International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 22 No. 2 Apr. 2016, pp. 431-438

http://www.ijisr.issr-journals.org/

Early Detection of Pulmonary Nodules Using Hierarchical Vector Quantization Scheme

Resmi R. Nair[1], B. Thamilvalluvan[1], M. Anto Bennet[2], P. Lavanya[3], S. Ragavi[3], and S. Arun Kumar[3]

[1]Assistant Professor, Department of Electronics and Communication Engineering, VELTECH, Chennai-600062, India

[2]Professor, Department of Electronics and Communication Engineering, VELTECH, Chennai-600062, India

[3]UG Student, Department of Electronics and Communication Engineering, VELTECH, Chennai-600062, India

ABSTRACT: Computer-aided detection (CADe) of pulmonary nodules is critical to assist radiologist in early detection of lung cancer from computed tomography (CT) scans. So in proposed system we use CADe system based on hierarchical vector quantization (VQ) scheme. On comparing with commonly-used simple thresholding approach, the high-level VQ yields accurate segmentation of lungs from chest volume and in identifying initial nodule candidates (INCs) within lungs, low-level VQ proves to be effective for INC detection and segmentation, as well as computationally efficient compared to existing approaches. This proposed system also reduces false positive detection. False positive reduction is conducted via the rule based filtering operation in combination with feature-based support vector machine classifier. This proposed system shows out performance and demonstrate its potential for early detection of pulmonary nodules via CT imaging

KEYWORDS: CADe system, INC detection, INC segmentation, Computed tomography scans, Initial nodule candidates, Juxtapleural nodule annotation, Thoracic CT images.

INTRODUCTION

A lung nodule is defined as a "spot" on the lung that is 3 cm (about 1½ inches) in diameter or less it is also called benign which is non-cancerous. If an abnormality is seen on an x-ray of the lungs that is larger than 3 cm, it is considered as "lung mass" instead of a nodule, and is more likely to be cancerous and called as malignant. On the existing system x-rays and ct images are used on detecting pulmonary nodules. Lung nodules which is 8-10mm in diameter are only visible on the chest-xray. Nodules which is 0-2mm are visible through CT scan.

Many techniques where used on detecting pulmonary nodules, inspired on previous techniques, we have proposed a hierarchical vector quantization (VQ) approach to address the preprocessing and INCs detection issues in an adaptive manner, aiming to overcome the drawbacks of global thresholding methods. Compared with the existing approaches, the hierarchical VQ can be an alternative with either comparable detection performance and less computational cost, or comparable cost and better detection performance. To reduce FPs in the detected INCs, we make use of both rule-based filtering and supervised learning. Expert rules are learned from prior knowledge of true nodules annotated by the radiologists, while the classification rule for SVM is learned from two dimensional (2D) and (3D) features extracted from the INCs.

So in proposed system we use CADe system based on hierarchical vector quantization (VQ) scheme. On comparing with commonly-used simple thresholding approach, the high-level VQ yields accurate segmentation of lungs from chest volume and in identifying initial nodule candidates (INCs) within lungs, low-level VQ proves to be effective for INC detection and segmentation, as well as computationally efficient compared to existing approaches.

VQ was originally used for data compression in signal processing, and becomes popular in a variety of research fields such as speech recognition, face detection, image compression and classification, and image segmentation. It allows for the modeling of probability density functions by the distribution of protype vectors. The general VQ framework evolves two processes: (1) the training process which determine the set of codebook vector according to the probability of the input data; and (2) the encoding process which assigns input vectors to the codebook vectors. The well-known Linde-Buzo-Gray (LBG) algorithm has been widely used for the design of vector quantizer.

The algorithm aims to minimize the mean squared error guarantees to converge to the local optimality. However the following properties of the LBG algorithm limit its application in image segmentation (1) it relies on the initial conditions, and (2) it requires an iterative procedure and long computation times. Hence in our previous work, the self-adaptive online VQ scheme was proposed to speed up the vector quantizer, where the training and encoding processes are conducted in a parallel manner.

