ISSN 2348 – 8034
Impact Factor- 3.155

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

REVIEW ON: SEGMENTATION AND CLASSIFICATION OF SPUTUM CELL FOR LUNG CANCER DETECTION

Money Sethi[*1] and Sukhdeep Kaur[2]
[*1]Adesh Institute of Engineering & Technology
[2]Assistant Professor, Adesh Institute of Engineering & Technology

ABSTRACT

Diagnosis of lung cancer in its primal stage is a major issue confronted by the medical world. For that proper points of interest are required from the images which must be got by a decent segmentation method. However, many common forms of techniques are available in market and all we need is the accuracy of segmentation of the nucleus from the ROI and also the time consumed for the same. In this paper, we will discuss the segmentation and various classification techniques which help in lung cancer detection. The classification techniques differentiate between the cells and background leading to cell lung cancer detection. Our method demonstrates on sputum cytology images. For the implementation of this proposed work we use Image Processing Toolbox under the MATLAB software.

Keywords: *Lung Cancer Detection, Segmentation, Feature Extraction and Classification.*

I. INTRODUCTION

1.1 Image Segmentation

Image segmentation is the procedure of dividing a digital image into multiple segments. The objective of segmentation is to improve and change the representation of an image into something that is more important and easier to investigate. Image segmentation is regularly used to find objects and boundaries in images. All the more exactly image segmentation is the procedure of allocating a label to each pixel in an image such that pixels with the same label share certain qualities. There are following types of image segmentation:

- **Threshold based segmentation:** Histogram thresholding and slicing procedures are utilized to segment the image. They may be connected directly to an image but can likewise be consolidated with pre-processing and post-processing systems.
- **Edge based segmentation:** With this system recognized edges in an image are accepted to represent object boundaries and used to recognize these items.
- **Region based segmentation:** Where an edge based system may endeavour to discover the object boundaries and then find the object itself by filling them in a region based system takes the opposite method by starting in the middle of an item and then "growing" outward until it meets the object boundaries.
- **Clustering techniques:** Although clustering is sometimes utilized as an equivalent word for segmentation systems and we utilize it here to signify procedures that are essentially utilized as a part of exploratory information examination of high-dimensional estimation designs. In this connection grouping routines endeavour to gathering together examples that are comparable in some sense. This objective is very much alike to what we are endeavoring to do when we section an image and without a doubt some clustering methods can promptly be sought image segmentation.
- **Matching:** When we recognize what an item we wish to distinguish in an image looks like we can utilize this learning to find the item in an image. This way to deal with segmentation is called matching.

The consequence of image segmentation is a set of segments that aggregately cover the whole image or an arrangement of shapes extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property such as color, intensity or texture. Image segmentation algorithms are classified into two types that are supervised and unsupervised. Unsupervised algorithms are fully automatic and partition the regions in feature space with high density. The different unsupervised algorithms are Feature-Space Based Techniques, Clustering (K-means algorithm, C-means algorithm, E means algorithm), Histogram thresholding, Image-Domain or Region Based Techniques (Split-and-merge techniques, Region growing techniques, Neural-network based techniques, Edge Detection Technique), Fuzzy Techniques, etc.

ISSN 2348 – 8034
Impact Factor- 3.155

1.2 Lung Cancer

In medical image processing, the detection of various types of cancer affecting different organs is a primary ongoing research area. Lung cancer is one of the most frequent cancers worldwide. Like breast cancer in female, lung cancer is the leading cancer site in males, comprising 17% of the total new cancer cases and 23% of the total cancer deaths. Two major types of Lung cancers are adenocarcinoma (including bronchi alveolar carcinoma) representing about 40% and squamous cell carcinomas representing about 25-30%. Current investigations into early detection and adjuvant chemotherapy heavily rely on the proper staging of patient cancer type. Before computer aided classification, the cell segmentation is a prerequisite. Cancer detection should be more accurate since its reports will be used for diagnosis. So many researchers have provided different techniques. Biopsy is a gold standard for cancer detection including lung cancer. The use of fine needles or biopsy equipments to the affected area is more painful to the patient. Normally, surgery is needed for the diagnosis of lung cancer which will be costly and painful. Another commonly used method for lung cancer detection is the analysis of CT images. But the information obtained from CT has the problem of intensity variation. Hence, there is chance to get fault results and it may lead to wrong diagnosis and other related problems, which harm medical practices. So there is a need of an alternate strategy for lung cancer detection and it led them to the analysis of cell images, which is better than the CT image analysis. For this analysis they used the sputum of the affected person which can be obtained without surgery or any other harmful method.

