ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology
National Conference on Emerging Trends in Engineering and Technology (NCETET'16)
Lourdes Matha College of Science & Technology, Thiruvananthapuram
Vol. 3, Special Issue 3, August 2016

A Modified Method for Segmentation of Digital Skin Lesion Images

Revathi V.L[1], Chithra A.S[2]

M.Tech Student, Dept of CSE, Lourdes Matha College of Science and Technology, Trivandrum, India[1]

Associate Professor, Dept of CSE, Lourdes Matha College of Science and Technology, Trivandrum, India[2]

Abstract: Skin cancer is the most dangerous form of cancer. Images of melanoma taken from dermatoscope. Dermoscopy images have great potential in the early diagnostic of malignant melanoma, which is a type of skin cancer, but their interpretation is time consuming and subjective, even for trained dermatologists. Automatic lesion segmentation is an important part of computer-based image analysis of pigmented skin lesions. Melanoma is the deadliest form of skin cancer if left untreated. There is a need for an automated system to assess a patient's risk of melanoma using photographs of their skin lesions. In this project, for improving the segmentation accuracy, a segmentation algorithm based on Probabilistic Fuzzy C Means clustering (PFCM) is proposed. PFCM produces memberships and possibilities simultaneously, along with the cluster centers for each cluster. This approach combines the spatial probabilistic information and the fuzzy membership function in the clustering process. The proposed probabilistic fuzzy c-means method can deal effectively with image segmentation in a noisy environment. Next, regions in the image are classified as normal skin or lesion based on the occurrence of representative texture distributions.

Keywords: Skin cancer, Image segmentation, Dermatoscope, PFCM.

I. INTRODUCTION

Skin cancer is the most common cancer in the United States. There are two types of skin cancer: malignant melanoma which is less common but more serious; and non-melanoma skin cancer, which is very common but not so serious. Malignant melanoma on its own can sometimes be referred to as 'skin cancer'. Malignant melanoma is a type of skin cancer which is less common and most dangerous form of skin cancer. Malignant melanoma accounts for 75 percent of all deaths associated with skin cancer in United States [2]. A mole (nevus) is a benign skin tumour that develops from melanocytes which are found in uppermost skin layer epidermis. These skin cells make a brown pigment called melanin. Melanin gives the skin its tan or brown colour. Melanoma can originate in any part of the body that contains melanocytes. Visual analysis of a melanocytic lesion is segmentation of all points in the image as part of the lesion. A skin lesion is a part of the skin that has an abnormal growth or appearance compared to skin around it. Dermatoscope is a device used to take the skin lesion images. Images taken from dermatoscope are called dermoscopic iages. If melanoma detects earlier then almost at curable stage otherwise death occur.

Currently available digital dermoscopic systems offer the possibility of computer storage and retrieval of dermoscopic images. Some systems even display the potential for Computer Assisted Diagnosis (CAD). As diagnostic accuracy with dermoscopy has been shown to depend on the experience of the dermatologist, CAD systems will help less-experienced dermatologists and provides a lower impact for inter-subject variability. The standard approach in automatic dermoscopic image analysis has usually three stages: 1) image segmentation; 2) feature extraction and feature selection; and 3) lesion classification. The segmentation stage is one of the most important because accuracy is the main characteristics of segmentation. However, segmentation is difficult because of the variation of lesion shapes, sizes, and colors and also with different skin types and textures.

In addition to that, some lesions have irregular boundaries and in some cases there is a smooth transition between the lesion and the skin. In automated diagnosis of skin lesions, feature design is based on the ABCD rule of dermatoscopy. ABCD represent the asymmetry, border structure, variegated color, and dermatoscopical structures and define the basis for a diagnosis by dermatologist.

To address the segmentation problem, several algorithms have been proposed. They can be broadly classified as thresholding, edge based or region-based methods. In thresholding method a fusion of global thresholding, adaptive thresholding, and clustering is used. Thresholding methods can achieve good results when there is good contrast between the lesion and the skins, thus the corresponding image histogram is bimodal, but usually fail when the modes from the two regions overlap. Edge-based approaches perform poorly when the boundaries are not well defined, for instance when the transition between skin and lesion is smooth. In these situations, the edges have gaps and the contour may leak through them. Another

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
National Conference on Emerging Trends in Engineering and Technology (NCETET'16)
Lourdes Matha College of Science & Technology, Thiruvananthapuram
Vol. 3, Special Issue 3, August 2016

difficulty is the presence of spurious edge points that do not belong to the lesion boundary. They are the result of artifacts such as hair, specular reflections or even irregularities in the skin texture and they may stop the contour preventing it to converge to the lesion boundary. Region-based approaches have also been used. Region-based approaches have difficulties when the lesion or the skin region are textured or have different colors present, which leads to oversegmentation. From these type of classification of image segmentation includes different segmentation techniques.

