IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016

# Efficient Detection and Segmentation of Pulmonary Nodules using Quantization Approach

**Laxmi Priya .K**

M. Tech Student, Dept. of ECE, LBS College of Engineering, Kasaragod, India

**Abstract:** Cancer is one of the most serious illnesses in the world, among which lung cancer is the deadliest form. The survival rate of lung cancer in five years is only 54% and the early diagnosis rate is merely 15%. Lung cancer spreads to the different parts of the body rapidly before it is being diagnosed. Therefore the early detection has a crucial role in increasing the survival rate. This paper proposes a vector quantization (VQ) based approach for the detection of lung nodules from computed tomography (CT) scan images. Lungs are extracted from the surrounding tissues using simple thresholding. VQ is performed in two levels, first level of VQ is for lung segmentation and the second level is for the segmentation of lung nodules. Morphological closing operation is performed to refine the lung mask and to ensure the detection of juxta pleural nodules. False positives are reduced using Support vector machine (SVM) classifier. Experimental results shows improved performance comparing to existing computer aided detection (CAD) systems.

**Keywords:** Vector quantization (VQ), Computed tomography (CT) scan, False positives (FP), Computer aided detection (CAD), Support vector machine (SVM).

## I. INTRODUCTION

Lung cancer is the main cause of cancer deaths all over the world. The prognosis of lung cancer is very poor because doctors cannot find this disease until it is in an advanced stage. Survival rate in 5 years is 54% for early stages of lung cancer, but only 1 to 5% in advanced stages. Burden of the cancer is increasing in every year. Therefore the detection of lung cancer in its early stage has great importance [1].

Lung cancer is very hard to cure if it is detected in its final stage. A computer aided system for early detection of pulmonary nodule will play a crucial role in increasing the survival rate.

A CAD system for nodule detection generally consist of 3 stages

- Pre-processing
- Nodule candidate detection
- Classification

Most of the existing methods make use of segmentation to detect the infected lungs. Shape and appearance analysis is conducted to identify TPs (True Positives).Adaptive thresholding, watershed segmentations [2] fuzzy and global thresholding etc are the commonly used approaches for segmentation [3]. Contrast difference between lungs and surrounding tissues are quite high. So intensity based thresholding is highly effective [4, 5]. But the determination of accurate threshold is a challenging task. KNN (K Nearest Neighbour), ANN (Artificial Neural Network), SVM (Support Vector Machine) and Bayesian are commonly used classifiers to identify TPs.

This paper is organized in to 4 sections. Section-I describes an introduction about this work. Section-II introduces vector quantization as an image segmentation tool which is commonly used for data compression. Principle component analysis which is used for dimensionality reduction is also reported in this section. SVM classifier and proposed technique for pulmonary nodule detection is reported in section II. Section III draws the experimental results. Conclusion is described in section IV.

## II. PULMONARY NODULE DETECTION

### A. Vector quantization for Image Segmentation

Vector quantization (VQ) is a method of data compression [6]. It is one of a fixed-to-fixed length algorithm. Design of a vector quantizer is considered to be a challenging problem in earlier days. In 1980 Linde, Buzo and Gray (LBG) proposed an algorithm based on a training sequence. Hence it got the name LindeBuzoGray (LBG) algorithm [7].

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016

LBG algorithm is similar to K-means clustering algorithm. A group of vectors are taken as its input and produces a representative vector set. The convergence of this algorithm depends on the initial code book and the selected threshold. Each set of representative vectors give rice to specific clusters or classes. The idea of vector quantization for image segmentation is to classify voxels based on its intensity vector. The local intensity distribution can be represented by a vector of voxel intensities. Fig.1 illustrates typical configuration of local intensity feature vector that can be used by the VQ algorithm. In this study, a 3D first order neighbourhood is applied to form the local intensity feature vector.

B. Principle Component Analysis

A large quantity of body voxels is present in CT scan images of chest CT volume, and we form local intensity vector for each voxel. Intensive effort of computing is needed to process such a big data. So a dimension reduction technique is applied to the local feature vector space to reduce the computational complexity. A linear Karhunen–Loeve (K–L) transformation [8] of our local intensity feature vectors via principal component analysis (PCA) is obtained and we choose only the first few principal components, which contribute 95% of the total variance, which in turn optimize the dimension. The selected PCs are retained to form the feature vector for vector quantization, and the remaining PCs will be neglected as it contains very little information.

