

## Human Skin Cancer Recognition and Classification by Unified Skin Texture and Color Features

Dr. Shubhangi D C<sup>1</sup>, Nagaraj<sup>2</sup>

<sup>1</sup>Professor, Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India,

<sup>2</sup>Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India,

**ABSTRACT:** In this paper we have proposed a novel method called automatic segmentation of skin lesion in conventional macroscopic images. Many approaches have been proposed to determine the skin cancer. An extensive literature survey is done to study the state-of-art techniques for skin cancer recognition; level set active contours (LSAC), skin lesion segmentation (SLS) and multidirectional gradient vector flow (MGVF) have given considerable results. A technique based on stochastic region merging (SRM) and region adjacency graph (RAG) is adopted in the proposed method. Segmenting the skin lesion from macroscopic images is a very challenging problem due to some factor such as, illumination variation, presence of hair, irregular skin color variation and multiple unhealthy skin regions. To solve all these factors we have introduced a new approach called novel iterative stochastic region merging likelihood for segmenting the skin lesion from macroscopic images based on the discrete wavelet transformation (DWT).

**Keywords:** Lesion, MRF, RAG, ROI, Skin Cancer, SRM.

Submitted Date 24 June 2013

Accepted Date: 29 June 2013

### I. Introduction

Skin cancer [1] is a malignant tumor, able to invade surrounding tissues and metastasize (or spread) to other parts of the body. There are mainly three different types of skin cancer: Basal cell carcinoma, Squamous cell carcinoma and Melanoma. Every year, new cases of skin cancer are logged than the combined incidence of cancers of the breast, prostate, lung and colon. The sun is the primary source of excessive ultraviolet (UV) radiation. Thus, immediate adverse effect of excessive exposure to sun results in sunburn and eye damage, longer effects include premature aging of the skin and skin cancer. The biological description about the layers of the human skin is shown in” Fig.1”.

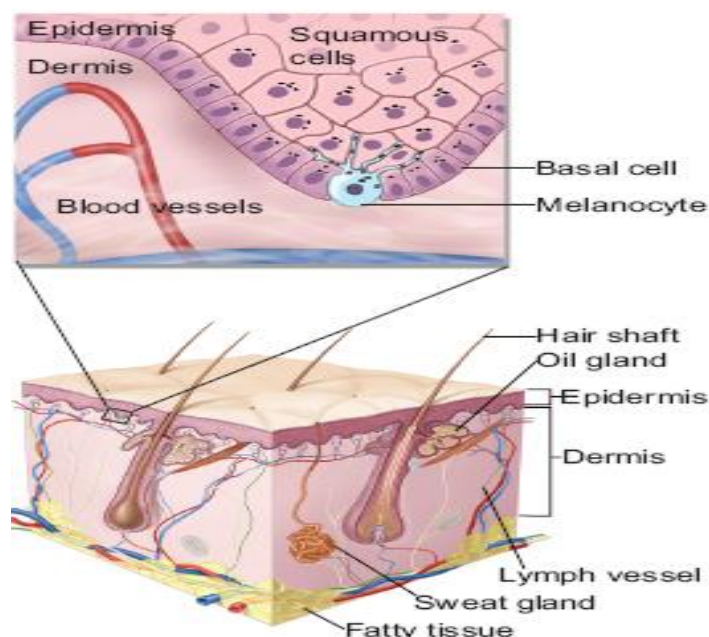


Fig.1 Layers of the skin-the epidermis and dermis

There are so many technologies to diagnose and treat the skin cancer. Remote clinical diagnosis is important in telemedicine, for improving the skin cancer diagnosis where the macroscopic images of the patient's defected skin area will be taken using conventional cameras by local clinician and those will be sent for remote examination through internet. In this way we have pre-screened the defected skin lesion with the help of conventional cameras and if the image analysis suggests a lesion that needs special attention then the patient has to be immediately referred to a dermatologist. There is another technology called computer aided system(CAS), to assist in the rapid clinical analysis and diagnosis of dermatological skin lesions(e.g., melanomas) [2],[3]. CAS is effective and it had been tested remotely using internet. It is used for conventional macroscopic images but not for the dermoscopic images. The segmentation of skin cancer part from macroscopic image is a very challenging problem due to the factors like color variation, noise and artifacts and weak boundary separation as shown in "Fig.2".

