ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

Analysis of Lung Nodule Classification with Feature Extraction

M.Jayalaxmi[1], S.Kalpana[2], Y.Sangamitra[3]

PG Student [VLSI], Dept. of ECE, Vivekanandha College of Engg for Women , Namakkal, Tamilnadu, India.

PG Student [AE], Dept. of ECE, Vivekanandha College of Engg for Women , Namakkal, Tamilnadu, India.

Assistant professor, Dept. of ECE, Vivekanandha College of Engg for Women , Namakkal Tamilnadu, India.

ABSTRACT: In this paper, the four types of lung nodules are classified, i.e. well circumscribed, vascularised, juxta-pleural and pleural tail in low dose computed tomography (LDCT) scan. This classifier analyses both lung nodule and surrounding anatomical structures. Also, it consists of three main stages as follows: (1) Multi level concentric partition through patch based image representation, (2) Feature set design for patch description of image, and (3) SVM classifier compute the classification probability based on level nodule and pLSA calculate the classification probability based on level context. The proposed method was evaluated on a publicly available dataset and clearly demonstrated promising classification performance.

KEYWORDS: Classification, patch division, feature design, SVM classifier.

I.INTRODUCTION

Lung cancer is a major cause of cancer-related deaths in humans worldwide. Approximately 20% of cases with lung nodules represent lung cancers. Therefore the identification of potentially malignant lung nodules is essential for the screening and diagnosis of lung cancer. Lung nodules are small masses in the human lung, and are usually spherical; however, they can be distorted by surrounding anatomical structures, such as vessels and the adjacent pleural. Intra-parenchymal lung nodules are more likely to be malignant than those connected with the surrounding structures, and thus lung nodules are divided into different types according to their relative positions. Lung nodules classification has four types: well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularized (V) with the nodule located centrally in the lung but closely connected to neighbouring vessels; juxta-pleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail. Sample images are shown in Fig 1, with the nodule encircled in red.

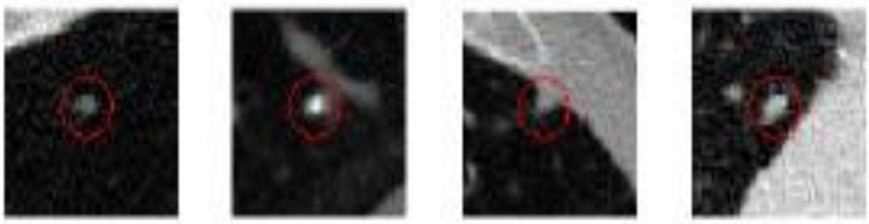

Fig -1: Transaxial CT images with the four types of nodules, shown from left to right, well-circumscribed, vascularized, juxta-pleural, and pleural-tail.

Computed tomography (CT) is the most accurate imaging modality to obtain anatomical information about lung nodules and the surrounding structures. In current clinical practice, however, interpretation of CT images is challenging for radiologists due to the large number of cases. This manual reading can be error-prone and the reader may miss nodules and thus a potential cancer. Computer-aided diagnosis (CAD) systems would be helpful for

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

radiologists by offering initial screening or second opinions to classify lung nodules. CADs provide depiction by automatically computing quantitative measures, and are capable of analyzing the large number of small nodules identified by CT scans.

II.RELATED WORK

While many studies have reported the detection and segmentation of lung nodules, there are limited data in lung nodule classification. Farag et al. reported on some of the initial studies in the classification problem. Here suggest, however, that improved performance could be obtained by better feature design and a more advanced classifier. In recent work, the design of an overlapping nodule identification procedure to help the classification, but this work mainly focused on identifying the nodules located in the intersections among different types. In prior work from our group we suggested that contextual information surrounding the lung nodules could be incorporated to improve nodule classification however, this method required a complicated segmentation process. Contextual information refers to the complicated anatomical structures around the nodules, within which some structures are only present in certain type of nodules, and some are common across more than one type. For example, W and V nodules are similar in location and shape, which makes it difficult to distinguish them, merely based on the nodule information. V nodules are closely connected to the neighboring vessels and W nodules are isolated from other structures, so identifying connected vessels from V nodules provides an important clue to separate these from each other. Contextual patterns are similarly important for the other nodule types and patch-based approaches can be effective in tackling such a problem.

