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

Classification of Skin Lesion Images Using Kernel Classifier

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 in medical science. Dermatoscope is a device which is used to capture the melanocytic skin lesion images. Dermoscopy images have great potential in the early diagnostic of malignant melanoma. Automatic skin lesion segmentation is an important part of computer-based image. 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. Proposed work is used for improving the segmentation accuracy and classification. Segmentation of skin lesion image is done on the basis of a k-means clustering. Classification as melanoma or others based on a kernel sparse based representation classifier.

Keywords: Skin cancer, Image segmentation, Dermatoscope, Kernel sparse classifier

I. INTRODUCTION

A skin lesion is a part of the skin. Skin lesion is an abnormal growth around skin. 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. Nearly all moles are harmless. Melanoma is a type of skin cancer which is less common than basal and squamous skin cancer but it is one of the most aggressive types of cancer[2]. It can be healed by surgical excision; if recognized in the early stage. So it becomes very important and useful to develop an automated melanoma recognition system. The skin lesion images can be taken from a device called, dermatoscope. Images from that device is called dermoscopy images.

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 typical architecture of CAD system includes selection of training samples, image pre-processing, segmentation, feature extraction and classification. The aim of the pre-processing step is to eliminate the background noise and improve the image quality for the purpose of determining the focal areas in the image. Image segmentation is an important step in image analysis, pattern recognition, and computer vision. An accurate segmentation of skin images can help the diagnosis to define well the region of cancer. Fig 1. Shows an example of skin lesion image [12].

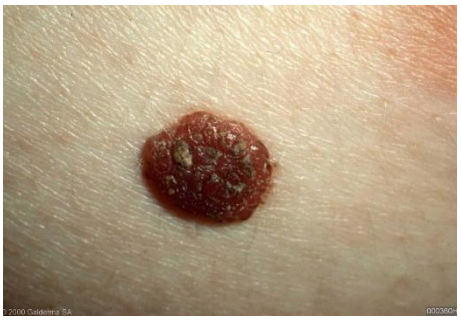

Fig. 1. Skin lesion image

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.Image Segmentation is a vital procedure of processing and understanding an image. It is the fundamental necessity of any computer based vision application on the grounds that individuals are mostly interested just in specific parts of the picture. Essentially it is characterized as the procedure of isolating the picture into distinctive parts of homogeneity.The standard approach in automatic dermoscopic image analysis has usually three stages such as, image segmentation, feature extraction and feature selection; and finally the 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.

The segmentation problem can be overcome through different segmentation algorithms. They can be broadly classified as thresholding, edge based or region-based

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

methods [6]. 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 skin, 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 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 over segmentation. From these type of classification of image segmentation includes different segmentation techniques.

The main 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 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 segmentation technique with the use of a classifier also. The classifier used is Kernel sparse based representation classifier (KSRSC).

II. RELATED WORK

There many algorithms are used for the segmentation of digital skin lesion images. Some systems only define the segmentation technique only. Segmentation using joint statistical texture distinctiveness [1] method only define the segmentation. So the classification has greater importance in skin cancer detection.

III. PROPOSED SYSTEM

The TDLS algorithm consists of two main steps [1]. 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. In this section, the first step is described in detail illustrates the overall process to learn the representative texture distributions and calculate the TD metric.

Existing sparse texture algorithms use sparse texture models for segmentation or classification of images with different texture patterns. Sparse texture models find a small number of texture representations, such as texture patches, to characterize an entire image. Sparse texture models learn important local texture details present in an image. Using a sparse texture model allows the image to be stored efficiently and allows for efficient computation of algorithms that involve textures from the image. There are many ways to learn the model, including clustering or by formulating the problem as an optimization problem. A common method to learn a sparse texture model is by employing a dictionary-learning algorithm, where a set of texture patches that can best match details in the original image is learned. To learn whether each texture distribution belongs to the skin or lesion class, a TD metric is formulated.