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of the image segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to evaluate. It is characteristically used to find objects and image boundaries (lines, curves, etc.) in images and processed of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. It is states to the splitting of an image into separate regions that are identical with respect to luminance, color, texture etc., and techniques can be categorized in to Histogram thresholding, clustering, Edge based detection; Region based detection, morphological detection, active contours etc. Texture based segmentation algorithms have been applied to dermoscopy images. Each classification contain different methods of segmentation.

Most of the segmentation algorithm defines only segmentation techniques not the classification methods. Proposed method includes a most accurate segmentation technique such as Probabilistic Fuzzy C-means clustering (PFCM) and a classifier used is Kernel sparse based representation classifier (KSRSC).

A. Edge-Based Segmentation

Edge-base or boundary-based segmentation methods [3] commonly refer to segmenting an image based on the edges among regions, by searching for edge pixels and connect them to form image contours. However for applying such methods two approaches are founded; manually, by using the mouse to draw lines that represent image boundaries among regions, and automatically, by implementing some edge detection filters, where the pixels classified into edge or non-edge according to the filter output result. Examples of some edge detection filters are: Laplacian of Gaussian filter, watershed segmentation algorithm [4].

B. Region-Based Segmentation

Region-based segmentation partitioned an image into regions or groups of similar pixel depending on some properties [7]. Its principle depends on the idea that neighbouring pixels within the same region have same value [4]. This idea can be implemented by comparing each pixel with its neighbours in a particular region [4], and according to the crucial similarity condition the pixel is decided to belong to a specific region [4]. In the segmentation process feature image is used instead of original input image, the feature image is represented with small neighbourhoods which forms a regions [8]. Region-based segmentation technique required the use of proper thresholding methods [6], and the noise has great impact on the output result [4]. Some skin classification methods based region based are: region growing, region splitting, region merging, split and merge, Neural Networks (NNs).

C. Threshold Based Segmentation

Point based or pixel based segmentation, also known as thresholding. It is the simplest approach for segmenting images, depending on gray-level values to segment image pixels. For skin color classification several algorithms have been suggested, which include piecewise linear classifiers, histogram based thresholding, Neural Networks (NNs).

II. EXISTING SYSTEM

Most of the segmentation algorithms for dermatological images or photographs use color information, either in a single channel or across three color channels, to find the lesion. Another approach to find skin lesions is to incorporate textural information, because normal skin and lesion areas have different textures. Textures include smoothness, roughness, or the presence of ridges, bumps or other deformations and are visible by variation in pixel intensities in an area. Features and measurements of a texture in an image are extracted and textures from different regions are compared. Existing system includes a segmentation algorithm based on texture distinctiveness (TD) to locate skin lesions in photographs. This algorithm is referred to as the TD lesion segmentation (TDLS) algorithm. The main contributions are the introduction of a joint statistical TD metric and a texture-based region classification algorithm. TD captures the dissimilarity between learned representative texture distributions.

A. Joint Statistical Texture Distinctiveness

This method is a texture based segmentation algorithm. This segmentation algorithm based on texture distinctiveness (TD) to locate skin lesions in photographs. This algorithm is referred to as the TD lesion segmentation (TDLS) algorithm. The main contributions are the introduction of a joint statistical TD metric and a texture-based Region classification algorithm.TD captures the dissimilarity between learned representative texture distributions. This region classification algorithm incorporates the texture information captured by the TD metric [1].

The TDLS algorithm mainly consists of two main steps. First, a set of sparse texture distributions that represent skin and lesion textures are learned. A TD metric is calculated to measure the dissimilarity of a texture distribution from all other texture distributions. Second, the TD metric is used to classify regions in the image as part of the skin class or lesion class [1].