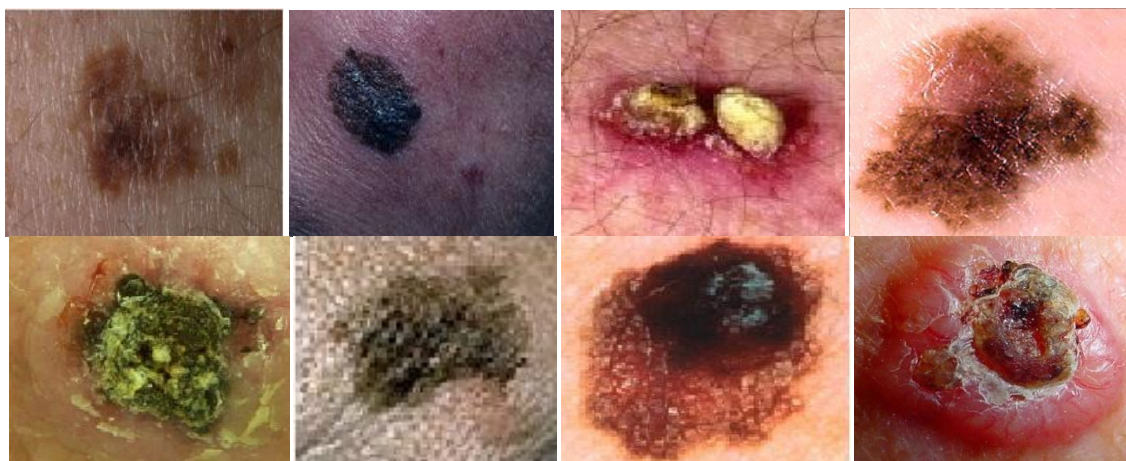


Fig.2 Macroscopic images of skin cancer regions

Many methods have been designed to solve these problems. Popular approaches to skin lesion segmentation are thresholding [3]-[6], and active contours [7]-[9]. Thresholding is most suitable when skin lesions have consistent characteristics and surrounding skin regions are homogeneous. Due to the presence of poor illumination, color variation finding the clear threshold is a challenging task. Active contours can differentiate the skin lesion region more accurately. Thresholding fails to handle noise and artifacts, variation in color and illumination which can be efficiently managed by Active contours. Markov random fields (MRF) [13] and other statistical based methods [10] [11] are studied which consider local spatial interactions and densities to segment skin lesion. Initially the image is divided into  $m \times n$  pixel [13]. Each region is assigned with a unique number, such regions will form adjacent matrix. Then we apply statistical region merging (SRM) [12] for skin lesion segmentation as proposed by celebi et al [11].

Above mentioned techniques leads to key drawbacks and to solve all these problems a new approach called "Iterative stochastic region merging" came into existence. Statistical region merging follows single pass strategy to segment in linear time/space but it does not yield accurate results compared to stochastic region merging. In this paper the iterative stochastic region merging method is introduced which can efficiently handle noise and artifacts, illumination, color, texture variation. It is improved to separate the weak boundary based on the thresholding level of the image, which are the key challenges to skin lesion segmentation in macroscopic image. The proposed method differs from current methods based on the statically region merging algorithm [12]. The proposed method uses merging criteria by introducing a new likelihood function that allocates stochastic region merging decision and exploits multi-pass strategy to get accurate results and segmentation. It is concluded that the proposed method is better and gives accurate results compared to the current methods.

