Biomedical Image Mining in Detection of Skin Cancer Disease

Sukhdeep singh*, Dipti bansal**

Abstract - *We explore this topic in biomedical that have been used in the several hospitals and dermatology clinics to detect the skin cancer or so called MELANOMA, brain tumors and kidney biopsies. In this paper, we introduce the state-of-art to study skin lesion classification characteristics and methods defining them. Further, to describe the features of digital image processing methods such as segmentation, color processing, texture processing, border detection etc. The paper reports the survey of early detection methods of skin lesions of skin cancer disease and compares the performance of various skin lesion classification problems.*

INDEX
Decomposition, Lesion segmentation ,soft thresholding, FCM Algorithm.

I. INTRODUCTION

In the progress of digital image processing, images have become more important. Out of all, skin detection plays an important role in the range of variety of digital image processing applications such as face tracking, gesture analysis and content based image of retrieval systems to various human-computer domains. In such these applications, the search part for objects of interest such as the body part or hands, can be reduced through the detection of skin regions. The simplest and most commonly used method for human skin detection methods is to define the boundary for different color components. Various ranges of threshold values for the single and multiple ranges for each color component are defined and the image pixel value those fall within the predefined ranges are selected as the skin pixels. In this method, for any given color space that may be small or large part in the space.

The focus of the biomedical scientific community for computer supported skin lesion characterization and inspection increased during previous years. Skin cancer is one of the most persistent types of cancer and one of the most malignant tumors. Its proportion has significantly increased than that of all other cancers with annual rates in order of 3-7% in fair skinned population in recent decades.[1] Between 2 and 3 million non-melanoma skin cancers and 1,32,000 melanoma skin cancers are reporting annually. One in every three cancers diagnosed is a skin cancer, and according to Skin cancer foundation survey, one in every five Americans are suffering from skin cancer problem [2].

By using the properties of the images of skin cancer we make three steps segmentation method. The steps of our method are as following:-

1.Pre processing:- The color images are the first transformed into an intensity image in such a way that the intensity of pixel shows the color difference of that particular pixel with the color difference of the background of image. The color of the background is to be taken as the median color of the pixels in the small part in the four corners of the image.

2.Initial segmentation:- In the obtained image, a threshold value is determined from the average intensity of high gradient pixels in that image. This threshold value is used to find the approximate image boundaries.

3.Region refinement:- in this part, region boundary is refined using edge information of the image. This involves the initializing approximate boundary to fit on the edges in its neighborhood. For each chrominance space, the detection efficiency of defect depends upon the capacity of the skin image distribution and most important the estimation between the skin and non skin distributions[3].

(A)

(B)

Fig 1: Examples of illumination variation in skin lesion images. In image (a), the illumination variation

ISSN: 2278 – 1323

changes horizontally, while in image (b), it changes vertically.[2]

II. SEGMENTATION AND FCM METHOD

2.1 Wavelet Transform (WT)

The wavelet transform always offering great design flexibility while trying to replace standard image processing techniques, wavelet transforms provides an efficient representation of the image by finely tuned to its intrinsic properties. By combining such representations with simple processing techniques in the transform domain, multi-resolution analysis can accomplish remarkable performance and efficiency for many image processing problems. Discrete non redundant wavelet transform plays a major role in image analysis techniques. But after the decomposition stage it introduces some artifacts. However one best example of redundant representation is Dyadic wavelet transform.Haar, Debauchees, sym, coif, bior etc. are various types of wavelets. Out of them Debauchees is very popular and used by many scholars in their research.

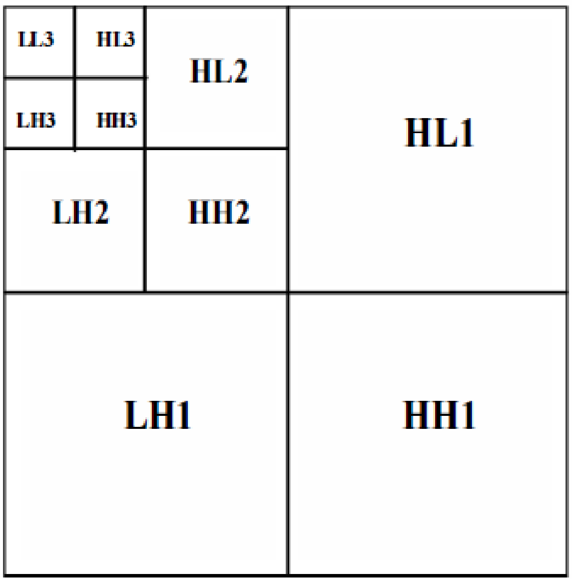

Figure 2.1: Image Decomposition (3 Level) using DWT[4]