A. Lung Nodule Classification with Multi-Level Patch-based Context Analysis

In this paper, a novel classification method for the four types of lung nodules, i.e., well-circumscribed, vascularized, juxta-pleural and pleural-tail, in low dose computed tomography (LDCT) scans. This method is based on contextual analysis by combining the lung nodule and surrounding anatomical structures, and has three main stages.

First, a improved superpixel clustering method based on quick shift is designed to generate the patch division; (2) multi-level partition of the derived patches is used to construct level-nodule (i.e., patches containing the nodules), and level-context (i.e., patches containing the contextual structures). A concentric level partition is thus constructed to tackle the rigid partitioning problem.

Second, a feature set of three components is extracted for each patch of the image that are as follows: (1) a SIFT descriptor, depicting the overall intensity, texture, and gradient information; (2) a MR8+LBP descriptor, representing a richer texture feature incorporating MR8 filters before calculating LBP histograms; (3) a multi-orientation HOG descriptor, describing the gradients and accommodating rotation variance in a multi-coordinate system.

Third, the category of the lung nodule image is finally determined with a probabilistic estimation based on the combination of the nodule structure and surrounding anatomical context: (1) SVM is used to compute the classification probability based on level-nodule; (2) pLSA with contextual voting is employed to calculate the classification probability based on level-context. The designed classifier can obtain better classification accuracy, with SVM capturing the differences from various nodules, and pLSA further revising the decision by analyzing the context.

B. An Efficient Content Based Image Retrieval Using Local Tetra Pattern

The LBP and the LTP extract the information based on the distribution of edges, which are coded using only two directions (positive direction or negative direction). Thus, it is evident that the performance of these methods can be improved by differentiating the edges in more than two directions. This observation has motivated to propose the four direction code, referred to as local tetra patterns (LTrPs) for CBIR. The LTrP describes the spatial structure of the local texture using the direction of the center gray pixel. In this work, propose a second-order LTrP that is calculated based on the direction of pixels using horizontal and vertical derivatives.

ALGORITHM:

Input: Query image; Output: Retrieval result

1. Load the image, and convert it into grayscale.
2. Apply the first-order derivatives in horizontal and vertical axis.
3. Calculate the direction for every pixel.
4. Divide the patterns into four parts based on the direction of the center pixel.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

5. Calculate the tetra patterns, and separate them into three binary patterns.

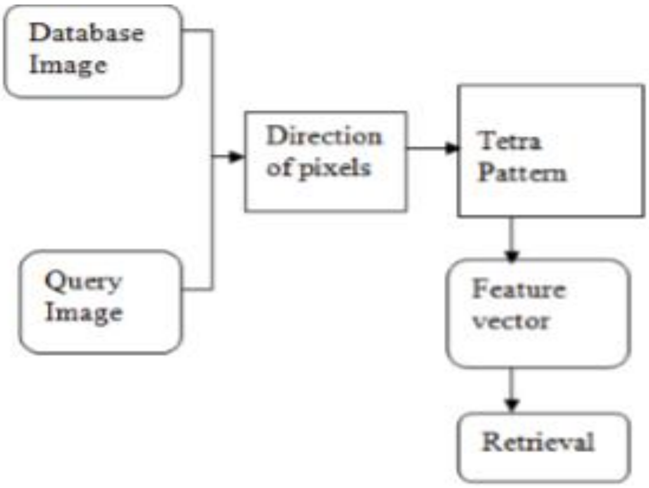

Fig-2: Proposed image retrieval system

C. Fuzzy Modeling of Brain Tissues in Bayesian Segmentation of Brain MR Images

Bayesian classifier is an unsupervised classifier designed based on Bayes' probability formula (equation I).

$$p(w|x) = \frac{p(x|w) \quad p(w)}{p(x)}$$

in which p(w) is the priori probability of class w, p(x|w) is the likelihood probability, p(x) is the probability of gray value x and p(w|x) is the posterior probability which makes the final classification. Each gray value x is classified as a member of class w if the posterior probability of it is highest in between all other classes. p(x) is constant for all the classes and so the classification result is independent of it.