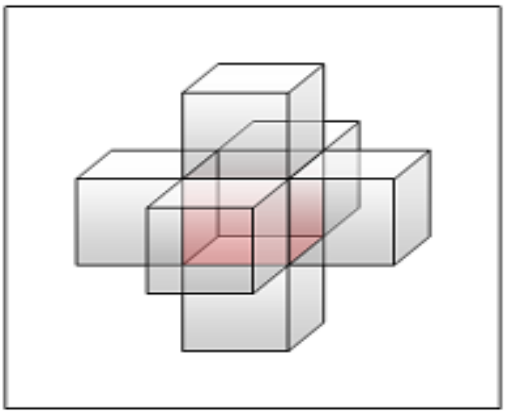


Fig. 1 Configuration of local intensity vector

C. SVM Classifier

SVM, Support vector machines are supervised learning method with specific learning algorithms which analyse data and recognize unique patterns, used for regression and classification. If we give a collection of training set, which belong to one of two categories, SVM training algorithm create a model which assigns new examples into one of the specified category and this makes SVM a non-probabilistic classifier. SVM [9, 10] can perform linear and nonlinear classification by using different kernel tricks. When our data sequences are not labelled accurately, supervised learning is not possible, and we go for unsupervised learning, this demands natural clustering of our data to groups, and assigns new data to these newly formed groups. These clustering algorithms which enhance the performance of support vector machines are called support vector clustering algorithm (SVC). Classifier block of this CAD system consist of an SVM classifier.SVM works based on various geometric, texture and intensity features extracted from CT scan images.

D. Proposed Method

This section introduces the detailed description of proposed CAD system. We have a pre-processor a detector and a classifier.

In the stage of pre processing, we started with the process of image enhancement which aims to improve the perception and interpretability of information in an image. This will helps us to provide best input for our automated techniques of image processing. Median filtering is adopted for contrast enhancement. In the block of pre processing, simple thresholding is used to extract chest volume from the field of view of the image. Usually outside of our chest volume will not have any anatomical structure. In detector block, high level VQ along with connected component analysis is applied to isolate two lungs from their surrounding anatomical structures. Nodules that grow nearer or originate from lung parenchyma are called juxta pleural nodule. In order to include these nodules initial lung mask is refined by morphological closing operation. Then a second level VQ is used to extract nodule candidates from lung volume. In order to separate nodules and non nodules SVM classifier is trained based on various features. Following figure shows the detailed flow chart of nodule detection system.

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016

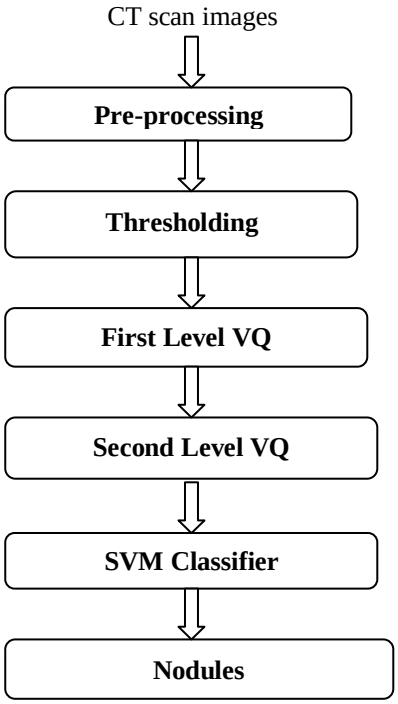



Fig. 2 Flowchart of the proposed system

Median filtering is a non linear operation which is often used in image processing to simultaneously remove noise and to preserve edges .proposed algorithm uses median filtering for contrast enhancement.
Lungs can be easily extracted from other anatomic structures by binary thresholding at -350HU.Selection of -350HU as the threshold enables the inclusion of juxta pleural nodules in the lung mask.
VQ can model the local statistics, and can analyses group of features, and then classifies each and every voxel in to the dimension-reduced feature space. VQ algorithm scans from the first voxel to the last voxel. VQ algorithm begins with only one class and the current number of classes k'= 1. The representative vector, C1 of first class is the local feature vector of the first voxel. Next step is the determination of squared Euclidian distance between intensity feature vector ωi of each and every voxel i and the representative vector Ck of every existing class k = 1, . . . ,k' .It is calculated using the following equation.