A. TD Metric

A TD metric is formulated using the learned sparse texture model. Since we are only interested in two classes, normal skin and lesion, but have learned many texture distributions, multiple texture distributions must represent the same class. To measure similarity of two texture distributions, we first measure the probability that the mean of one texture distribution is a realization of the mean of the other texture distribution, which is defined as $l_{j,k}$. Because we assume that the texture distributions are Gaussian, $t^r_j$and $\Sigma_j$are the mean and covariance of distribution T^r_j. The metric $l_{j,k}$is asymmetric, because when comparing most pairs of distributions, $\Sigma_i \neq \Sigma_j$. The measure of similarity $L_{j,k}$given the average of $l_{j,k}$and $l_{k,j}$. After $L_{j,k}$has been calculated for each pair of texture distributions, they are normalized to be between 0 and 1,

$$l_{j,k} = 1/(\sqrt{(2\pi)^{n \times n \times a}|\Sigma_j|}) \exp(-(1/2(t^r_j - t^r_k)^T \Sigma_j (t^r_j - t^r_k)))$$

$$L_{j,k} = \tfrac{1}{2}(l_{j,k} + l_{j,k})$$

They are interested in finding distinct texture distributions. For example, lesion texture distributions are both dissimilar from the normal skin texture distributions and also from other texture distributions, due to color variegation and textural patterns found in skin lesions. The probability that a texture distribution is distinct from another texture distribution is given by $d_{j,k}$:

$$d_{j,k} = 1 - L_{j,k}$$

Using the texture distributions and probabilities of distinctiveness, a weighted graphical model can be constructed to characterize all pair-wise relationships. The graphical model is defined as G = {V, E }. V represents the set of vertices for the graphical model, which are the texture distributions associated with each pixel in the

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

image. E represents the set of edges between every pair of texture distributions, which are given a weight based on the probability of distinctiveness, $d_{j,k}$.

A TD metric $D_j$is used to capture the dissimilarity of texture distribution $T^r_j$from other texture distributions. The metric is defined and measures the expected distinctiveness of T^r_jgiven the photograph I , where P (T^r_k |I) is the probability of occurrence of a pixel being associated with a texture distribution T^r_k. P (T^r_k|I) is estimated using the histogram of the number of pixels associated with each texture distribution across the entire image,

$Dj = \Sigma^k_{k=1} d_{j,k} P(T^r_k|I)$

In the case of normal skin texture distributions, the dissimilarity of one skin texture distribution from other skin texture distributions is very small. The TD metric for skin texture distributions is small overall. Lesion texture distributions are dis-similar from other skin and lesion texture distributions, so the textural distinctiveness metric is large.

A brighter pixel corresponds to a higher TD metric. In both figures, the lesion is predominately white, meaning that the lesion texture distributions have higher TD metrics, as expected. In, there are two texture distributions that correspond to the lesion class and have high TD. However, in, some normal skin pixels to the right of the lesion also have high TD. This can occur when there are unique texture patterns in normal skin areas. This commonly occurs, motivating the region classification step of the TDLS algorithm. The region classification step allows the algorithm to be more robust and minimize misclassification of pixels.

B. Region Classification

The second main step in the TDLS algorithm is to find and classify regions in the input image as being part of the lesion based on the sparse texture distributions and their associated TD metric. First, the image is over segmented, which results in the image being divided into a large number of regions. Next, each region is independently classified as representing normal skin or lesion based on the textural contents of that region. Finally, postprocessing steps refine the lesion segmentation.

1. Initial Regions

The corrected lesion image is divided into a large number of regions. This initial over segmentation step is incorporated to increase the TDLS algorithm's robustness to noise. Further-more, it allows for the use of an efficient and fast classification algorithm to find which regions belong to the skin or lesion class [9]. The initial over segmentation algorithm is adapted from the statistical region merging (SRM) algorithm. The main difference is that the SRM algorithm uses the image in the RGB color space, while the TDLS algorithm converts the photograph to the XYZ color space, as mentioned. The advantages of using the SRM algorithm as the initial over segmentation algorithm are that it directly takes into account pixel location, is simple and is computationally efficient.

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. A four-connected graph is constructed so that each pixel in the photograph is connected with its neighbors. The pixels are sorted based on their similarity with their neighbouring pixel. Both horizontal and vertical neighbouring pixels are considered when sorting the pixels. The merging predicate determines whether two regions are merged together, based on pixel intensities. The predicate depends on the difference between average pixel intensity for each channel for the two regions. Furthermore, it depends on the number of pixels in the regions. It includes a tunable parameter Q to change the likelihood that two regions are merged. The parameter Q is set to 128 following experimental testing. Additional details are available in the paper by Nock and Nielson.

The result of the initial over segmentation step is a map of several regions which correspond to the normal skin or lesion classes. To reduce the number of regions, all segments that touch the edges of the photograph are merged into a single region. This is based on the assumption that the lesion is not touching the edges of the photograph, which is reasonable for situations where the photographs are captured in controlled, clinical environments. As such, regions touching the edges are all likely to be part of the normal skin class.

