

Image Processing for Skin Cancer Features Extraction

Md.Amran Hossen Bhuiyan, Ibrahim Azad, Md.Kamal Uddin

{amran_apece,ibrahimazad_apece,cse_kamal_ud}@yahoo.com

Department of Computer Science & Telecommunication Engineering, Noakhali Science & Technology University, Bangladesh.

Abstract— The application of image processing for diagnostics purpose is a non-invasive technique. There is currently a great interest in the prospects of automatic image analysis method for image processing, both to provide quantitative information about a lesion, which can be relevance for the clinical, and as a standalone early warning tool. In order to achieve an effective way to identify skin cancer at an early stage without performing any unnecessary skin biopsies, digital images of melanoma skin lesions have been investigated. To achieve this goal, feature extraction is considered as an essential-weapon to analyze an image appropriately. In this paper, different digital images have been analyzed based on unsupervised segmentation techniques. Feature extraction techniques are then applied on these segmented images. After this, a comprehensive discussion has been explored based on the obtained results.

Keywords: Clustering, Dilation, Erosion, Features, Image Processing, Melanoma, Supervised, Segmentation.

I. INTRODUCTION

Human Cancer is a complex disease caused primarily by genetic instability and accumulation of multiple molecular alternations [1],[2]. Current diagnostic and prognostic classifications do not reflect the whole clinical heterogeneity of tumors and are insufficient to make prediction for successful treatment and patient outcome [3],[4]. Most of the currently applied anti-cancer agents do not greatly differentiate between cancerous and normal cells. In addition cancer is often diagnosed and treated too late, when the cancer cells have already invaded and metastasized into other parts of the body. At the time of clinical presentation, a great percentage of patients with breast, lung, colon, prostate, and ovarian cancer have hidden and over metastatic colonies [5]. At this stage, therapeutic modalities are limited in their effectiveness. Due to these problems, cancer has overtaken heart disease as the leading cause death for any age in all over the world.

Among many types of cancer, Skin cancers are the most common form of cancers in human [6]. It is severe among the faired-skinned population in Europe, North America, and Australia. There are two major types of skin cancer, name malignant melanoma and non-melanoma (basal cell, squamous cell, and markel cell carcinomas, etc.) [7]. Melanoma is more dangerous and can be fatal if not treated. If melanoma is detected in its early stages, it is highly curable, yet advanced melanoma is lethal.

It is well-known that early finding and treatment of skin cancer can reduce the mortality and morbidity of patients. Digital Dermoscopy is widely considered as one of the most cost effective means to identify and classify skin-cancer. An automatic dermoscopy image analysis system [8] has usually three stages: (1) Proper Segmentation, (2) feature extraction and selection, and (3) lesion recognition. The proper segmentation is the most important, since it affects the precision of the subsequent steps. Supervised segmentation is somewhat easy to implement by varying its parameters for variety of lesion shapes, sizes, and colors along with diverse skin types and textures. But the unsupervised segmentation is a difficult task due to the above mentioned properties.

Although the significant research effort have gone into developing computerized algorithms to segment dermoscopic image properly, still some significant limitations exists in each efforts. And very few studies dealt with the automatic segmentation of dermoscopic image.

In order to achieve an effective way to identify skin cancer at an early stage without performing any unnecessary skin biopsies, digital images of melanoma skin lesions have been investigated. To achieve this goal, feature extraction is considered as an essential-weapon to analyze an image appropriately. In this thesis work, different digital images have been analyzed based on unsupervised or automatic segmentations segmentation techniques. Feature extraction techniques are then applied on these segmented images. After this, a comprehensive discussion has been explored based on the obtained results.

II. SEGMENTATION

. The segmentation is the most important stage for analyzing image properly since it affects the accuracy of the subsequent steps. However, proper segmentation is difficult because of the great verities of the lesion shapes, sizes, and colors along with different skin types and textures. In addition, some lesions have irregular boundaries and in some cases there is smooth transition between the lesion and the skin. To address this problem, several algorithms have been proposed. They can be broadly classified as thresholding, edge-based or region-based methods. In this thesis, three methods of segmentation have been discussed. The methods are:

- Otsu's method.
- Gradient Vector Flow (GVF)
- Color Based Image Segmentation Using K-mean Clustering.