This section will provide details of the proposed CADe system for lung nodules. The top-level block diagram of the proposed CADe system is depicted in Fig.1. In the preprocessor block, the chest volume is extracted from the field-of-view (FOV) of the image volume by simple thresholding (where the outside of the chest volume does not have anatomical structures). In the detector block, the two lungs are first separated from their surrounding anatomical structures via high level VQ and a connected component analysis. In order to include juxta-pleural nodules (i.e., nodules grow near, or originated from the parenchyma wall), the initial lung mask obtained by the above separation operation is refined by a morphological closing operation. Then low level VQ is employed to identify and segment the INCs within the extracted lung volume. In the classifier block, the obvious FPs are firstly excluded from the INCs via rule-based filtering operations, and then the SVM classifier is trained to further separate nodules from non-nodules based on the 2D and 3D features of the INCs. More details on the proposed CADe system are given by the following sections.

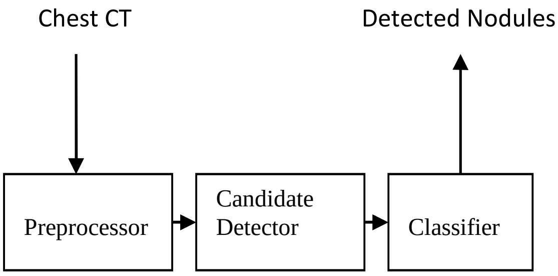

Fig 1. Top-level block diagram of the proposed CADe system for pulmonary nodules.

LITERATURE SURVEY

Surveying is the process of gathering each and every information about a think to know its growth. A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic, The purpose of literature is to demonstrate that the person is knowledgeable in the area of expertise that they requires.

CANCER STATISTICS, BY R. SIEGEL, D. NAISHADHAM, AND A. JEMAL

Maintaining a statewide cancer registry that meets both National Program of Cancer Registries and Centers for Disease Control and Prevention (CDC) high quality data standards and North American Association of Central Cancer registries (NAACCR) gold certification is accomplished through collaborative funding efforts.

Early lung cancer action project: Overall design and findings from baseline screening by 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

The choice of treatment for patients with cancers diagnosed as a result of screening is selected by the treating physician in conjunction with the participant. However, each participating institution must be committed to document, for each diagnosed case of lung cancer, the timing and nature of the intervention(s) (if any) and also the prospective course in respect to manifestations of metastases. The development and refinement of the screening protocol has been a concern of the ELCAP (Early Lung Cancer Action Program) Group for more than two decades, and it has been updated in the framework of the International Conferences organized by this Group and in the resultant international consortium on screening for lung cancer, I-ELCAP.

LUNG IMAGING AND COMPUTER AIDED DIAGNOSIS BY A. EL-BAZ AND J. SURI

An image-based CAD system for early detection of prostate cancer using DCE-MRI is introduced. Prostate cancer is the most frequently diagnosed malignancy among men and remains the second leading cause of cancer-related death in the USA with more than 238,000 new cases and a mortality rate of about 30,000 in 2013. Therefore, early diagnosis of prostate cancer can improve the effectiveness of treatment and increase the patient's chance of survival. Currently, needle biopsy is the gold standard for the diagnosis of prostate cancer. However, it is an invasive procedure with high costs and potential morbidity rates. Additionally, it has a higher possibility of producing false positive diagnosis due to relatively small needle biopsy samples.

GUIDELINES FOR MANAGEMENT OF SMALL PULMONARY NODULES DETECTED ON CT SCANS: A STATEMENT FROM THE FLEISCHNER SOCIETY BY H. MACMAHON, J. H. M. AUSTIN, G. GAMSU, C. J. HEROLD, J. R. JETT, D. P. NAIDICH, E. F. PATZ, AND S. J. SWENSEN

There is no clear consensus regarding the definition of a pulmonary nodule. Yet, "nodule" is one of the most common words found in chest CT reports. A committee of the Fleischner Society on CT nomenclature defined a pulmonary nodule as "a round opacity, at least moderately well marginated and no greater than 3 cm in maximum diameter".