2. Distinctiveness-based Segment Classification

Following the initial over segmentation step, each region must be classified as belonging to the normal skin class or lesion class based on a criterion. The classification step is illustrated, where y is the resulting segmentation map. Each element in y is either 1 (lesion) or 0 (normal skin), depending on the classification results for that element's corresponding region. The threshold is denoted by τ and it represents the decision boundary between the normal skin and lesion class. The feature used to discriminate between the two classes is the regional textural distinctiveness metric D_R. This metric is based on the TD across a region,

$Y(R) = \{1, \text{if } D_R >= \tau \text{ (lesion)}$
$\{0, \text{otherwise (normal skin)}$

A TD metric D is calculated for each texture distribution based on the probability of it being similar to other texture distributions. This information is combined with the contents of each region to determine a regional TD metric, D_R. D_R represents the average TD across region R, where P (T^r_j|R) is the probability of a pixel being associated with the j^{th} texture distribution in region R .Again, P (T^r_j|R) is estimated using the histogram of the number of pixels associated with each texture distribution across the region R,

$D_R = \Sigma^K_{j=1} D_j P(T^r_j|R)$

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

Finally, a threshold τ is defined to divide the set of representative texture distributions into two classes, normal skin and lesion, and is also based on the TD metrics. There are many ways to find two classes from a one-dimensional set of features.

3. Segmentation Refinement

After the regions are classified as being normal skin or lesion, the following post processing steps are applied to refine the lesion border: 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. The shape and amount that the binary mask is expanded is controlled by a structuring element, which is a disc with a radius of 5 pixels in the TDLS algorithm. Next, 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.

C. Kernel sparse representation based classifier

Sparse representation-based classifier (SRC), a combined result of machine learning and compressed sensing (CS), shows good classification performance on face image data.SRC could not well classify the data with the same direction distribution. The same direction distribution means that the sample vectors belonging to different classes distribute on the same vector direction. Kernel sparse representation- based classifier (KSRC), based on SRC and the kernel trick which is a usual technique in machine learning. KSRC is a nonlinear extension of SRC and can solve the drawback of SRC [10].

In KSRC the data in the input sample space are implicitly mapped into a high or even infinite dimensional kernel feature space by using some nonlinear mapping associated with a kernel function. Since we can only access the kernel feature space in terms of the kernel function, the kernel-based dimensionality reduction method is required to reduce the dimensionality of the feature space.

Kernel approach can change the distribution of samples by mapping samples into a high dimensional feature space. This change possibly has two effects if an appropriate kernel function is selected. On the one hand, some linear inseparable samples in the original feature space become linear separable in the high dimensional feature space. This leads to superiority of the KERNEL-NN classifier over the NN classifier. On the other hand, a test sample can be represented as the linear combination of training samples from the same class as itself more accurately in the high dimensional feature space than original. Then the nonzero entries of sparse representation coefficient vector of the test sample are more associated with training samples of the same class. This results in better classification ability of sparse representation based classifier (SRC) [11].

Algorithm

1. Input: the matrix of training samples $A \varepsilon R^{mxn}$, a test sample $y \varepsilon R^{m}$ and a kernel function.
2. Normalize the columns of A to have unit l2-norm.
3. Calculate $B^{T}B$ and $B^{T}\phi(y)$.
4. Solve the minimization problem.
5. Compute the residuals $R_k(y)$ (k=1,2,. . .,c) .
6. Output: identity (y)= arg_k min ($R_k(y)$).

Kernel sparse based representation classifier give better accuracy than other logistic classifier.

IV. 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 statistical texture distinctiveness and with a kernel classifier.Kernel sparse representation based classifier is used for the classification of different types of skin cancer as, melanoma or benign. It also helps to improve the segmentation accuracy of the system.

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] http://www.cancer.org/cancer/skincancer-melanoma/overviewguide/melanoma-skin-cancer-overview-what-is-melanoma
[3] Hassana Grema Kaganami, ZouBeiji, "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] RafiqulZaman 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] HassanaGremaKaganami, ZouBeiji, "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] Li Zhang, Member,, We-Da Zhou, Pei-Chann Chang, Jing Liu, Zhe Yan, Ting Wang, and Fan-Zhang Li., "Kernel sparse representation based classifier", IEEE Transactions on Signal processing, Vol. 60, noO. 4, April 2012.
[11] Jun Yin, Zhonghua Liu, Zhong Jin, Wankou Yang, "Kernel sparse representation based classification", Neurocomputing 77 (2012) 120–128.
[12] Derm Quest, (2012). [Online]. Available: http://www.dermquest.com