2.2 Thresholding

The wavelet thresholding method relies on the basic idea that the energy of a signal will often be concentrated in a few coefficients in wavelet domain while the energy of noise is spread among all coefficients in wavelet domain. Therefore, the nonlinear thresholding function in wavelet domain willtend to keep a few larger coefficients representing the signal. while the noise coefficients will tend to reduce to zero.

.

2.3 Behaviour of Hard and soft thresholding

Two standard thresholding policies are: *hard-thresholding*, and *soft-thresholding*. In both cases, the coefficients that are below a certain threshold are set to zero. In hard thresholding, the remaining coefficients are left unchanged

$Thard(w)$ ={ 0, if $|w| \leq T$ *and* w, if $|w| > T$. }

In soft thresholding, the magnitudes of the coefficients above threshold

are reduced by an amount equal to the value of the threshold

$Tsoft(w)$ = { 0, if $|w| \leq T$, $\mathrm{sgn}(w)(|w| - T)$, if $|w| > T$. }

We can say that in both cases each wavelet coefficient is multiplied by a given *shrinkage factor*, which is a function of the magnitude of the coefficient (Fig. 2.12).

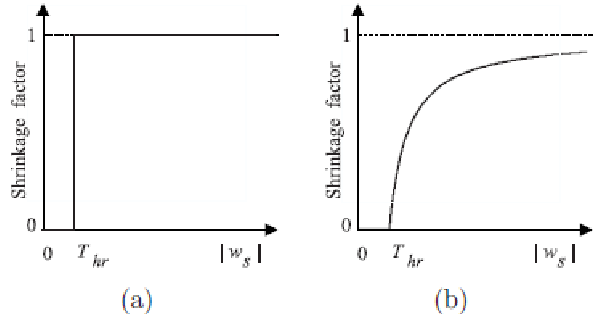

Figure 2.2: Shrinkage factors that multiply the wavelet coefficients in (a) hard-thresholding and (b) soft-thresholding[55].

In soft thresholding, the estimates are biased: large coefficients are always reduced in magnitude; therefore, the mathematical expectations of their estimates differ from the observed values. The reconstructed image is often oversmoothed. On the other side, a disadvantage of the hard thresholding is its abrupt discontinuity: estimates have a larger variance and may be highly sensitive to small changes in the data. In practice, especially when the noise level is high, hard thresholding yields abrupt artifacts in the reconstructed image. Due to this, in image processing applications the soft thresholding is usually preferred over the hard one

Step 3) Applying Fuzzy-c-means clustering for initial segmentation

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981)is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

$$1 \leq m < \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an

ISSN: 2278 – 1323

iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \qquad \ldots\ldots\ldots\ldots 3$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad \ldots\ldots\ldots\ldots 4$$

This iteration will stop when

$$\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon \qquad \ldots\ldots\ldots\ldots 5$$

where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

The algorithm is composed of the following steps:

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$

2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

3. Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4. If $\| U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP; otherwise return to step 2.

FCM has been used to locate the seed point in the image after which region growing has been applied for lesion region segmentation. The region growing algorithm has been explained in next step.

Applying Region growing algorithm

The basic idea of region growing is a collection of pixels with similar properties to form a region [5] . First, we need to find a seed pixel as a started point for each of needed segmentation. And then merge the same or similar property of pixel (Based on a pre-determined growing or similar formula to determine) with the seed pixel around the seed pixel domain into the domain of seed pixel. These new pixels as a new seed pixel to continue the above process until no more pixels that satisfy the condition can be included, and then the region has grown. In the practical application of this method we need to address three questions: first, chose or determined a group of seed pixel which can correctly represent the required region; second, fixed the formula which can contain the adjacent pixels in the growth; third, made rules or conditions to stop the growth process [6] [7]. The advantage of region growing algorithm is easy to complete and compute. Similar to the threshold, the region growing methods are rarely used alone; it is often used with other segmentation methods. The practical method of this subject combines the watershed algorithm and region growing algorithm for color image segmentation. The disadvantage of region growing: first, it needs human interaction to obtain the seed point, so that the user needs to implant a seed point in every region which needs to extract; second, The patterns of regional growth are also sensitive to noise as result the extracted region has empty or links the separate region under the case of local effect [8]. This article according a certain rules to automatic select seed pixels as well as effective solve the first question. We carry on the regional growing on the basis of the watershed segmentation algorithm; this method effectively solved the second questions. Domain decomposition technique makes seed region continually split into four rectangular regions until the internal of every region is similar. Region merging is often combines with the region growing and domain decomposition in order to merge the similar sub-region into a domain as large as possible. The disadvantage of domain decomposition technique may cause destruction of the border.