DETECTION OF PULMONARY NODULES USING MTANN IN CHEST RADIOGRAPHS BY PREETHA.J, G. JAYANDHI

Computer aided scheme for pulmonary nodule detection in chest radiographs are to detect the pulmonary nodules (lung nodules) that are overlapped by ribs and clavicles and to minimize the false positive results caused by the ribs. Computed Tomography is used to detect lung nodules, but X-Rays are preferred due to low cost and low radiation dose. But x-rays does not effectively detect lung nodules because ribs and clavicles suppress the nodules and produce a false positive result to the radiologists. The purpose is to develop the CADe scheme with improved sensitivity and specificity by use of Virtual Dual Energy (VDE) chest radiographs. Ribs and clavicles in the chest radiographs (X-ray images) are suppressed with MTANN.

CONCLUSION

The above discussion gives the clear information on what the paper is based on. On inspired on the previous work, we proposed a CADe system using hierarchical vector quantization which is used for early detection of pulmonary nodules and also reduce false positive detection using rule-based filtering.

PROPOSED SYSTEM

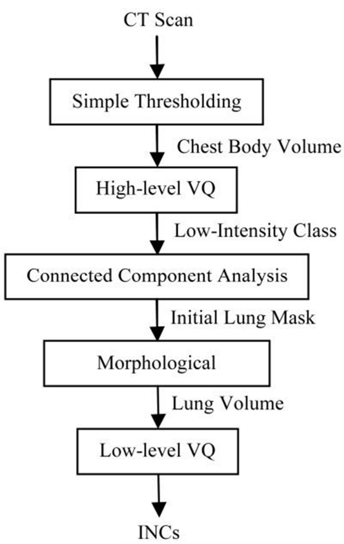

Fig 2.block diagram of proposed hierarchical vector quantization scheme for INCs detection

SELF-ADAPTIVE VQ ALGORITHM

VQ was originally used for data compression in signal processing, and becomes popular in a variety of research fields such as speech recognition, face detection, image compression and classification, and image segmentation. It allows for the modeling of probability density functions by the distribution of protype vectors. The general VQ framework evolves two processes: (1) the training process which determine the set of codebook vector according to the probability of the input data; and (2) the encoding process which assigns input vectors to the codebook vectors. The well-known Linde-Buzo-Gray (LBG) algorithm has been widely used for the design of vector quantizer. The algorithm aims to minimize the mean squared error guarantees to converge to the local optimality. However the following properties of the LBG algorithm limit its application in image segmentation (1) it relies on the initial conditions, and (2) it requires an iterative procedure and long computation times. Hence in our previous work, the self-adaptive online VQ scheme was proposed to speed up the vector quantizer, where the training and encoding processes are conducted in a parallel manner.

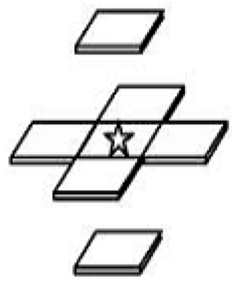

fig 3.a the 3D first order neighborhood to form a 7D local intensity vector

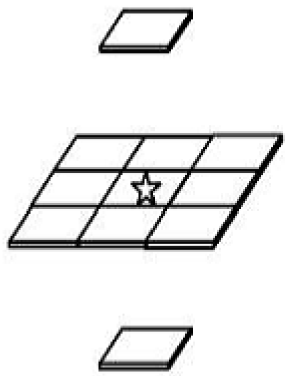

fig 3.b the 3D up-to-second order neighborhood to form a 11D local intensity vector