## II. Related Work

There are number of approaches in literature for the evaluation of skin lesion segmentation. An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm, remote user uploads a dermoscopic image: separates the tumor area from the surrounding skin [3]. Measurement of image features for diagnosis of melanoma requires that first the lesions be detected and localized in an image. It is essential that lesion boundaries are determined accurately. Image edges are then used to localize the boundary in that area. A closed elastic curve is fitted to the initial boundary, and is locally shrunk or expanded to approximate edges in its neighborhood in the area of focus. [4]. Unsupervised border detection in dermoscopy skin lesion images can be found which is based on a modified version of the JSEG algorithm [5] [11]. Type-2

fuzzy logic exhibits considerably better performance over some popular methods in determining the threshold value. Hence, this inspires that type-2 fuzzy logic techniques may be employed for accurate segmentation of dermoscopic pigmented skin lesion images [6]. Several algorithms have been proposed for the segmentation of skin lesions in dermoscopic images. They can be broadly classified as thresholding, edge based or region-based methods [7]. The gradient vector flow (GVF) snake algorithm was investigated to find the skin lesion borders of dermoscopy images automatically. The percentage border error obtained for the GVF-based method is more consistent for the benign and melanoma lesions examined than Pagadala's color thresholding-based approach[8] [9]. Mean shift based fuzzy c-means algorithm incorporates a mean field term within the standard fuzzy c-means objective function [10]. Statistical region merging, a statistical basis for image segmentation by region merging following a particular order in the choice of regions. This approach can be efficiently approximated in linear time/space, leading to a fast segmentation algorithm tailored to processing images described using most common numerical pixel attribute spaces [12]. Cross entropy (CE), into the MRF theory for medical image segmentation is based on the theory of rare event simulation, is general and stochastic [13]

### III. System Design

In system design there are two phases, training phase and testing phase. The overall system design is explained in “Fig.3”.

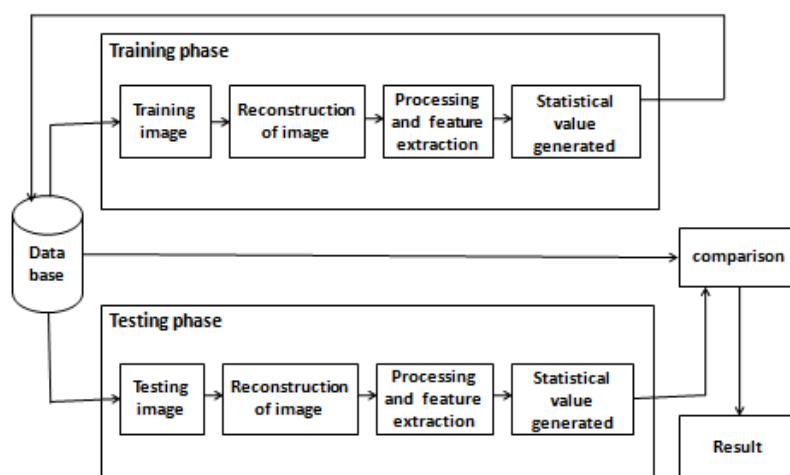


Fig.3 System design

The different images of skin cancer disease are stored in database.

$$I_D = \{i_{d1}, i_{d2}, i_{d3}, \dots\} \tag{1}$$

Where, I- Standard set

i- Individual image

D- Standard data set

d- Individual data set

In the training phase images are chosen from different data sets which are of different diseases for the training purpose.

$$T_D = \{t_{d1}, t_{d2}, t_{d3}, \dots\} \tag{2}$$

Where, T- Standard training

t- Individual image

D- Standard data set

d- Individual data set

Trained images are to be reconstructed by converting those images into lab image where resizing, image compression, gray scale conversion and reshaping have to be done. Later, adjusting the counters and extracting the edges are done from reconstructed images. Further processing and feature extraction are analyzed as shown in “Fig.4” which generates the statistical values and will be stored in database. Testing phase also employs the same sequence steps applied on the information stored in the database. Finally the evaluation will be made with reference to the trained image sets and results will be obtained.