Although fully unsupervised segmentation would be desirable, GVF method requires some degree of user interaction in order to achieve good performance. The following sections describe each of the methods in details.

A. Otsu's Method

Otsu's method is optimal for thresholding objects from the background. This technique is based on a discriminate analysis which partitions the image into two classes. Given an image represented in L gray levels {0,1,2,...,L}, Otsu's thresholding method [9] partitions the image pixels into two classes C0={0,1,2,...,t} & C1={t+1,t+2,...,L-1}. Let the number of pixels in the *i*th gray level be *n_i*, and *n* be the total number pixels in a given image. The probability of occurrence of gray level *i* is defined as:

$$p_i = \frac{n_i}{n}$$

C₀ and C₁ are normally corresponding to the object of interest and the background, the probabilities of the classes are W₀ and W₁,

$$W_0 = \sum_{i=0}^t p_i \quad \text{and} \quad W_1 = \sum_{i=t+1}^{L-1} p_i$$

Thus, the mean of the two classes can be computed as:

$$\mu_0(t) = \sum_{i=0}^t \frac{ip_i}{W_0(t)} \quad \text{and} \quad \mu_1(t) = \sum_{i=t+1}^{L-1} \frac{ip_i}{W_1(t)}$$

An optimal threshold *t** can be obtained by maximizing the class variance:

$$t^* = \text{Arg} \left\{ \underset{0 \leq i \leq L-1}{\text{Max}} \frac{\sigma_B^2}{\sigma_T^2} \right\}$$

Where, Class Variance,

$$\sigma_B^2 = W_0 (\mu_0(t) - \mu_T(t))^2 + W_1 (\mu_1(t) - \mu_T(t))^2$$

$$\text{Total Variance, } \sigma_T^2 = \sum_{i=0}^{L-1} (i - \mu_T)^2$$

$$\text{Total Mean, } \mu_T = \sum_{i=0}^{L-1} ip_i$$

$$\text{So, } t^* = \text{Arg} \left\{ \underset{0 \leq i \leq L-1}{\text{Max}} \left[W_0 (\mu_0(t) - \mu_T(t))^2 + W_1 (\mu_1(t) - \mu_T(t))^2 \right] \right\}$$

Otsu's method of thresholding gray level images is efficient for separating an image into two classes where two types of fairly distinct classes exist in the image.

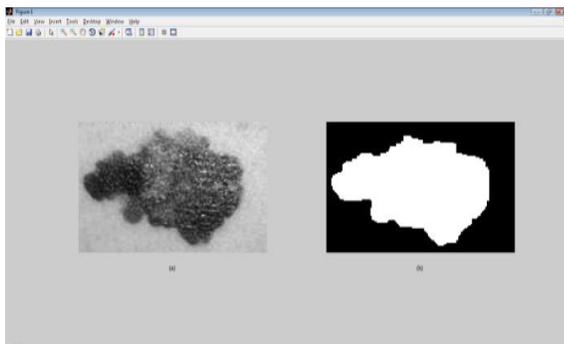


Fig.1: (a) Grayscale version of RGB image; (b) Segmented image after applying Otsu's method

B. Gradient Vector Flow (GVF)

The GVF snake is a well-known algorithm proposed in [10] which has been successfully used in many medical imaging problems. The object boundary is approximated by an elastic contour $X(s) = (X(s), Y(s))$, $S \in [0,1]$ which is initialized in the image domain by the user or the heuristic criteria. The elastic contour is then modified according to the differential equation:

$$\frac{dX(s,t)}{dt} = F_{int}(X(s,t)) + V_{int}(X(s,t))$$

Where, F_{int} is an internal force, similar to the one used in traditional snakes to keep the shape continuity and smoothness and $V = (u(x, y), v(x, y))$ is the GVF field. The GVF field is a regularized version of the image or edge gradient which allows long range attraction of the contour toward the object boundary even if the contour is located in a homogeneous region where the gradient is zero. V can be obtained by minimizing the energy:

$$E = \iint \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |v - \nabla f|^2 dx dy$$

