the hessian matrix and clustering algorithm", 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007,

[11]. Tony Lindeberg," Edge detection and ridge detection", *Int. J. of Computer Vision, vol 30, number 2, 1998.*