**Processing And Feature Extraction**

Initially MRF is applied on the input image that divides the image into MxN regions, which leads to ROI. Color, texture, skewness, kurtosis and area features are considered in classification and recognition. Based on the color human skin tone, hair, and defected parts (wound) varies. RGB and Lab a\*b\* color planes are considered.

$$\Delta E_{ab}^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \tag{3}$$

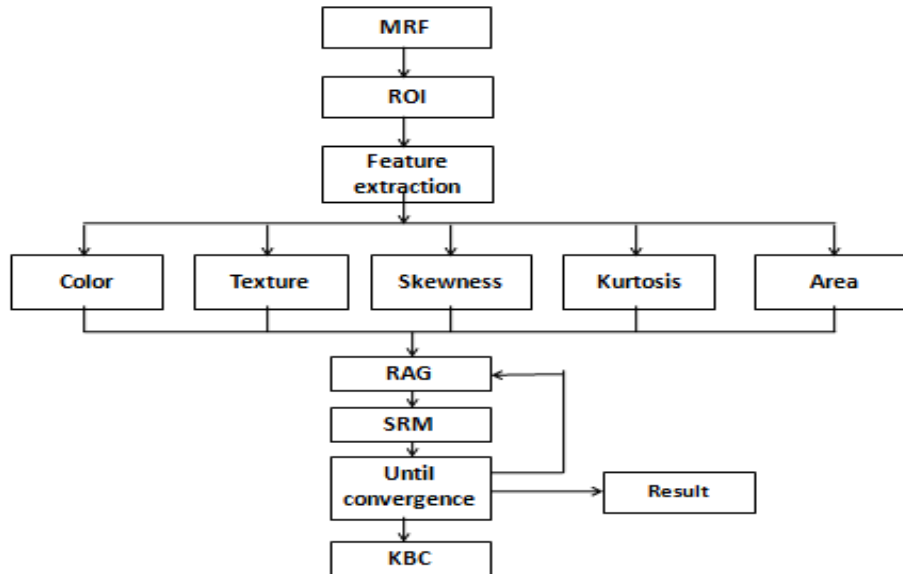


Fig.4 Processing and feature extraction

The GLCM is used for texture extraction. It contains important information about the texture in the examined area of the image. The gray level co-occurrence matrix (GLCM) is just the tool to start and then get the indicators:

$$\text{Entropy} = \sum_{i,j=1}^N P(i,j) \log(P(i,j)) \tag{4}$$

$$\text{Energy} = \sum_{i,j=1}^N P(i,j)^2 \tag{5}$$

$$\text{Contrast} = \sum_{i,j=1}^N P(i,j) P(i,j)^2 \tag{6}$$

The skewness of a random variable  $X$  is denoted or  $skew(X)$ . It is defined as:

$$\text{Skew}(X) = \frac{E(X-M)^3}{n^3} \tag{7}$$

Kurtosis Positive kurtosis means distribution is sharper than a normal distribution. Negative kurtosis means distribution is flatter than a normal distribution. The kurtosis of a random variable  $X$  is denoted or  $Kurt(X)$ :

$$\text{Kurt}(X) = \frac{E(X-M)^4}{n^3} \tag{8}$$

After convergences apply KBC (knowledge based classifier) to classify the values of the image (statistical) based on the knowledge available (pixel value, texture, color, skewness and kurtosis). The values are grouped separately depends on the matching pixels and classify the type of cancer.

**IV. Algorithm**

1. Initially divide the image I into MxN blocks where each block is consisting of single pixel.
2. Each block is assigned a unique number and termed as regions.
3. Construct an initial region adjacency graph where each region r is compared with its neighbors.
4. **repeat** step 3
5. Begin representation of feature from the block  $b_i$  of image I,  
 $Queue = \sum_{i=0}^n r_i(I)$
6. Then terminate feature extraction and restore all the adjacent region pairs.
7. Place all adjacent region pairs into a priority queue based on ascending regional expectation differences
8. **repeat** step 7
9. Remove region pairs  $R_a$  and  $R_b$  from priority queue. Restore with a probability  $\alpha (R_a, R_b)$  based on the proposed region merging likelihood function.
10. After restoring the image update the adjacency graph.