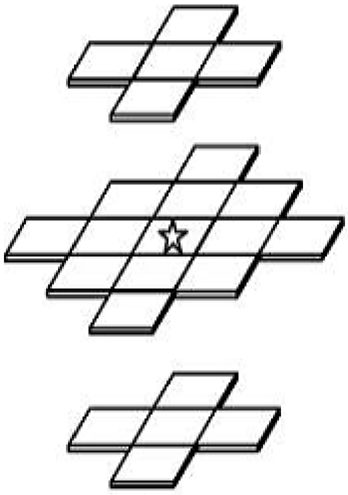

Fig 3.c the 3D up-to-third order neighborhood

Medical image segmentation is the key step toward quantifying the shape and volume of different types of tissues in a given image modality which are used for 3D display and feature analysis to facilitate diagnosis and therapy. The main idea of VQ for image segmentation is to classify voxels based on their local intensity distribution rather than voxels-based intensities. The local intensity distribution can be described by a vector of local intensities.fig 3.a,3.b,3.c illustrates three typical configurations of the local intensity vector that can be used by the VQ algorithm and in this study the 3D first order neighborhood was chosen to form the local intensity feature vector.

VQ models the local statistics, analyzes group features and classifies each voxel in the dimension-reduced feature space. VQ scans from the first voxel to the last one in a contiguous manner. For a given volume of interest (VOI), there is only one class at the beginning (i.e., the current total number of classes $K' = 1$) and its representative vector c_1 is the local feature vector of the first voxel. Since the initial value setup is data-oriented, the VQ algorithm is fully automatic. For each following voxel i, the squared Euclidean distance between its local feature vector ω_i and the representative vector c_k of every existing class $k = 1, \ldots, K'$ is calculated by

d (ωi, ck) =∑p=1p (ωip−ckp) 2

(1)

Here the local feature vector ω_i and each representative vector c_k are both of dimensions P, which was previously determined by PCA. The finite set codebook CB = $\{c_1, c_2, \ldots, c_{K'}\}$ is then exhaustively searched for the nearest code vector $c_{\min}$ such that

d(ωi,cmin)=min1≤k≤K′{d(ωi,ck)}

(2)

Let T denote the threshold for inter-cluster distance, so that if $d(\omega_i, c_{\min}) > T$, a new class $K' = K' + 1$ is generated subject to the constraint of maximum class number K. Otherwise, if $d(\omega_i, c_{\min}) = d(\omega_i, c_k) < T$ or$K' = K$, the representative vector c_k of the current class k is updated after adding a new member x_i into the class k. After a whole scan of the VOI, the representative vector, prior probability, and covariance of each cluster are generated. Meanwhile, all voxels have been classified under the nearest neighbor rule, where the exhaustive search under condition of eq(2) ensures the optimal classification result. As described above, our VQ algorithm only depends on two parameters: K and T. The maximum class number K can be determined according to radiologists' prior knowledge of how many major tissue types are perceived in the specific VOI. For instance in chest CT images, the lungs usually consist of four major tissue types: low and high frequency parenchyma, blood vessels, and nodules. And their average intensities increase from the lowest to the highest. To set an appropriate value for the classification threshold T is more crucial than the setting of K. If T is too large, only one class could be obtained. On the other hand, if T is too small, redundant classes might occur. According to extensive numerical experiments, a robust choice for T would be the maximum principle component variance of the local intensity vector series. In addition, to avoid the situation of resulting class number being less than the expected number of tissue types, the class separation threshold T may be tuned to be the second or third maximum principal component variance. Since the similarity threshold T is estimated from each CT scan, the algorithm is self-adaptive. The proposed VQ-based image segmentation algorithm is outlined as follows.