The main factor other than seed point on which region growing methods depend, is growing formula. Initial seeds grow by integrating their neighborhoods which fulfil some similarity criteria [9][10] which is either determined heuristically or from some simple algorithm like Otsu's adaptive threshold technique. The region growing formula should be capable of fulfilling the following homogeneity criteria:

- Pixels must be homogeneous with respect to some properties, inside one region.
- Pixels must have distinct properties with respect to pixels from other region. Region growing formula should also be time efficient and able to segment a wide range of images.

ISSN: 2278 – 1323

III. PARAMETERS USED

(i)Sensitivity is the proportion of patients *with* disease who test positive. In probability notation: $P(T^+|D^+) = TP / (TP+FN)$.

(ii)Specificity is the proportion of patients *without* disease who test negative. In probability notation: $P(T^-|D^-) = TN / (TN + FP)$.

IV. RESULTS AND DISCUSSIONS

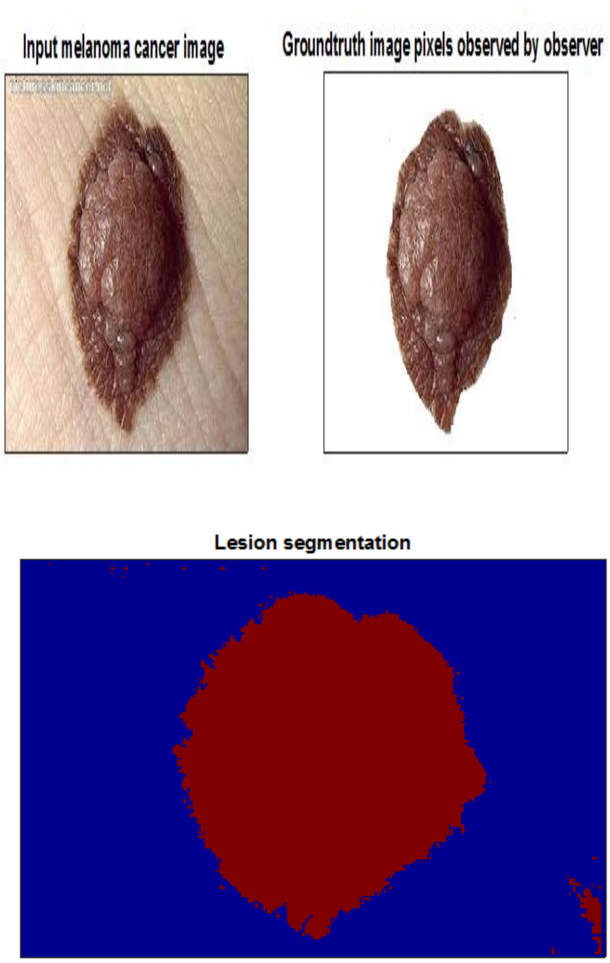

Figure4.1: (a) input melanoma image (b) ground truth image (c) lesion segmentation