- 11. **until** priority queue is empty
- 12. Decrease Q by half
- 13. **until** convergence

### V. Materias And Methods

The work requires the images of the cancer skin part which are taken from the standard resources. With the aid of the Matlab-8 and Digital Image Processing methods the work will give us the accurate results. It indeed referred with an expert person before concluding.

### VI. Results And Discussion

The following experimental results have been obtained by performing the digital image processing techniques. The image data set  $I_D$  has been constructed by the different skin cancer types as  $I_{BCC}$ ,  $I_{SCC}$ ,  $I_{mel}$  and  $I_{act}$ . The process of human skin cancer recognition and classification by unified skin texture and color features has been performed. The resulting name of diseases after passing an input testing image is displayed as the result for the experiment and stored for the feature enhancement.

This is the initial GUI which accepts the image having Skin Cancer for testing purpose. The image has been given as input to the system as input testing image  $T_D$ .



Fig.5 Input image

The image has been reconstructed by using discrete wave late transformation (DWT) method during which the image compression, denoising and gray image conversion will be performed.

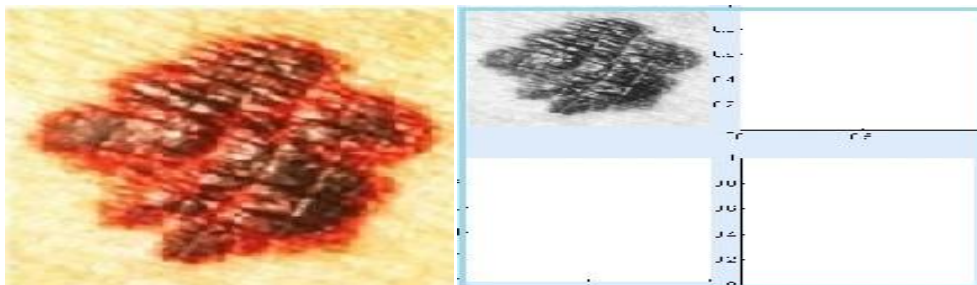


Fig.6 Reconstructed Image

Reconstructed image is processed to perform the subsequent events like RAG and SRM which indeed processed after the creation of MRF which has been stored temporary data stack which will be extracted during feature extraction process

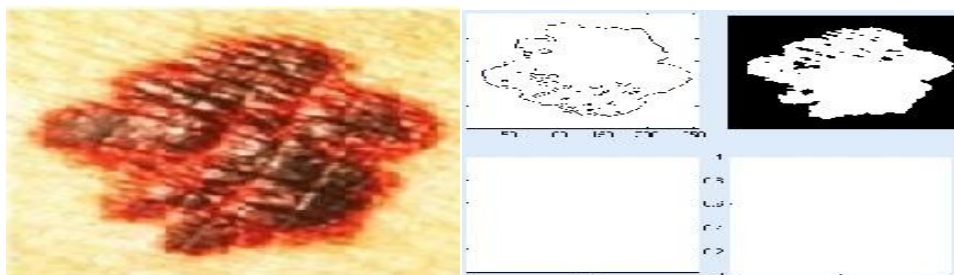


Fig.7 Region extracted by SRM and MRF

During the feature extraction process the stored pixel value of the infected part of the human skin are extracted and the features like mean, variants, standard deviation, kurtosis, skewness etc., are calculated for each

occurring pixels these values are stored in the system data base which are useful in classification and recognition phases.