- Perform PCA to obtain the K-L transformation matrix for the target VOI, determine the reduced dimension P for the local intensity vector space, and calculate the K-L transformed local intensity vector $\omega_i = \{\omega_{i1}, \omega_{i2}, \ldots, \omega_{iP}\}$ for each voxel $i = 1, \ldots, I$.
- Set the classification threshold T as the maximum principal component variance, and set a value for the maximum class number K based on prior anatomical knowledge.
- $i = 1$, set the first voxel label $v_1 = 1$, its local intensity vector ω_1 as the representative vector c_1 for the first class, $n_1 = 1$ as the number of voxels belonging to class 1, and $K' = 1$ as the current number of classes.
- $i = i + 1$, calculate the squared Euclidean distance $d(\omega_i, c_k)$ between the local intensity vector ω_i of the current voxel and the representative vector c_k for each existing class $k = 1, \ldots, K'$.
- Let $d(\omega_i, c_m) = \min_{1\le j\le K'}\{d(\omega_i, c_j)\}$, if $d(\omega_i, c_m) < T$ or $K' = K$, the label for the i-th voxel is $v_i = m$. c_m is updated by $c_m = (n_m{*}c_m + \omega_i) / (n_m + 1)$, and $n_m = n_m + 1$. Otherwise, a new class $K' = K' + 1$ is generated with representative vector $c_{K'} = \omega_i$, and the current voxel is labeled as $v_i = K'$ s.t. $K' <= K$.
- Repeat from step 4 until $i = I$ to complete a whole scan.
- If $K' < K$, repeat steps 1) to 6) for another whole scan while setting the classification threshold T to be the second or third maximum principal component variance until $K' = K$.

In this paper, we mainly focus on demonstrating the merits of our hierarchical VQ scheme in the detection of INCs, from where a novel CADe system is proposed to efficiently detect pulmonary nodules

INCS DETECTION VIA A HIERARCHICAL VQ SCHEME

A very important but difficult task in the CADe of lung nodules is the detection of INCs, which aims to search for suspicious 3-D objects as nodule candidates using specific strategies. This step is required to be characterized by a sensitivity that is as close to 100% as possible, in order to avoid setting a priori upper bound on the CADe system performance. Meanwhile, the INCs should minimize the number of FPs to ease the following FP reduction step. This section presents hierarchical VQ scheme for automatic detection and segmentation of INCs.To further investigate the maximum class number K, we conducted repeated experiments of applying VQ to lung voxels classification with different K values.

RULE-BASED FILTERING OPERATIONS

It is challenging to thoroughly separate nodules from attached structures due to their similar intensities, especially for the juxta-vascular nodules (the nodules attached to blood vessels). Since the thickness of blood vessels varies considerably (e.g., from small vein to large arteries), a 2-D morphological opening disk with radius of 1 up to 5 pixels was adopted to detach vessels at different degrees.

FEATURE-BASED SVM CLASSIFICATION

A supervised learning strategy is carried out using the SVM classifier to further reduce FPs. Our feature-based SVM classifier relies on a series of features extracted from each of the remaining INC after rule-based filtering operations.

EXPERIMENTAL RESULTS AND DISCUSSION

Most existing approaches utilize the outcome from simple thresholding to extract lungs from the chest volume. Though thresholding is computationally inexpensive, the associated side effect, called “salt and pepper” noise, diminishes the computational advantage. Furthermore, because of relatively high image contrast between pathologic abnormalities and normal lung parenchyma, it is known that the conventional thresholding methods fail to extract complete lungs from such scans. The proposed high level VQ scheme can avoid the failure on lung boundary corrections, as shown by the rectangular regions marked in fig.4. Although both thresholding and VQ are intensity-based approaches, VQ classifies each voxel based on its local intensity features rather than the single voxel intensity used by simple thresholding. Moreover, most simple thresholding approaches set a uniform threshold for all CT scans, which is unrealistic, while the similarity threshold in VQ is adaptively determined for each scan. This makes VQ more robust to intensity inhomogeneity and image noise.