$$d(\omega_i c_k)=\sum_{p=1}^{P}(\omega_{ip}-c_{kp})^2 \qquad (1)$$

Here, dimension P, of the local feature vector ωi and each representative vector ck are determined by PCA. Now the codebook [7] CB = {c1, c2…....Ck'} is exhaustively searched for the nearest code vector Cmin such that

$$d(\omega_i c_{min})= \min_{1\le k\le K'}\{d(\omega_i c_k)\} \qquad (2)$$

Let T represents the threshold of intercluster distance. If d(ωi ,cmin) > T, then a new class , K'= K' + 1 is created subject to the constraint of maximum class number K. Otherwise it is added into the first class with proper updation of parameters in VQ algorithm [11].Algorithm is listed below

1) Determine the dimension P of the local intensity vector space using PCA. Let the K-L transformed local intensity vectors be $\omega_i = \{\omega_{i1}, \omega_{i2} \ldots\ldots \omega_{ip}\}$ where i = 1. . . I.
2) Chose the threshold value T as the maximum PC variance. Set the class number k as 2 for first level of VQ and k as 4 for second level of VQ.
3) For the first voxel ie for i = 1, set the voxel label $v_1$ = 1, and the representative vector $c_1$ of the first class is assigned as the local intensity vector $\omega_1$ and set $n_1$ = 1 as the number of voxels belonging to class- 1.Now the current number of class $k'$ = 1.
4) For i = i + 1, calculate the squared Euclidean distance $d(\omega_i c_k)$ between the voxel's local intensity vector $\omega_i$ and the representative vector $c_k$ for each and every class k = 1, . . . , $K'$.
5) Let $d(\omega_i c_m) = \min_{1\le k\le K'}\{d(\omega_i c_k)\}$, if $d(\omega_i c_m) <$ T or $K'$ = K, the label for the ith voxel is $v_i$ = m. cm is updated by $c_m = (n_m * c_m + \omega_i)/(n_m + 1)$, and $n_m = n_m + 1$.If the condition is not met, a new class is introduced with its representative vector $c_{K'} = \omega_i$, and label it as $v_i = K'$ such that $K' <=$ K.

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016

6) Repeat the steps from 4) for the entire image.
7) If current number of classes $K' < K$, then repeat steps from 1 to 6 for another scan and modify initial classification threshold T to the variance of the second or higher order principle components until reaching the target number of classes $K' = K$.
After isolating the lungs from the chest volume, next step is extraction of pulmonary nodules by second level VQ. Here second level VQ aims a more diversified classification. Histogram analysis of lung tissues predicts the existence of four types of tissues, low frequency parenchyma, high frequency parenchyma, blood vessels and INCs. So the class number for low level VQ is set to 4.
The feature-based SVM classifier operates on a series of features extracted from each of INC [12, 13] after second level VQ. Table I list out the details of extracted features. Geometric, intensity and texture features are extracted. Inclusion of texture feature improves the performance of SVM classifier .It will maintain the performance of the system in the level where we extract hessian and 3D features

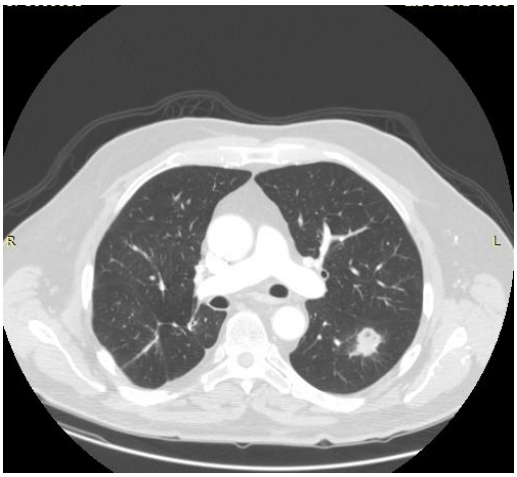
TABLE I OVERVIEW OF FEATURES FOR SVM CLASSIFIER

| No : | features | category |
|---|---|---|
| 1 | Area | 2D geometric |
| 2 | Diameter | 2D geometric |
| 3 | Eccentricity | 2D geometric |
| 4 | Circularity | 2D geometric |
| 5 | Mean | 2D intensity |
| 6 | variance | 2D intensity |
| 7 | Skewness | 2D intensity |
| 8 | Kurtosis | 2D intensity |
| 9 | Energy | 2D texture |
| 10 | contrast | 2D texture |
| 11 | Correlation | 2D texture |
| 12 | Homogeneity | 2D texture |