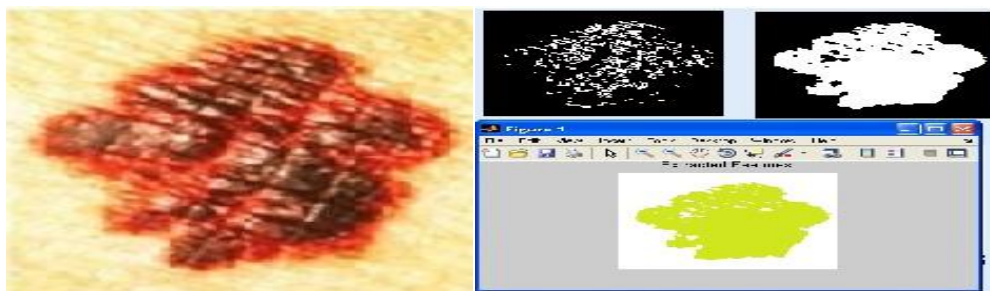


Fig.8 Feature Extraction

The resulting values after the feature extraction are applied on the input image given and the diseased part of the human skin image is classified as shown.

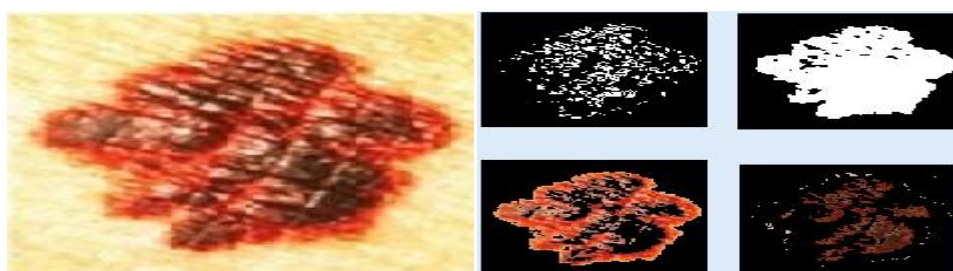


Fig.9 Convergence reached final region

The values of all the images which have been stored earlier are compared with the input image given and the nearest value matching will be displayed based on nearest neighbor method. The result is stored and processed for the future work.



Fig.10 Result of the system

Multidirectional gradient vector flow (MGVF): The multi-direction GVF snake extends the single direction GVF snake and traces the boundary of the skin cancer even if there are other objects near the skin cancer region. The LSAC method was chosen as it represents the state of the art in region based active contour methods, as well as the SLS and MGVF methods, which were shown to provide strong segmentation accuracy for various types of skin cancer regions, thus acting as good indicators of the level of segmentation accuracy that can be achieved by current techniques in this area. It should be noted that the parameters of the LSAC, SLS, and MGVF methods have been set based on that presented in their respective literatures, given that the goal is for fully automatic skin lesion segmentation without manual intervention.