As it can be seen, this is an example of a variation formulation of regularization. The parameter which adjusts the tradeoff between the first and second term of the integrand and is set according to the level of noise present in the image. Also, where the value of the edge gradient is small, energy is dominated by the sum of the partial derivatives of the gradient. In this case, setting $v = \nabla f$ minimizes the energy. Using the calculus of variations, it can be shown that the GVF field can be found by solving the pair of Euler equations stated below:

$$\mu \nabla^2 u - (u - f_x) (f_x^2 + f_y^2) = 0$$

$$\mu \nabla^2 v - (v - f_y) (f_x^2 + f_y^2) = 0$$

Here, ∇^2 is the Laplacian operation. The initialization of the GVF snake is automated. A circle with a given radius is placed on the image.

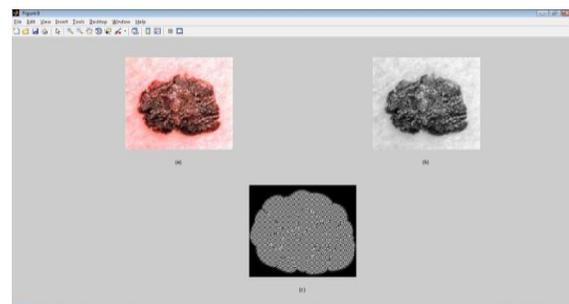


Fig.2: (a) Original RGB image; (b) Grayscale version of the RGB image; (c) Segmented image after using GVF method.

C. Color Based Image Segmentation Using K-mean Clustering

Image segmentation techniques can be differentiated into the following basic concepts: pixel oriented, contour-oriented, region-oriented, model-oriented, and color-oriented and hybrid. Color segmentation of image is a crucial operation in image analysis and in many computer vision, image interpolation, and pattern recognition system. The performance of color segmentation may significantly affect the quality of an image understanding system [11].

This segmentation process is divided into two stages. First enhancing color separation of medical image using de-correlation stretching is carried out and then the regions are grouped into set of three classes using k-mean clustering algorithm. Using this two step process, it is possible to reduce the computational cost avoiding feature calculation for every pixel in the image. Although the color is not frequently used for image segmentation, it gives a high discrimination power regions present in the image.

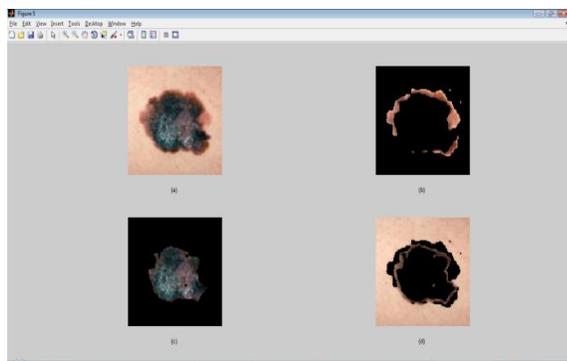


Fig.3: (a) RGB image; (b) objects in cluster 1; (c) objects in cluster 2; (d) objects in cluster 3.

III. FEATURE EXTRACTION

In automated diagnosis of skin lesions, feature extraction is based on the so-called ABCD-rule of dermatoscopy [12],[13]. ABCD represent the asymmetry, border structure, color variation, and dermatoscopical structure so called diameter of the lesion and define the basis for a diagnosis by a dermatologist.

A. Asymmetry

An important aspect of shape understanding is symmetry, which is very useful in pattern analysis. For a symmetric pattern, one needs only one half of the pattern with the axis of symmetry. If a part of the pattern is missing or noisy, with the help of symmetry one can complete the pattern or rid the pattern of noisy. To check for the degree of symmetry, there are two values of asymmetry feature i.e. Asymmetry Index (AI) and Lengthening Index. In this thesis, asymmetry index has been calculated.

Asymmetry Index is computed with the following equation:

$$AI = \frac{\Delta A}{A} \times 100 \quad (1)$$

Where, A= Area of the total Image. ΔA = Area difference between total image and lesion area.