Herein, the fuzzy classifier at second phase of the algorithm uses the Bayesian segmented image to improve it in subtle parts of borders between tissues. EM generated Gaussian probability distribution functions of gray values in each tissue are used as membership functions. 8-neighbors of each pixel are considered as neighboring system. Mamdani's max-min fuzzy inference system and Centroid Of Area (COA) are used in designing fuzzy classifier. Connected pixels of the same type constitute an object. Left/Right and Up/Down edges of neighboring rectangle are considered as opposite sides and accordingly, objects of each pair as opposite objects.

The following rules are the most important fuzzy rules are used in the algorithm.

• If "neighbours are WM", then "new center is WM"
• If "neighbours are GM", then "new center is GM"
• If "neighbours are CSF", then" new center is CSF"
• If "neighbours are WM", then "new center is not a CSF"
• If "neighbours are CSF", then "new center is not a WM"
• If "old center is WM" AND "number of new objects is more than old ones", then "new center is WM"
• If "old center is GM" AND "number of new objects is more than old ones", then "new center is GM"
• If "old center is CSF" AND "number of new objects is more than old ones", then "new center is CSF"
• If "neighbours have opposite separated WM objects", then" new center is WM"
• If "neighbours have opposite separated GM objects", then "new center is GM"
• If "neighbours have opposite separated CSF objects", then "new center is CSF"

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

D. Adaptive Multilevel Patch Based Lung Nodule Classification

Once the decision tree has been constructed, classifying a test record is straightforward. Starting from the root node, we apply the test condition to the record and follow the appropriate branch based on the outcome of the test. It then lead us either to another internal node, for which a new test condition is applied, or to a leaf node. When we reach the leaf node, the class lable associated with the leaf node is then assigned to the record, it traces the path in the decision tree to predict the class label of the test record, and the path terminates at a leaf node labeled no.

A decision tree is a sequence of binary splits of data. Several popular indexes are proposed to determine the best variable and best place on which to split a node in learning decision tree. Decision tree classifier is powerful for automatic feature selection. But one decision tree is unstable and only provides limited modeling of the joint statistics of data. The ends of training, all weak learners are combined with different weights to make decision. In this work, we use decision tree classifier for classification of lung nodules, since they work well for automatic feature selection and are efficient to train and apply.

The training steps of the decision trees are given as follows:

1. Initialize t = 1.
2. Learn decision tree Tt of K nodes based on the weighted distribution wt = [wt ,1, wt ,2,..., wt ,N] .
3. Assign the weighted log-likelihood ratio to each node Tt ,k for k =1, 2, ... ,K:
4. Update the weights of training data
5. Normalize the weights
6. If t < M, set t = t + 1 and go to step 2.

E. Context Curves for Classification of Lung Nodule Images

A context curve is thus calculated for each lung nodule image based on the superpixel labeling result. Formally, for image I, superpixel labeling output is firstly partitioned into circular sections SEC = {seck : k = 1,...,K} with K sections. Then, the foreground ratio FR= {frk : k =1,...,K} is computed for each section seck, i.e. frk = foreground(seck)/all(seck) (1) where foreground and all are the numbers of foreground and all pixels respectively. The final feature vector F(I)= {fk = frk : k =1,...,K} is obtained by combining the foreground ratios of all sections from inside to outside The overall procedure is illustrated in Fig.3. The circular partition is obtained by grouping the pixels that have the same distance to the centroid of lung nodule together. The total number of sections, K, is determined by the size of image, which is the radius of the last concentric circle that reaches the edge of the image.