## III.EXPERIMENTAL RESULTS

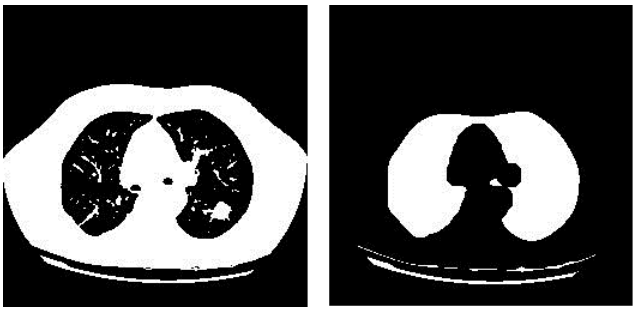
The experiments were conducted on publically available LIDC data base. The LIDC–IDRI CT scan images were collected from different scanner manufacturers. So their spatial resolution and imaging parameters will be different. e.g., slice thicknesses: varies from 0.45–5.0 mm; in-plane pixel size varies from: 0.461–0.977 mm; total slice number varies from: 96–605 slices; tube peak potential energies: 120–140 kV; and tube current: 40–627 mA. CT scan images collected are in DICOM format with a size of 512x512. Proposed algorithm is tested for 40 datasets obtained from LIDC. System shows sufficient detection power on juxta pleural nodules.

Following section shows our experimental results obtained while scanning patients CT slices. Lung mask obtained by global thresholding and first level vector quantization is depicted in figure.4. The 11th slice of a patient's CT scan having a total of 128 slices are shown here.



(a)

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016



(b) (c)

Fig. 2 a) chest CT slice b) Lung mask formed by simple thresholding c) lung mask formed by VQ

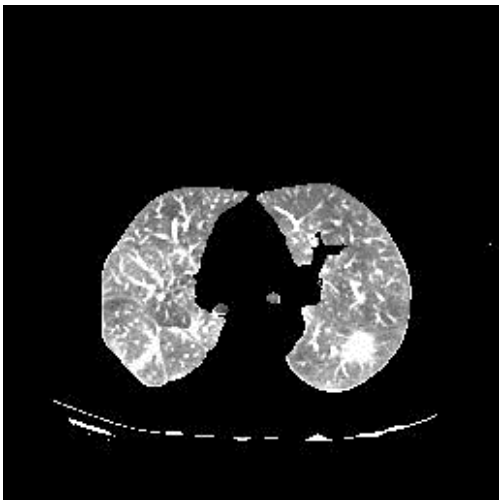


Fig. 3 segmented lungs after first level of VQ

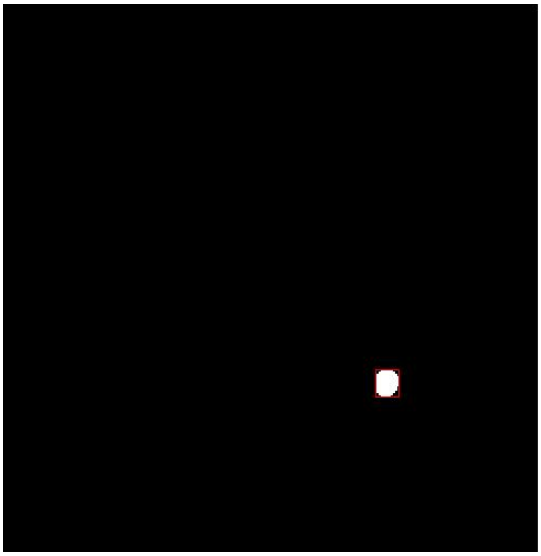


Fig. 4 Pulmonary nodule

## IV. CONCLUSION

A vector quantization based image segmentation approach is proposed for efficient detection of pulmonary nodules from CT scan images. The first level of VQ proves to be capable to replace the commonly used simple thresholding approach for extraction of lungs with higher accuracy. The following second level of VQ illustrates adequate detection power for juxta pleural and non GGO nodules. VQ based approach shows better performance than existing techniques. A sophisticated feature based SVM classifier is used to avoid the FPs. Geometric, intensity and texture features are

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
nCORETech
LBS College of Engineering, Kasaragod
Vol. 3, Special Issue 1, February 2016

extracted for SVM classification. Inclusion of texture features improves the efficiency of SVM classifier. While comparing to existing CAD systems which is evaluated on the same LIDC database, our method showed a comparable detection capability with a lower computational cost. The proposed CAD system yields comparable detection accuracy and more computational efficiency than existing systems, which enables our CAD system for clinical utility.