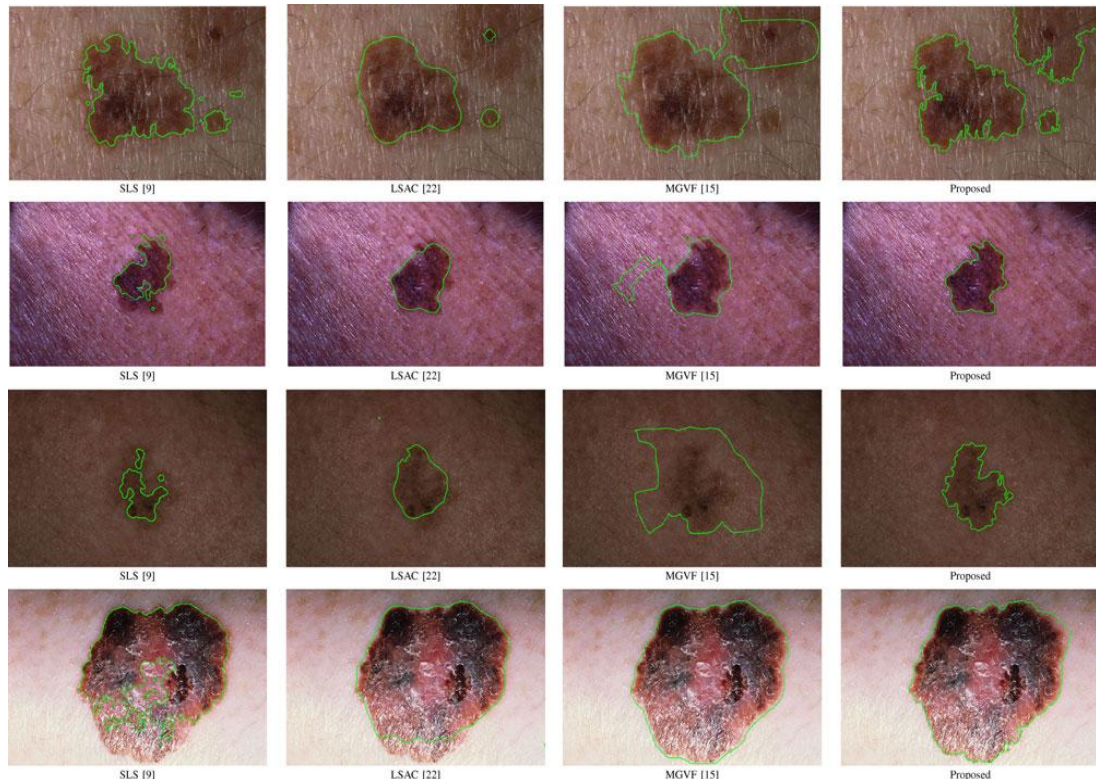


Fig.11 Sample skin lesion segmentation results for Tests 1, 2, 3, and 4 (first to fourth rows, respectively)

TABLE I

Segmentation performance metrics on 60 real macroscopic images, based on measurements made by the first expert

Method	SE (%)	TDR (%)	FPR (%)
LASC[22]	35.24	70.49	1.83
SLS[9]	27.98	75.28	2.94
MGVF[15]	26.74	95.92	15.78
DWT	25.32	96.78	18.38

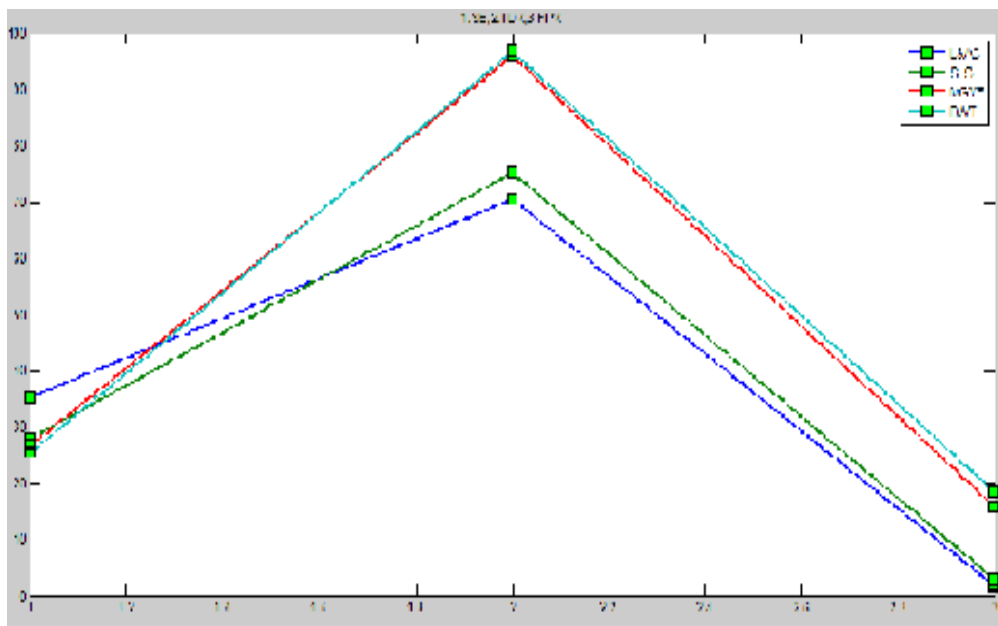


Fig.12 The resulting comparison graph 1.SE 2.TDR 3.FPR

These are the formula used for all the above techniques used for detection of skin lesion and are compared with proposed values.