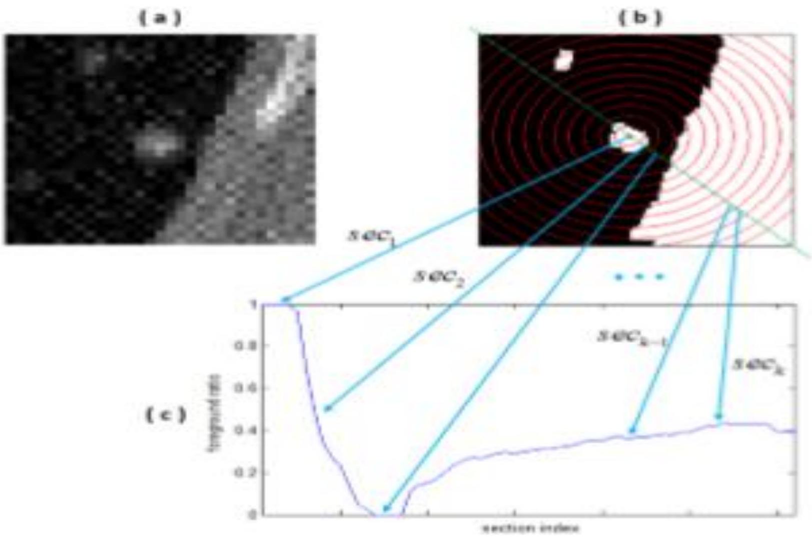

Fig-3: The procedure of calculation of context curve. (a) The original sample CT slide. (b) The superpixel labeling result with the circular partition. (c) The context curve. The red concentric circles indicate the context partition based on the distance of each pixel to centroid of the lung nodule. The light blue arrows indicate the foreground ratio for each section.

III. PROPOSED SYSTEM

The proposed system consists of four types of lung nodules classification such as well circumscribed, vascularised, juxta- pleural and pleural tail. Also, we improve the performance of texture descriptions of image by Local Tetra Pattern (LTrP). It has three main stages as follows as:

First, a improved superpixel clustering method based on quick shift is designed to generate the patch division; (2) multi-level partition of the derived patches is used to construct level-nodule (i.e., patches containing the nodules), and level-context (i.e., patches containing the contextual structures). A concentric level partition is thus constructed to tackle the rigid partitioning problem.

Second, a feature set of three components is extracted for each patch of the image that are as follows: (1) a SIFT descriptor, depicting the overall intensity, texture, and gradient information; (2) a MR8+LTrP descriptor, representing a richer texture feature incorporating MR8 filters before calculating LTrP histograms; (3) a multi-orientation HOG descriptor, describing the gradients and accommodating rotation variance in a multi-coordinate system.

Third, the category of the lung nodule image is finally determined with a probabilistic estimation based on the combination of the nodule structure and surrounding anatomical context: (1) SVM is used to compute the classification probability based on level-nodule; (2) pLSA with contextual voting is employed to calculate the classification probability based on level-context. The designed classifier can obtain better classification accuracy, with SVM capturing the differences from various nodules, and pLSA further revising the decision by analyzing the context.

IV.CONCLUSION

This paper presents a supervised classification method for lung nodule LDCT images. The designed proposed method can overcome the problem of the lung nodule overlapping with the adjacent structures. The proposed system first preprocesses the image and extracts the features. These features were used for classification of lung nodules into four categories: juxta-pleural, well-circumscribed, vascularized and pleural-tail, based on the extracted information. Finally, SVM classifier are utilized for classification of lung nodules.Also, the performance of texture descriptions of image by Local Tetra Pattern(LTrP) was enhanced. The results from the experiments on the low-dose CT lung image are showed a promising performance for the proposed method.

REFERENCES

[1] Fan Zhang ,Yang Song, Weidong Cai, Min-Zhao Lee, Yun Zhou, Heng Huang,Shimin Shan, Fulham M.J., Feng D.D., 2014 , " Lung Nodule Classification With Multilevel Patch Based Context Analysis", IEEE transaction on Biomedical Engineering,volume:61, Page(s): 1155 – 1166

[2] Ms. Shraddha D. Jumade, Prof. Sheetal S. Dhande , IJCSIT 2013,"An Image Indexing and Retrieval Method Using Local Tetra Pattern for Content-Based Image Retrieval (CBIR)",, Vol. 6 (3) , 2061-2064.

[3] Fan Zhang, Yang Song, Weidong Cai, Yun Zhou, Shimin shan & Dagan Feng, IEEE 2013, "Context curves for classification of lung nodule images".

[4] Ali Farzan, Syamsiah Mashohor, Abd Rahman Ramli, Rozi Mahmud, IEEE 2010" Fuzzy modelling of Brain tissues in Bayesian Segmentation of Brain MR images", pg no 77-80.

[5] M H Hasna, Jobin Jose ,IRJET 2015 "Lung nodule classification using multilevel patch-based context analysis and decision tree classifier" vol. 02, 2290-2294.

[6] Y. Song, W. Cai, Y. Wang and D. Feng, in Proc. IEEE, 2012, " Location classification of lung nodules with optimized graph construction", pp. 1439-1442.

[7] Abina Aliyar, Veena Vijayan , " Adaptive Multilevel Patch Based Lung Nodule Classification" IJESC 2015, ISSN-2321 -3361.