## REFERENCES

[1] C.I.Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D. M. Libby, M. W. Pasmantier, J. Koizumi, N. K. Altorki, and J. P. Smith, "Early lung cancer action project: Overall design and findings from baseline screening,"Lancet, vol. 354, no. 9173, pp. 99–105, 1999. R.Siegel,D Naishadham and A.Jemaal (2013) 'Cancer Statistics', vol.63,pp11-30,2013.

[2] Sayali Satish Kanitkar and N.D Thombare, "Detection Of Lung Cancer Using Marker Controlled Watershed Transform", International Conference on Pervasive Computing, (ICPC), 2015.

[3] I. Sluimer,M. Prokop, and B. van Ginneken, "Toward automated segmentation of the pathological lung in CT," IEEE Trans. Med. Imaging, vol. 24, no. 8, pp. 1025–1038, Aug. 2005.

[4] J. Pu, J. K. Leader, B. Zheng, F. Knollmann, C. Fuhrman, F. C. Sciurba, and D. Gur, "A computational geometry approach to automated pulmonary fissure segmentation in CT examinations," IEEE Trans. Med.Imaging., vol. 28, no. 5, pp. 710–719, May 2009.

[5] S. Ukil and J. M. Reinhardt, "Anatomy-guided lung lobe segmentation in X-ray CT images," IEEE Trans. Med. Imaging, vol. 28, no. 2, pp. 202–214, Feb. 2009.

[6] A. Gersho and R.M. Gray, Vector Quantization And Signal Compression,Boston, MA, USA: Kluwer Academic, 1992.

[7] C. C. Chang and W. C. Wu, "Fast planar-oriented ripple search algorithm for hyperspace VQ codebook," IEEE Trans. Image Process., vol. 16, no. 6, pp. 1538–1547, Jun. 2007.

[8] Yingbo Hua and Wanquan Liu," Generalized Karhunen–Loeve Transform", IEEE Signal Processing Letters, Vol. 5, No. 6, June 1998.

[9] C. C Chang and C. J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–27, 2011.

[10] S. G. Armato, G.McLennan, L. Bidautb, M. F. McNitt-Gray, C. R.Meyer, A. P. Reeves, B. Zhaoc, D. R. Aberle, C. I. Henschked, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. van Beeke, D. Yankelevitzd, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann,G. E. Laderach, D. Maxd, R. C. Pais, D. P.-Y. Qingf, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batrah, P. Caligiurii, A. Farooqid, G.W. Gladish, C. M. Jude, R. F.Munden, I. Petkovskaj, L. E. Quint, L. H. Schwartzk, B. Sundaram, L. E. Doddl, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallamm, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, S.Vastagh, B.Y. Croft, and L. P. Clarke, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A complete reference database of lung nodules on CT scans," Med. Phys., vol. 38, pp. 915–931, 2011.

[11] HaoHan, Lihong Li, Fangfang Han, HaO Zang, William Moore and Zhengrong Liang, "Fast and Adaptive Detection of Pulmonary Nodules in Thoracic CT Images Using a Hierarachical Vector Quantization Scheme", IEEE Journal Of Biomedical And Health Informatics, Vol.19, NO.2, March 2015.

[12] L. Li, D. Chen, H. Lu, and Z. Liang, "Segmentation of brain MR images: A self-adaptive online vector quantization approach," Proc. SPIE, 2001, vol. 4322, pp. 1431–1438.

[13] B. Sahiner, L. M. Hadjiiski, H. P. Chan, J. Shi, T. Way, P. N. Cascade, E. A. Kazerooni, C. Zhou, and J.Wei, "The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: Comparison on a data set annotated by multiple radiologists," Proc. SPIE, vol. 6514, pp. 65140L-1–65140L-7, 2007.