B. Border Irregularity

In order to calculate border irregularity, there are different measures such as: compactness index, fractal index, edge abruptness, pigment transition. In this thesis work, compact index, fractal dimension, and edge abruptness has been calculated.

- Compact Index:

Density index (compact index/CI) is the measurement of the most popular form of barrier which estimates unanimous 2D objects. However, this measure is very sensitive to noise along the boundary. This can be determined by using the following equation:

$$CI = \frac{P_L^2}{4\pi A_L} \quad (2)$$

Where, P_L = Perimeter of the Lesion. A_L = Area of the Lesion.

- Fractal Dimension :

Fractal has characteristics self-similarity, and properties to the scale/size. Each section has a fractal which is different scale with the whole fractal. These characteristics causes suitable for fractal compression techniques. Another characteristic is fractal dimension. Dimension size generally an integer, such as the line has 1 dimension, the field has 2 dimensions, and cube has 3 dimensions and so on. However, fractal dimension is a strange as it may worth fraction. This fractal dimension can be used as a characteristic of an image.

Fractal dimension can be calculated by the method of calculation of the box (box-counting). To find the fractal dimension of an image, the Hausdorff dimension calculation method is simpler and effective one. Let's discuss this method with the help of an example:

Consider a squiggly line on a sheet of paper. Cover this line with 2-dimension cube of side e and let N (e) is the smallest of e-sided cubes that can cover this line. The dimension of this line is then:

$$D = \lim_{e \rightarrow 0} \frac{\ln(N(e))}{\ln(\frac{1}{e})} \quad (3)$$

Using equation (3), fractal dimension of an image can easily be calculated.

- Edge Abruptness:

Lesion with irregular boundaries (abruptness edge) has a large difference in radial distance (e.g. distance between the center d_2 and the barrier P.GL).

Barrier irregularity is estimated by analyzing the distribution of radial distance difference:

$$C_r = \frac{\frac{1}{P} \sum_{p \in C} d_2(P, GL - m_d)^2}{m_d^2} \quad (4)$$

Where, m_d is the mean distance of d_2 between the centered point and the barrier P.GL.

C. Color Variation

One early sign of melanoma is the emergence of color variations in color. Because melanoma cells grow in grower pigment, they are often colorful around brown, or black, depending on the production of the melanin pigment at different depth in the skin. The descriptors of color are mainly statistical parameters calculated from different color channels, like average value and standard deviation of the RGB or HSV color channel. In this thesis work, color variance of the RGB image has been calculated using HSV channel.

D. Diameter:

Melanoma tends to grow larger than common moles, and especially the diameter of 6mm. Because the wound is often irregular forms, to find the diameter, draw from all the edge pixels to the pixel edges through the midpoint and averaged.

IV. RESULTS OF SEGMENTATION PROCESS AND FEATURE EXTRACTION

Fig.4.1 (a) shows the original RGB image. Then, the grayscale image transformation applied on it. The resultant grayscale intensity image has been displayed on Fig.4.1 (b). After this, the Otsu's thresholding method has been applied on the grayscale intensity image. This gives the desired segmented image which has been shown on Fig.4.1(c).

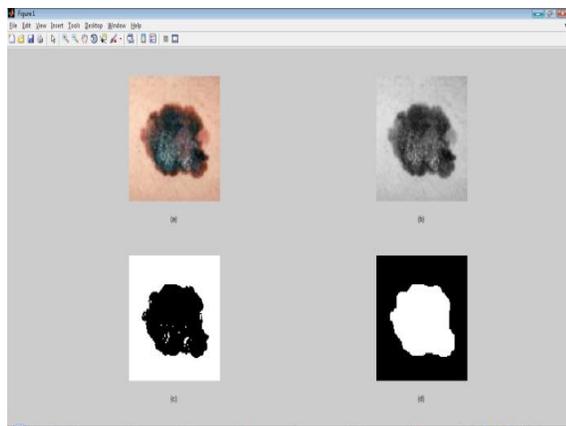


Fig.4.1: (a) Original RGB image; (b) Grayscale version of the RGB image; (c) Segmented image after using Otsu's thresholding method; (d) final segmented image after dilation and erosion.