1. The segmentation error (SE) as defined by:

$$SE = \frac{A(SR \oplus GT)}{A(GT)} \times 100\% \quad (9)$$

Where  $\oplus$  denotes XOR operation and  $A(\cdot)$  denotes the area.

2. The true detection rate (TDR) as defined by:

$$TDR = \frac{A(SR \cap GT)}{A(GT)} \times 100\% \quad (10)$$

3. The false positive rate (FPR) as defined by:

$$FPR = \frac{A(SR \cap \bar{GT})}{A(\bar{GT})} \times 100\% \quad (11)$$

The SE metric provides a good indication of the overall segmentation performance, but does not give a good indication on specific characteristics of the segmentation methods. To complement this overall metric, the TDR metric provides a good indication of the under segmentation error while the FPR metric provides a good indication of the over segmentation error. As such, the use of these three metrics provides an overall picture of the different aspects of the segmentation algorithms.

## VII. Conclusion

By adopting the proposed image processing technologies like stochastic region merging (SRM) and region adjacency graph (RAG) the detection of the affected human skin cancer disease has been improved to the larger extent by reducing segmentation error (SE) and increasing the true detection rate (TDR) and false positive rate (FPR) as compared to other techniques available. The proposed method also performs the task efficiently with less consumption of time in determining the infected and uninfected part of human skin. But the major drawback of the work is raised when user tries to give the diseased image of untrained type and it also fails during new type of skin cancer image input. Future work can be made to overcome the above mentioned major drawbacks with addition of larger data set and automatic system.

## References

- [1] Skin Cancer Foundation, "Skin Cancer Facts," [Online]. Available: <http://www.skincancer.org/Skin-Cancer-Facts/>, Accessed: Jul. 4, 2010.
- [2] H. Oka, M. Hashimoto, H. Iyatomi, and M. Tanaka, "Internet-based program for automatic discrimination of dermoscopic images between melanoma and Clark nevi," *Br. J. Dermatology.*, vol. 150, no. 5, p. 1041, 2004.
- [3] H. Iyatomi, H. Oka, M. Emre Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Comput. Med. Imag. Graph.*, vol. 32, pp. 566–579, 2008.
- [4] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, and A. Huntley, "Segmentation of skin cancer images," *Image Vis. Comput.*, vol. 17, pp. 65–74, 1999.
- [5] M. Celebi, Y. Aslandogan, W. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Res. Tech-nol.*, vol. 13, pp. 454–462, 2007.
- [6] M. Yuksel and M. Borlu, "Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 976–981, Aug. 2009. \*
- [7] M. Silveira, J. Nascimento, J. Marques, A. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.
- [8] B. Erkol, R. Moss, R. Stanley, W. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Res. Technol.*, vol. 11, pp. 17–26, 2005.
- [9] J. Tang, "A multi-direction GVF snake for the segmentation of skin cancer images," *Pattern Recognit.*, vol. 42, no. 6, pp. 1172–1179, 2009.
- [10] H. Zhou, G. Schaefer, A. Sadka, and M. E. Celebi, "Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 26–34, Feb. 2009.
- [11] M. Celebi, H. Kingravi, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, J. Malter, J. Grichnik, A. Marghoob, H. Rabinovitz, and S. Menzies, "Border detection in dermoscopy image using statistical region merging," *Skin Res. Technol.*, vol. 14, pp. 347–353, 2008.
- [12] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1452–1458, Nov. 2004. [19] M. Celebi, H. Iyatomi, G. Schaefer, and W. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.* vol. 33, pp. 148–153, 2009.
- [13] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York: Springer-Verlag, 2001.
- [14] Q. Yu and D. A. Clausi, "IRGS: Image segmentation using edge penalties and region growing," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 30, no. 12, pp. 2126–2139, Dec. 2008.
- [15] C. Li, C. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.