4.1 Accuracy in terms of sensitivity

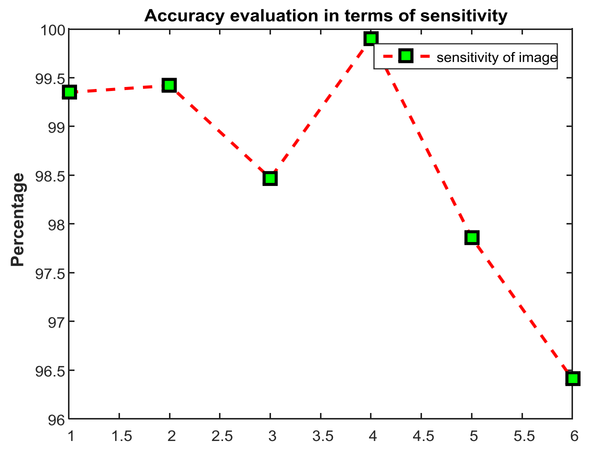

Figure4.1: graphical representation of senstivity

4.2 Accuracy in terms of specitivity

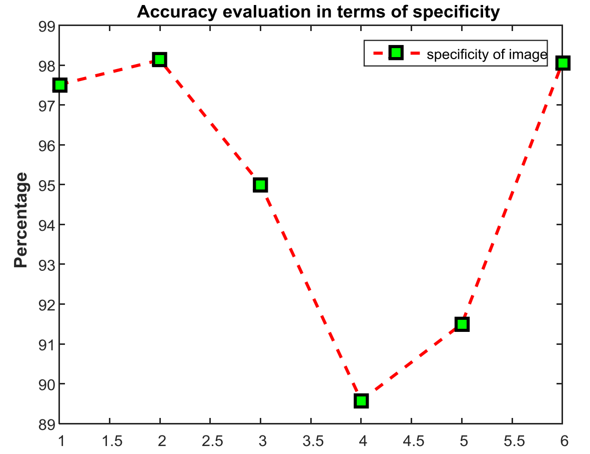

Figure4.3: graphical representation of specificity

V. CONCLUSION

Detecting skin cancer acts as a latest and burning topic in the field of biomedical science. And the study of literature review of different skin lesion and segmentation methods. Segmentation methods employs three steps i.e. preprocessing, segmentation and region refinement. Previous methods defines the skin lesion using methods such as ABCD rule , 2D histogram method and oversegmentation which leads to the problem of eye detector algorithm dependent and system complexity. Hence, there is a need for easy detection of skin lesion with less time consuming efforts which we would be trying in the future course of our dissertation work.

VI. REFERENCES

[1] R.Marks, "Epidemiology of melanoma," *Clin. Exp. Dermatol.* , vol. 25, pp. 459-63, Year 2000.

[2] World Health Organization, Ultraviolet Radiation and the iNTER SUN Programme. (2015. Jan.) .http:// www.who.int/uv/faq/skin cancer/en/

ISSN: 2278 – 1323

[3] L.Xu, M.Jackowaski, A. Goshtasby, D. Roseman, S. Bines, C.Yu, A.Dhawan, A. Huntley , " Segmentation of skin cancer images" *Elsevier image and vision computing,* Vol.17 Page no. 65-74, Year 1999.

[4] Rafael C.Gonzalez, Richard E.Woods and Steven L.Eddins, "Digital Image Processing Using MATLAB", Pearson Education (Singapore) Pte Ltd., Indian Branch, 482 F.I.E. Patparganj, Delhi 110092, India, 2004.

[5] J. Tang "A color image segmentation algorithm based on region growing", Proc. ICCET, vol. 6, pp.634-37, 2010.

[6] Z. He, et al, "The Application of Watershed in Segmentation of Overlapping Cells," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 10, pp. 840–49, Oct. 2002.

[7] Wilson, E., D.W. Sutter, and R.W. MahThe Application of Watershed in Segmentation of Overlapping Cells.. 2004. Honolulu, HI, United States: Acta Press, Anaheim, CA, United States.

[8] Bergmann, E.V., B.K. Walker, and B.K. Levy, The Research on Stereo Matching Algorithm based on Region-growth,"*Journal of Guidance, Control, and Dynamics*, p. 483-91, 1987

[9] Luis Garcia Ugarriza, Eli Saber, SreenathRaoVantaram, Vincent Amuso, Mark Shaw, and RanjitBhaskar, "Automatic Image Segmentation by Dynamic Region Growth and Multiresolution Merging " *IEEE transactions on image processing*, VOL. 18, NO. 10, October 2009.

[10] S. A. Hojjatoleslami and F. Kruggel, "Segmentation of Large Brain Lesions ", *IEEE transactins on medical imaging,* Vol. 20, No. 7, July 2001.

AUTHORS

***First Author** – Sukhdeep Singh completed his B.Tech. From RBIEBT KHARAR under Punjab Technical University. He is pursuing M.Tech in Electronics and Communication from University College of Engineering, Punjabi University, Patiala.

****Second Author** – Dipti Bansal working as Assistant Professor in University College of Engineering, Punjabi University,Patiala.

ISSN: 2278 – 1323