In GVF segmentation method, the gradient vector flow (GVF) algorithm has been applied on grayscale intensity image. In this thesis, initially the GVF regularization coefficient

parameter is 0.25 which can vary for different irregular images. For proper segmentation, the number of iterations that has been used is 100 which may vary for different irregular images. The achieved segmented image is then undergoes through morphological dilation and erosion operations. Fig.4.2 shows the details of GVF segmentation method.

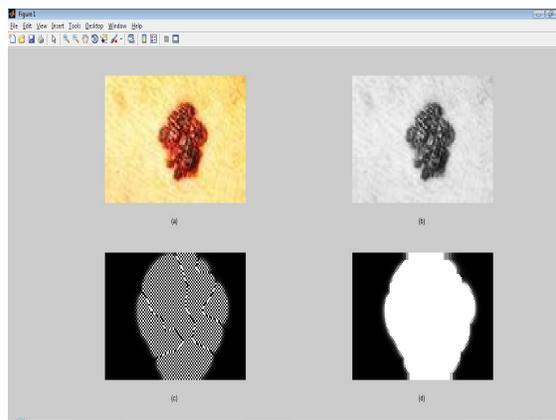


Fig.4.2: (a) Original RGB image; (b) Grayscale version of the RGB image; (c) Segmented image after using GVF method; (d) final segmented image after dilation and erosion.

In this thesis work, the Colored-based Segmentation using k-mean Clustering segmentation method is used to cluster the objects into three clusters by using Euclidean metric. The following figures show the object in cluster 1, object in cluster 2, cluster 3 and it has also show the blue nuclei.

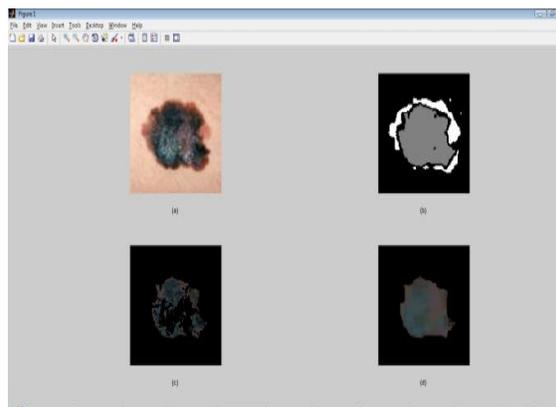


Fig.4.3: (a) Original RGB image; (b) Image labeled by cluster index; (c) blue nuclei; (d) final segmented image after dilation and erosion

In this thesis work, a set of 25 skin lesion images has been used as input data which contains both the melanoma skin cancer images and also noncancerous images. The performance of different techniques has been evaluated and extracted the desired features on these skin lesions. Fig.4.4 shows the users interface for observing the features value. The experimental finding of extracted feature is discussing from the following section.

A. Asymmetry Index:

Asymmetry is one of most important feature in image processing that helps to realize the degree of symmetry. To get proper asymmetry index value, it is required to segment the image properly which works as an input for calculating asymmetric index. In this thesis works, the asymmetric index has been calculated by using the equation (1) for all the skin lesions. After analyzing all the value of asymmetric index, this thesis came to a conclusion that a properly segmented skin lesion can be considered as asymmetric if its asymmetric index value falls within the range of 10%-25%. It has also been examined that some properly segmented image also fall beyond or below this reason.

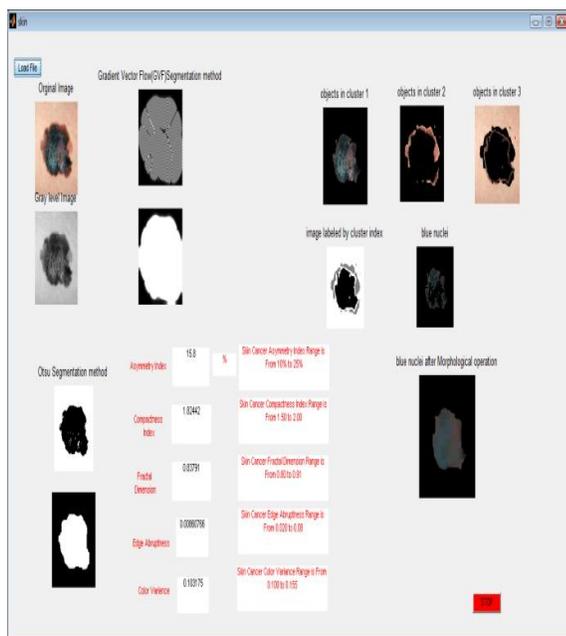


Fig.4.4: User interface for observing features value

B. Border Irregularity:

It has already been discussed that the border irregularity can be represented by using different measures. The experimental findings of all the measures discuss below:

- **Compactness Index:**

The compactness index of all the lesions have been calculated by using the equation (2). The proper segmentation is also required to achieve the true value of compactness index. Based on the obtained data, one can decide that the value of compactness index for the melanoma skin cancer fall within the range of 1.5-2.0.

- **Fractal Dimension:**

The computation of the fractal dimension of 2-dimensional image has become an important issue in various image processing applications. The fractal dimension is a measure of the similarity in a multiresolution representation. This technique also require proper segmented image as an input. In this thesis works, the fractal dimension of all the skin lesions

has been calculated by using Hausdorff dimension calculation method. The fractal dimension of all the skin lesions for melanoma skin cancer exhibits a range between 0.83-0.91. One skin lesions can be consider as a melanoma if its fractal dimension value fall within this range along with some other feature fulfillments.

- **Edge Abruptness:**

For calculating edge abruptness, boundary of the image has been traced. Based on this boundary, the radial distance between the center and the barrier has been calculated which exhibits the edge abruptness value. The experimental findings in this case showed that the value of edge abruptness of melanoma skin cancer fall within the range 0.02-0.08. It is also examined that some non-melanoma also exhibits this range. For this reason, other features should consider to determine a melanoma skin cancer.

C. Color Variation:

To calculate color variance, first step is to convert the RGB color map into an HSV color map. The column of the output matrix represents hue, saturation, and value respectively. In second step, the color variance of the HSV color map has been calculated which represent the actual color variance for an image. In this way, the color variance of all the skin lesions has been calculated. From the color variance data, one can draw a conclusion that the color variance of all the melanoma skin lesions is varying within the range of 0.10-0.15.

V. RESULT ANALYSIS

Image segmentation is the first step in early detection of cancer. To analyze skin lesions, it is necessary to accurately locate and isolate the lesions. In this thesis work, three unsupervised segmentation method for skin lesions have been discussed.

The Otsu's method is the oldest and simplest one. It has shown the best segmentation results among the three methods. It is fully unsupervised that does not require any change of parameter for different skin lesions. The second one, the gradient vector flow succeeds in active contour to boundary concavities, even with the presents of noise. The main drawback is its execution speed. It takes a long time to converge to objects. This method is not fully unsupervised. It required changing its parameter for different skin lesions. On the other hand, using color based segmentation; it is possible to reduce the computational cost avoiding feature calculation for every pixel in the image. Although the color is not frequently used for image segmentation, it gives a high discriminative power of regions present in the image. This kind of image segmentation may be used for mapping the change in land use land cover taken over temporal period in general but not in particular.

Feature extraction is considered as the most critical state-of-the-art skin cancer screening system. In this thesis, the feature extraction is based on ABCD-rule of dermatoscopy. Algorithms for extracting features have been disused. All the features have been calculated based on Otsu's segmentation method.

For asymmetry, asymmetry can be defined in different measures. But, it is still undefined that which measures would be the best for asymmetry determination. Since asymmetry is a critical feature in the diagnosis of skin cancer, asymmetry index was calculated in this thesis as a measure to determine the degree of symmetry. Asymmetric index has been calculated by comparing absolute area difference to the total area of lesion shape. The obtained results showed that this method is quite effective under the proper segmentation. If the image is not segmented properly which is the case of non-melanoma cancer, it will show erroneous results.

For border irregularity, three measures have been discussed: the compactness index, the fractal dimension and the edge abruptness. Among these three measures, fractal dimension showed better performance. But, for accurate measurement the combinations of these three have to be considered to decide whether the border is irregular or not.

