Content-Based Image Retrieval System for Differential Diagnosis of Lung Cancer

Ashis Kumar Dhara¹, Chanukya Krishna Chama¹, Sudipta Mukhopadhyay¹, Niranjan Khandelwal²

¹Dept. of Electronics & Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, West Bengal, India
²Dept. of Radiodiagnosis, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Email: dear.ashis79@gmail.com, ckchama@iitkgp.ac.in, smukho@ece.iitkgp.ernet.in, khandelwaln@hotmail.com

Abstract

We have designed a content-based image retrieval (CBIR) system using chest CT images for differential diagnosis of lung cancer. The objective of CBIR system is to retrieve similar nodules from large chest CT image database for a given query nodule. This tool can also be used for training of junior radiologists by visualization of nodules having different shape and size. Lung cancer is a disease with significant prevalence in several countries around the world. Pulmonary nodules are potential manifestation of lung cancer. A pulmonary nodule is defined as approximately round opacity, with moderately distinct margin and not greater than 3 cm in maximum diameter. The five years survival of a patient diagnosed with lung cancer can be increased from 10%-15% to 65%-80% if the pulmonary nodules are detected at an early stage. The main challenge for designing such system lies in segmentation of nodules adjacent to the pleural wall or other structure.

There is large number of images generated by hospitals and clinics every day. These images play very important role in diagnosis of diseases, medical research and education. CBIR system could facilitate the development of interactive computer aided detection technology that would exploit the wealth of data stored in the archive. Finding similar images or reference is one way to assist radiologist during daily clinical practice. Budding radiologists can explore their perception by visualization of all possible sites of lesions for given query nodule and directly provide a diagnosis report without assistance of expert radiologists. The CBIR system is validated on one public database (Lung Image Database Consortium, LIDC) and one private database taken from PGIMER Chandigarh. The average precision achieved for LIDC data set is 72.18% and for PGIMER data set is 78.29%. Feedback system is integrated with CBIR system for inclusion of knowledge of expert radiologists to reduce the semantic gap. A team of radiologists from PGMIR, Chandigarh is working as medical partner to improve the system performance.

Keywords: CBIR system, CT images, lung cancer, pulmonary nodule, LIDC, precision and recall.
Objective

- Design and development of Content based image retrieval system of lung nodules from lung chest computed tomography (CT) images
- System will retrieve and display similar nodules for a given query nodule
- Nodule CBIR system can be used as training tool for budding radiologists
- This tool can also be used for differential diagnosis of lung diseases

1. Introduction

According to the statistics from American Cancer society, lung cancer is the primary cause of cancer related death in United States [1]. In United States 222,520 cases was diagnosed as lung cancer in 2010 [2]. According to SEER 17 data, the age-adjusted incidence of cancer of the lung and bronchus was 62.5 per 100,000 persons per year, from 2003 to 2007. The five years survival of a patient diagnosed with lung cancer at later stage is only 10%-15% [3]. If pulmonary nodules are detected at early stage then five years survival rate can be increased up to 65%-80% [4]. This makes early detection and volumetric analysis of lung nodules major front of research in the war against lung cancer. A pulmonary nodule is defined as approximately round opacity having maximum diameter less than 3 cm [5]. Depending on intensity profile pulmonary nodule are of two type Solid nodule and ground glass nodule.

Solitary pulmonary nodule (SPN) has uniform intensity distribution whereas ground glass nodule (GGN) has hazy increase in lung attenuation on CT studies that doesn’t obscure the underlying pulmonary parenchymal architecture shown in Fig. 1(a), Fig.1 (b). Compared to other imaging modalities, CT provides the most comprehensive information on nodule morphology. It depicts the nodule itself as well as its surroundings and distant morphological abnormalities in the lung parenchyma. Detection is easy for isolated nodule but the detection becomes difficult if nodules locate at pleural of vascular attachment shown in Fig. 1(c), Fig. 1(d). In response to a query nodule the CBIR system will retrieve similar type of nodule and display in 3-D viewer. Classification of retrieved nodule is also possible using this system.