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
National Conference on Emerging Trends in Engineering and Technology (NCETET'16)
Lourdes Matha College of Science & Technology, Thiruvananthapuram
Vol. 3, Special Issue 3, August 2016

The grouping of lesion features are done by using k-means clustering. K-means clustering is used as an initial step to increase the robustness and to speed up the number of iterations required for the finite mixture model to converge. K-means clustering finds K clusters of texture data points that minimize the sum of squared error between cluster members and the cluster mean. To improve clustering efficiency Expectation and Maximization algorithm is used. Then use the SRM (Statistical Region Merging) method to extract cluster features. SRM contains two main steps: a sorting step and a merging step.

SRM sorts pixels in an image to determine the order in which pixels are compared, and then merges pairs of pixels into regions based on their similarity. To maximize lesion region adaptive thresholding technique is used. After the regions are classified as being normal skin or lesion, postprocessing steps are applied to refine the lesion border such as morphological dilation and region selection. First, the morphological dilation operator is applied to fill holes and smooth the border. Morphological dilation is a process that expands binary masks to fill small holes.

Morphological dilation to remove non lesional area from the image. Second, since multiple noncontiguous regions may have been identified as part of the lesion class, the number of regions is reduced to one. While it is possible to have multiple lesions in a single image, it is necessary to reduce the number of lesions for the feature extraction step. To eliminate the small regions, the number of pixels in each contiguous region is counted. The contiguous region with the largest number of pixels is assumed to correspond to the lesion class and any other regions are converted to the normal skin class. This gives the final lesion segmentation. Because of the use of k-means clustering center point of each cluster to be calculated. This will leads to lagging to find clusters.

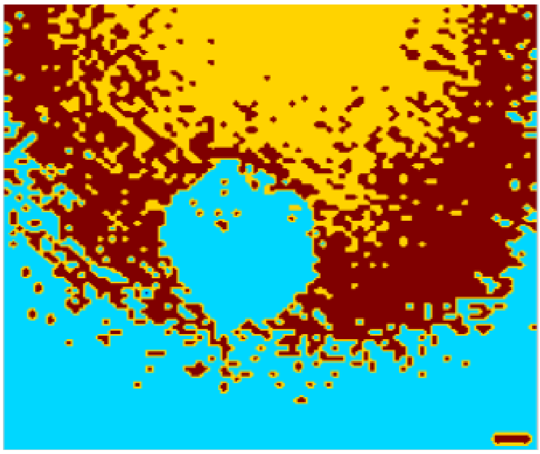

Fig.1 K-means clustered image

Fig 1 Shows the image after k-means clustering

III. PROBABILISTIC FUZZY C-MEANS CLUSTERING

The proposed method include TDLS algorithm with Probabilistic Fuzzy c-means clustering instead of k-means clustering. Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics and data compression. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions.

Clustering can be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties.
The term clustering, there are a number of terms with similar meanings, including automatic classification, numerical taxonomy, botryology (from Greek βότρυς "grape") and typological analysis. The main differences are often in the usage of the results: while in data mining, the resulting groups are the matter of interest, in automatic classification the resulting discriminative power is of interest. This leads to misunderstandings between researchers coming from the fields of data mining and machine learning, since they use the same terms and often the same algorithms, but have different goals.

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to one cluster. Thus, points on the edge of a cluster, that is in the cluster to a lesser degree than points in the center of cluster. Any point x has a set of coefficients giving the degree of being in the k^{th} cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster [9][10].

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
National Conference on Emerging Trends in Engineering and Technology (NCETET'16)
Lourdes Matha College of Science & Technology, Thiruvananthapuram
Vol. 3, Special Issue 3, August 2016

The degree of belonging, w_k (x), is related inverse to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. The fuzzy c-means algorithm is similar to the k-means algorithm:

- Choose a number of clusters.
- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε, the given sensitivity threshold) :
  - Compute the centroid for each cluster, using the formula above.
  - For each point, compute its coefficients of being in the clusters, using the formula above.

Using a mixture of Gaussians along with the expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes. Fuzzy c-means has been a very important tool for image processing in clustering objects in an image. FCM algorithm to improve the accuracy of clustering under noise.

A. Probabilistic Fuzzy C-means clustering:

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is commonly used in pattern recognition. This method is based on minimization of the following objective function:

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

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

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

This iteration will stop when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k is 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 $U^{(k)}$, $U^{(k+1)}$

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

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

Advantages

1) Gives best result for overlapped data set and comparatively better than k-means algorithm.
2) Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center. The image after probabilistic fuzzy c-means clustering as shown in Fig.2.