For color variation, the color variance measurement is quite adequate. The results of color variance showed that the color variation of the entire melanoma skin cancer image maintain a certain range. This color variance measurement is very useful in the precise evaluation of perceptual closeness between different colors.

For diameter measurement, it has been found in many articles that the diameter greater than 6mm can be considered as melanoma. Since the image used in this thesis were mostly obtained from different sources and photographed with varying magnifications, lesions size could hardly be estimated from the image accurately. For this reason, the diameter of the lesions has not been calculated in this thesis. This is an important feature of melanoma, so inclusion of this analysis is mandatory.

VI. CONCLUSION

Incident rates of melanoma skin cancer have been rising since last two decades. So, early, fast and effective detection of skin cancer is paramount importance. If detected at an early stage, skin has one of the highest cure rates, and the most cases, the treatment is quite simple and involves excision of the lesion. Moreover, at an early stage, skin cancer is very economical to treat, while at a late stage, cancerous lesions usually result in near fatal consequences and extremely high costs associated with the necessary treatments.

When a skin lesion is suspected as melanoma, it must go through all four analyses. If the suspected skin lesion go through only the three of these, it might show erroneous results about its being melanoma or not. For this reason, all the four measures have to be considered to decide whether a skin lesion is melanoma or not.

After all, the best way to lower the risk of melanoma is to limit the exposure to strong sunlight and other source of Ultraviolet light. Take care of all the necessary measures such

as: protecting skin with clothing, wearing hat, using sunscreen, staying in the shade (etc). Moreover, always stay alert about skin and do monthly skin-self exams to reduce the chance of getting any skin cancer which is a risk to human life.

References

- [1] Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* 100:57–70.
- [2] Hahn WC, Weinberg RA. 2002. Modeling the molecular circuitry of cancer. *Nat. Rev. Cancer* 2:331–41.
- [3] Liotta L, Petricoin E. 2000. Molecular profiling of human cancer. *Nat. Rev. Genet.* 1:48–56.
- [4] Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. 2002. Clinical proteomics: translating benchside promise into bedside reality. *Nat. Rev. Drug Discov.* 1:683–95.
- [5] Menon U, Jacobs IJ. 2000. Recent developments in ovarian cancer screening. *Curr. Opin. Obstet. Gynecol.* 12:39–42.
- [6] A.W.Kopf, T.G. Salopek, J. Slade, A.A. Marghood, R.S. Bart, Techniques of cutaneous examination for the detection of skin cancer, *Cancer Supplement* 75 (2) (1994) 684-690.
- [7] I. Papamichail, N. Nikolaidis, I. Glava, K. Lentzas, Marmagkiolis, et al. Merkel cell carcinoma of the upper extremity: case report and an update, *J. Surg. Oncol.* 6 (32) (2008).
- [8] M.E. Celebi, H.A. Kingravi, B. Uddin, et al., A methodological approach to the classification of dermoscopy images, *Comput. Med. Imaging Graph.* 31(2007) 362-373.
- [9] N. Otsu. "A Threshold Selection Method from Gray Level Histogram." *IEEE Trans SMC*, vol.9, pp 62-66, 1979.
- [10] C.Yu and J.Prince, "Snakes, shapes, and gradient vector flow." *IEEE Trans. Image Process.*, vol.7, no.3, pp.359-369, Mar.1998.
- [11] H C Chen et al, Visible color difference-based quantitative evaluation of color segmentation. *IEEE Proceeding, Vis image signal process* vol.153 No.5 Oct 2006 pp 598-609.
- [12] F. Nachbar, W. Stolz, T. Merkle, A.B. Coggnetta, T. Vogt, M. Landthaler, P.Bilck, O. Braun-Falco, and G.Plewig, " The ABCD rule of dermatoscopy: High Prospective value in the diagnosis of doubtful melanocytic skin lesion," *J.Amer.Acas.Dermatol.*, vol. 30, no.4, pp.551-559, Apr.1994.
- [13] R.J Friedman and D.S Riegel, "The clinical features of malignant melanoma," *Dermatologic Clin.*, vol. 3, pp. 297-307, July 1982.