Fig.1 showing different types of nodules (a) Solitary nodule (b) Ground glass nodule (c) Pleural nodule and (d) Nodule with vascular attachment

2. Existing CBIR system for lung nodules

The reported CBIR system particularly on lung CT images are ASSERT, BRISC etc. One of the earliest CBIR systems that focuses on lung CT images is the ASSERT project at Purdue University, which was first published in 1999. It investigated image features such as co-
occurrence statistics, shape descriptors, Fourier transforms and global gray level statistics. It used nearest-neighbor and multidimensional hashing for calculating similarity and the best precision reported by this system was 76.3%. Kawata et al. [6] developed a CBIR system on lung nodule in 2004 considering shape descriptors and density histograms to classify and retrieve 3-D lung CT volumes but precision and recall of this CBIR system was not reported. Disney et al. [7] in 2007 developed an open source pulmonary nodule image retrieval framework from chest CT images named as BRISC. They used gray-level co-occurrence features, Haralick feature, Gabor filter and achieved precision of 88% when one nodule is retrieved. Automated nodule segmentation interface and shape based feature of nodule were not explored in their studies. Different modules of designing nodule CBIR system are preprocessing, nodule detection, false positive reduction, nodule segmentation, feature extraction, similarity measure and retrieval of similar nodules.

3. Framework for proposed CBIR system

Large amount of work has done to develop computer assisted diagnosis (CAD) systems for detection and analysis of pulmonary nodule but the CBIR system for pulmonary nodule has not explored much. CBIR systems have identified as an important research topic in radiology to facilitate diagnostic decision support for medical image interpretation. The budding radiologists can clarify their doubts on different possible similar patterns of pulmonary nodules with the help of CBIR system for pulmonary nodule. The objective of proposed CBIR system for lung nodule is to assist budding radiologists and to reduce the uncertainty of the radiologist in identifying suspicious pulmonary nodules by providing a visual comparison of a given nodule to similar nodules of known pathology. The block diagram of proposed CBIR system for retrieval of CT images is shown in Figure 2. The different steps are Query formation, visual content extraction (in form of a feature vector), checking similarity between feature vector for the query image and retrieval of similar CT images from database and displaying the particular retrieved image in 3-D viewer.

![Figure 2 Block diagram of CBIR system for pulmonary nodule](image)

In proposed CBIR system user will provide a pathology bearing region (PBR) from CT image as query and the system will retrieve similar images for given query image from the database. After similarity measure, the most similar nodule with the query nodule will be tagged as Rank-1 and next similar image will be tagged as Rank-2 and so on. Then user will click on particular retrieved nodule (Rank-1, Rank-2) to visualize in 3D viewer as shown in Figure 3 and Figure 4.
Figure 3 Graphical user interfaces for providing query nodule

Figure 4 Graphical user interfaces for displaying the retrieved nodule (Rank-1)
4. Results

CBIR system is tested on two databases, one public database named Lung Image Database Consortium (LIDC) and other is private image data set taken from PGMIR Chandigarh. We have taken 30 subjects from LIDC data set containing 45 solitary pulmonary nodules and 25 subjects from PGMIR data set containing 40 solitary pulmonary nodules for evaluation of the performance of nodule CBIR system. The performance of CBIR system has provided in terms of precision and recall where precision is the ratio of relevant retrieved nodule to all retrieved nodule and recall is the ratio of relevant retrieved nodule to all relevant nodule present in the database. The average precision achieved for LIDC data set is 72.18% and for PGMIR data set is 78.29%.

5. Conclusion and Future work

We presented a CBIR system considering lung nodule as pathology bearing region for differential diagnosis and training of budding radiologists. The performance of CBIR can be enhanced by improving the knowledge base through collecting more feedback from expert radiologists. Further research on nodule characterization should be focused on the integration of multiple features like patient history and several histopathological information as well as details imaging information CT scan.

5. Acknowledgements

This work has been supported by Department of Information Technology (DIT), Govt. of India Grant number 1(3)2009 - ME & TMD. The authors are grateful to Dr. N. Kalra, Dr. M. Garg, Dept. of Radio-diagnosis, PGIMER, Chandigarh, for providing chest CT data sets and giving valuable advice to our research.

7. References

[1] Cancer Facts and Figure 2009 by American Cancer Society, http://www.cancer.org