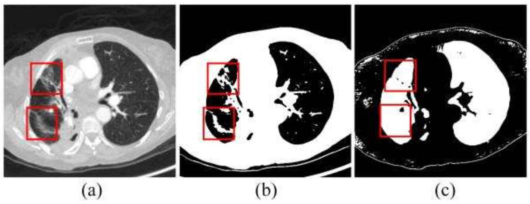

Fig 4.Comparison of simple thresholding and VQ for lung segmentation. Panel (a) shows a raw CT image with dense pathologies on the left lung lobe. Panel (b) illustrates the lung mask obtained by simple thresholding, and panel (c) obtained lung mask

By the high level VQ scheme, the obtained initial lung mask corresponds to the largest and the second largest (if left and right lungs are disconnected due to pathologic abnormalities) connected components in the low-intensity class, where the holes inside the extracted lung mask are filled by a flood-fill operation. Furthermore, in order to include juxta-pleural nodules into consideration, a 3D morphological closing operation using a spherical structuring element of radius 15 mm is applied to close the boundary in the binary lung mask.

CONCLUSION

In this paper, a novel CADe system was proposed for early detection of pulmonary nodules in chest CT scans. Based on our previous work of self-adaptive online VQ for image segmentation, we developed a hierarchical VQ scheme for INCs detection. The high level VQ proves to be feasible to replace the commonly-used simple thresholding scheme for extraction of the lungs in terms of higher accuracy, comparable processing time and automation level. The following low level VQ illustrates adequate detection power for non-GGO nodules, and is computationally more efficient than the state-of-the-art approaches. In this study, simple expert rules were firstly employed to exclude obvious FPs from being considered by the sophisticated feature-based SVM classifier, and further reduced the computational complexity. The SVM classification results indicated that gradient features contributed the most against any of the other three groups of features (geometric, intensity, and Hessian features). The forward feature selection strategy showed that the SVM classifier performed the best in the “gradient + intensity” feature space rather than in any other feature combination spaces. The optimal operating point of the SVM classifier for the best feature subset yielded a sensitivity of 92.7% and a specificity of 93.3%. In terms of the free-response ROC analysis, the proposed CADe system achieves an overall sensitivity of 82.7% at 4.0 FPs per scan. Compared with existing CADe systems evaluated on the same lung image LIDC database, our approach showed a comparable detection capability but a lower computational cost. In particular, we reported the performance of our system for the detection of juxta-pleural nodules. The outcome from our CADe system, with an overall sensitivity of 89.2% at a specificity level of 4.14 FPs/scan, is promising for tackling this challenging detection task. In a nutshell, the proposed hierarchical INCs detection approach is fast, adaptive and fully automatic. The presented CADe system yields comparable detection accuracy and more computational efficiency than existing systems, which demonstrates the feasibility of our CADe system for clinical utility.

REFERENCES

[1] Hao Han, Lihong Li, Senior member, IEEE. Fangfang Han, Bowen Song, William Moore, and Zhengrong Liang, Fellow, IEEE. Fast and Adaptive Detection of Pulmonary Nodules in thoracic CT images using a hierarchical vector quantization scheme.IEEE journal of bio medical and health informatics vol_19 no.2,march 2015.

[2] Siegel R, Naishadham D, Jemal A. Cancer statistics. CA Cancer J. Clin. 2013; 63:11–30.

[3] Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, Libby DM, Pasmantier MW, Koizumi J, Altorki NK, Smith JP. Early lung cancer action project: Overall design and findings from baseline screening. Lancet. 1999; 354:99–105.

[4] El-Baz A, Suri J. Lung imaging and computer aided diagnosis. Taylor & Francis; 2011.

[5] MacMahon H, Austin JHM, Gamsu G, Herold CJ, Jett JR, Naidich DP, Patz EF, Swensen SJ. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology. 2005; 237:395–400.

[6] Van Ginneken B, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study. Med. Image Anal. 2010; 14:707–722.

[7] El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, Soliman A, Abdollahi B. Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. Int. J. Biomed. Imag.2013; 2013(942353).

[8] Teramoto A, Fujita H. Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter. Int. J. CARS. 2013; 8:193–205.

[9] Tan M, Deklerck R, Jansen B, Bister M, Cornelis J. A novel computer-aided lung nodule detection system for CT images. Med. Phys. 2011;38:5630–5645.

[10] Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2011;2(27).

[11] Armato SG, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A complete reference database of lung nodules on CT scans. Med. Phys. 2011;38:915–931.