B. Summary of proposed TDLS algorithm

1) Convert the corrected image to the XYZ color space [1].
2) For each pixels in image I , extract the texture vector t_s to obtain the set of texture vectors T (1).
3) Cluster the texture vectors in T, with PFCM clustering, to obtain the representative texture distributions.
4) Calculate the probability that two texture distributions are distinct $d_{j,k}$ for all possible pairs of texture distributions.
5) Calculate the textural distinctiveness metric D_j
(7) For each texture distribution.
6) Apply the SRM algorithm to find the initial regions.
7) Calculate the region distinctiveness metric $D_R$for eachinitial region.
8) Calculate the threshold τ between the normal skin andlesion classes.
9) Classify each region as normal skin or lesion based on the results of steps 7 and 8.
10) Apply a morphological dilation operator to the initial lesion classification.

11) For each contiguous region in the initial segmentation, count the number of pixels in the region.
12) As the final lesion segmentation, return the contiguous region consisting of the most pixels.

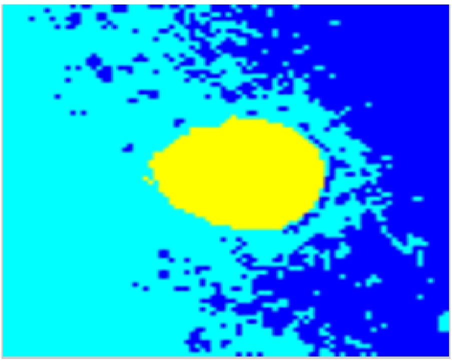

Fig. 2 Fuzzy C-means clustered image

V. CONCLUSION

Segmentation is the classification of the input image into skin and non-skin pixels based on skin texture. Classification also have very importance in segmentation of melanocytic skin lesion from digital images. Proposed a segmentation algorithm based on Probabilistic Fuzzy C Means clustering (PFCM). The proposed PFCM produces memberships and possibilities simultaneously, along with the cluster centers for each cluster. Also the proposed fuzzy clustering provides fuzzy membership function in the clustering process. The proposed fuzzy clustering algorithm gives higher segmentation accuracy when compared with other techniques.

REFERENCES

[1] Jeffrey Glaister, Alexander Wong, and David A. Clausi, "Segmentation of Skin Lesions From Digital Images Using Joint Statistical Texture Distinctiveness", IEEE Trans. on Biomedical Engg., vol. 61, no. 4, April 2014.
[2] D.S. Rigel, R. J. Friedman, and A. W. Kopf, "The incidence of malignant melanoma in the United States: Issues as we approach the 21st century," J. Amer. Acad. Dermatol., vol. 34, no. 5, pp. 839–847, 1996.
[3] Hassana Grema Kaganami, Zou Beiji, "Region-Based Segmentation versus Edge Detection", IEEE Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1217-1221, 2009. DOI: 10.1109/IIH-MSP.2009.13
[4] Yu-Hsiang Wang, "Tutorial: Image Segmentation", Graduate Institute of Communication Engineering National Taiwan University, Taipei, Taiwan. Available: http://disp.ee.ntu.edu.tw/ meeting/%E6%98%B1%E7%BF%94/Segmentation%20tutorial.pdf
[5] Rafiqul Zaman khan, Noor Adnan Ibraheem, "Survey on Gesture Recognition for Hand Image Postures", Canadian Center of Computer and Information Science, Vol. 5(3), pp. 110-121, May 2012. doi:10.5539/cis.v5n3p110.
[6] Hassana Grema Kaganami, Zou Beiji, "Region-Based Segmentation versus Edge Detection", IEEE Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1217-1221, 2009. DOI: 10.1109/IIH-MSP.2009.13
[7] http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/ RAMANI1/node19.html
[8] Krishna Kant Singh, Akansha Singh, "A Study Of Image Segmentation Algorithms For Different Types Of Images", International Journal of Computer Science Issues IJCSI, Vol. 7(5), September 2010.
[9] Soumi Ghosh, Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms ", ((IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013
[10] V. S. Rao and Dr. S. Vidyavathi, "Comparative Investigations and Performance Analysis of FCM and MFPCM Algorithms on Iris data", Indian Journal of Computer Science and Engineering, vol.1, no.2, 2010 pp. 145-151.