II. PREVIOUS WORK

Fuyong Xing, Lin Yang "Robust Cell Segmentation for Non-Small Cell Lung Cancer" (IEEE, 2013): This paper proposed a novel and strong touching cell segmentation calculation for non-small cell lung disease which first restricts cell seeds utilizing distance transform-based voting and thereafter employs a repulsive balloon snake model to accurately segment cells with introduction utilizing the detected seeds. GDMSPV algorithm is utilizations and GDMSPV makes more exact localization of cancer cells.

Fatm Taher, Naoufel Werghi and Hussain Al-Ahmad "Extraction of Sputum Cells using Thresholding Techniques for Lung Cancer Detection" (IEEE, 2012): This paper manages an enhanced variant of a sifting thresholding calculation for separating the sputum cell from the crude sputum picture for lung tumor early discovery. In this technique the issue is seen as a division issue concentrating on the extraction of such sputum cells from the pictures. This will be finished by fragmenting the picture into sputum cell district which incorporates the cores, cytoplasm and the foundation that incorporates all the rest. These cells can then be investigated to check whether they are harmful or not.

Fatma Taher and Rachid Sammouda "Lung Cancer Detection by using Artificial Neural Network and Fuzzy Clustering Methods" (IEEE, 2011) : This paper introduces two division routines, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) grouping calculation, for fragmenting sputum shading pictures to identify the lung tumor in its initial stages. It was found that the HNN division results are more precise and dependable than FCM grouping in all cases. The HNN succeeded in removing the cores and cytoplasm areas. However FCM fizzled in recognizing the cores, rather it recognized just piece of it. Notwithstanding that, the FCM is not delicate to power varieties as the division lapse at union is bigger with FCM contrasted with that with HNN. The HNN will be utilized as a premise for a Computer Aided Diagnosis (CAD) framework for right on time location of lung malignancy. Later on, we plan to consider a Bayesian choice hypothesis for the identification of the lung tumor cells, trailed by adding to a model taking into account the thought of watershed calculation which joined the thought of edge location and area based way to deal with concentrate the homogeneous tissues spoke to in the picture.

Xiong Wang, Daniel R. Bauer, Russell Witte and Hao Xin "Microwave-Induced Thermo acoustic Imaging Model for Potential Breast Cancer Detection" (IEEE, 2012): In this study, added to a complete microwave incited thermo acoustic imaging (TAI) model for potential bosom growth imaging application. Acoustic weights produced by diverse bosom tissue targets are examined by limited contrast time-area recreations of the whole TAI procedure including the encouraging radio wire, coordinating instrument, fluidic environment, 3-D bosom model, and acoustic transducer. Recreation results accomplish quantitative connections between the information microwave crest force and the subsequent particular retention rate and also the yield acoustic weight. Microwave recurrence reliance of the acoustic flags because of distinctive bosom tissues is set up over a broadband recurrence run (2.3–12 GHz).

ISSN 2348 – 8034
Impact Factor- 3.155

III. CLASSIFICATION TECHNIQUES

In medical image processing, there are a number of classification techniques used for detection of lung cancer some of them mentioned in this paper are discussed below.

Bayesian classifier: This method was proposed by Naïve Bayes in 1950. In machine learning, this method remained popular for text categorization, the problem of judging document as being to one category or the other. It is competitive in its domain and is more advanced method. A Naive Bayes classifier or Bayesian classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. It is used for judging whether a particular document belongs to one category or the other and we can categorize object belongs to one category or the other. For example a fruit may be considered as an apple of it is red, round and is of 4 inches in diameter. Even if these features depend upon each other or upon the existence of other features the Naïve Bayers classifier will consider all these properties to contribute independently to the probability that this fruit is an apple. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. The Naïve Bayes classification has worked well in many complex real-world problems. In 2004, analysis of the Bayesian classification problem has shown that there are some theoretical reasons for unreasonable efficiency of naive Bayes classifiers. Still, a comprehensive comparison with other classification methods in 2006 showed that Bayes classification is outperformed by more current approaches. It only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

Support vector machine: The SVM algorithm was invented by Vladimir N. Vapnik in 1963. SVM are supervised learning (machine learning which includes construction and study of algorithms that can learn from and make predictions on data) model with associated learning algorithm that analyze data and recognize pattern used for classification. If we have a lots of training examples and each example belongs to one of the two categories. SVM training algorithm builds a model which will determine whether the given example belongs to which one of the two categories. It represents example as points in spaces so that examples of separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into same space and predicted to belong to a category based on which side of the gap they fall on. They are suitable for binary classification tasks. More formally, SVM constructs hyper plane or a set of hyper-planes which can be used for classification. A good separation is achieved by the hyper-plane that has largest distance to the nearest training data points of any class. They are used to solve the real world problems. They are helpful in text and hypertext categorization. Classification of images can be performed by using SVM. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes

Artificial Neural Network: Artificial neural networks are composed of interconnecting artificial neurons and artificial neural networks may either be used to gain an understanding of biological neural networks or for solving artificial intelligence problems without necessarily creating a model of a system of real biological. It is supervised learning model that is it takes set of inputs and outputs to be trained. The real, biological nervous system is highly complex: artificial neural network algorithms attempt to abstract this complexity and focus on what may hypothetically matter most from an information processing point of view. Good performance (e.g. as measured by good predictive ability or low generalization error) or performance mimicking animal or human error patterns and it can then be used as one source of evidence towards supporting the hypothesis that the abstraction really captured something important from the point of view of information processing in the brain.

Hopfield neural network: Hopfield Neural Network (HNN) is another classifier which is one of the artificial neural networks, which has been proposed for Segmenting both gray-level and color images. If one have the segmentation problem for gray-level images as minimizing a suitable energy function with HNN, it derived the network architecture from the energy function, and classify the sputum cells into nuclei, cytoplasm and background classes.

Fuzzy Technique: It is one of the most difficult tasks in image analysis & computer vision. It is to classify the pixel in an image correctly, when there is no crisp boundaries between objects in an image thus in order to address this difficulty, fuzzy technique is used. Fuzzy technique classifies pixel values with great extent of accuracy & it is basically suitable for decision oriented applications like tissue classification & tumor detection etc. Fuzzy clustering divides the input pixels into clusters or groups on the basis of some similarity criterion, such that similar pixels belong to same cluster.

ISSN 2348 – 8034
Impact Factor- 3.155

IV. CONCLUSION

In this paper, we classify and discuss main image segmentation techniques which have been used for lung cancer detection. These techniques distinguish between the area of interest (cells) and non-area of interest (back-ground). We are planning to create a work which will include a segmentation technique which will recognize some of the cells in the image. This segmentation technique may not determine all the cancer cells in the image but will help in training a classifier which will recognize all of the remaining cancer cells present in the image leading to an efficient cancer detection method. The results will be compared with the results obtained by one of the above image classifier such as Bayesian classifier and Support Vector Machine.

REFERENCES

1. *Xing, Fuyong, and Lin Yang. "Robust cell segmentation for non-small cell lung cancer." In Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, pp. 386-389. IEEE, 2013.*

2. *Taher, Fatma, NaoufelWerghi, and Hussain Al-Ahmad. "Segmentation of sputum color image for lung cancer diagnosis based on mean shift algorithm." In AFRICON, 2013, pp. 1-6. IEEE, 2013.*

3. *Fatma Taher et al.,"Automatic Sputum Color Image Segmentation forLung Cancer Diagnosis", KSII Transactions on Internet and Information Systems Vol. 7, No. 1, Jan 2013*

4. *Taher, Fatm, NaoufelWerghi, and Hussain Al-Ahmad. "Extraction of sputum cells using thresholding techniques for lung cancer detection." In Innovations in Information Technology (IIT), 2012 International Conference on, pp. 36-41. IEEE, 2012.*

5. *Smitha P et al., "A review of medical image classification techniques, "International conference on VLSI, communication & instrumentation (ICVCI) 2011 proceedings published by International journal of computer applications, 34-38 pages.*

6. *A. Jemal, F. Bray, M. M. Center, J. Ferlay, E.Ward, and D. Forman, "Global cancer statistics," CA: Cancer J. Clin., vol. 61, no. 2, pp. 69–90, 2011.*

7. *W. D. Travis, E. Brambilla, M. Noguchi, A. G. Nicholson, K. R. Geisinger, and Y. Yatabe. et al, "International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma," J. Thorac. Oncol., vol. 6, no. 2, pp. 244–285, 2011.*

8. *H. Kong, M. N. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting," TMI, vol. 30, no. 9, pp. 1661–1677, 2011.*

9. *E. Bernardis and S. X. Yu, "Finding dots: Segmentation as popping out regions from boundaries," CVPR, pp. 199–206, 2010.*

10. *Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images." TBME, vol. 57, no. 4, pp. 841–852, 2010.*

11. *S. Issac Niwas et al.,"Complex wavelet as nucleus descriptors for Automated Cancer cytology classifier system using ANN", American Journal of Biomedical Engineering 2010, DOI: 10.5923/j.ajbe.20100203.08.*