Volume 5, Issue 6, June 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Medical Image Feature Extraction for Computer Aided Diagnosis of Lung Cancer

Shahid Eqbal
Assoc. Prof.,
ECE Department, GCET,
Greater Noida, U.P., India

Dr. M. A. Ansari
Asst. Prof.
EE Department, GB University,
Greater Noida, U.P., India

Abstract— ***Different approaches and numerous methods were developed by researchers for the purpose of computer aided diagnosis of lung problems. Depending on the type of lung disorder, each of the lung pathologies may require a specific approach to follow in order to characterise the disease. Since the focus of this paper is on lung cancer diagnosis & classification, one way is to look for lung nodules – which are round masses of tissue in the lungs, and can be early signs of cancer – and try to investigate whether they are benign or malignant. Also by means of measuring the characteristics of these tumour masses, one can predict their aggressive behaviour (i.e. how high is their metabolic activity). Major approaches in lung nodule detection and classification are discussed in this paper.***

Keywords— MIP, Image Acquisition, Preprocessing, Segmentation, Feature Extraction, Computer Aided Diagnosis, Lung Nodules, CT Scan, ROI, Genetic Algorithm (GA), Artificial Neural Network (ANN), Malignant, Benign.

I. INTRODUCTION

Medical image processing is one of the fastest growing areas within medicine at present, both in the clinical setting in hospitals. Medical image processing is the technique and process used to create images of the human body for clinical purposes or medical science. It is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. It can be seen as the solution of mathematical inverse problems. This means that cause is inferred from effect. This is very important to help improve the diagnosis, prevention and treatment of the diseases. Medical imaging is a part of biological imaging and incorporates radiology, nuclear medicine, investigative radiological sciences, endoscopy, thermography, medical photography and microscopy [27]. Medical imaging is the technique, process and art of creating visual representations of the interior of a body for clinical analysis and medical intervention. Medical imaging seeks to reveal internal structures hidden by the skin and bones, as well as to diagnose and treat disease. Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify abnormalities. Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are usually considered part of pathology instead of medical imaging. As a field of scientific investigation, medical imaging constitutes a sub-discipline of biomedical engineering, medical physics or medicine depending on the context: Research and development in the area of instrumentation, image acquisition (e.g. radiography), modeling and quantification are usually the preserve of biomedical engineering, medical physics, and computer science; Research into the application and interpretation of medical images is usually the preserve of radiology and the medical sub-discipline relevant to medical condition or area of medical science (neuroscience, cardiology, psychiatry, psychology, etc.) under investigation. Many of the techniques developed for medical imaging also have scientific and industrial applications [19]. Now-a-days, almost all areas of medical diagnosis are impacted by the digital image processing. When an image is processed for visual interpretation, the human eye is the judge of how well a particular method works. Clinical application demanding Radiotherapy plan, for instance, often benefits from the complementary information in images of different modalities. For medical diagnosis, Computed Tomography (CT) provides the best information on denser tissue with less distortion. Magnetic Resonance Image (MRI) provides better information on soft tissue with more distortion. With more available multimodality medical images in clinical applications, the idea of combining images from different modalities become very important and medical image fusion has emerged as a new promising research field [18]. Medical image processing (MIP) has been undergoing a revolution in past decades with the advent of faster, more accurate mass invasive devices [20]. This has driven the need for corresponding software development which in turn has provided impetus for new algorithms in signal and image processing [21]. Over the recent years, analysis of images such as segmentation, Edge Detection, Boundary detection, classification, clustering and texture property extraction were attracts the attention of many Researchers in the image processing and pattern recognition area. When compared to ordinary images the medical images, consists of so many information, in which the feature extraction is very difficult. Medical images, such as CT, MRI, show the information inside the patient body by non-invasive method, so that it is much helpful for doctor's diagnoses and less painful for patients. However the raw data can only give the material to doctor, the doctor has to decide by himself which is important which is not. The computer-aided diagnoses is to use computer to process the medical images to extract the useful information so that the doctor can make a diagnoses decision easier and quicker. But it is very difficult to locate the problems in medical images if it contains noise or the image is not in a proper format due to irregular structure of human body. Applying image processing technologies plays a pivotal role in processing and analyzing the images and also in forming the images. Detection of edges in an image helps us to understand the image feature. Since edges often occur at image locations representing object boundaries, edge detection is extensively used in image segmentation. The CAD system consists of five stages such as acquisition of TRUS image of prostate, preprocessing, segmentation, feature extraction and classification. The overview of the CAD system is depicted in figure 1 [17].

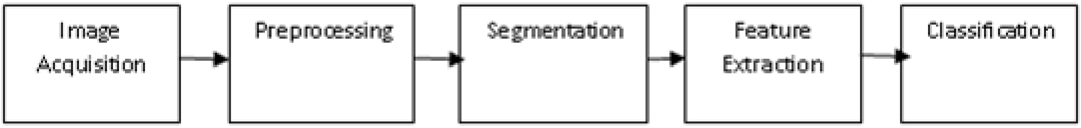

Figure 1: CAD System

All standard Computer vision aims to duplicate the effect of human vision by electronically perceiving and understanding an image. Giving computers the ability to see is not an easy task. The specific implementation of a computer vision system also depends on if its functionality is pre-specified or if some part of it can be learned or modified during operation. There are, however, typical functions which are found in many computer vision systems.

A. Image Acquisition

A digital image is produced by one or several image sensors, which, besides various types of light-sensitive cameras, include range sensors, tomography devices, radar, ultra-sonic cameras, etc. Depending on the type of sensor, the resulting image data is an ordinary 2D image, a 3D volume, or an image sequence. The pixel values typically correspond to light intensity in one or several spectral bands (gray images or color images), but can also be related to various physical measures, such as depth, absorption or reflectance of sonic or electromagnetic waves, or nuclear magnetic resonance.

B. Pre-Processing

Before a computer vision method can be applied to image data in order to extract some specific piece of information, it is usually necessary to process the data in order to assure that it satisfies certain assumptions implied by the method. Examples are:

1. Re-sampling in order to assure that the image coordinate system is correct.
2. Noise reduction in order to assure that sensor noise does not introduce false information.
3. Contrast enhancement to assure that relevant information can be detected.
4. Scale-space representation to enhance image structures at locally appropriate scales.

C. Feature Extraction

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called *features extraction*. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input (example, in medical imaging, extract anatomical boundaries before comparison with normal template and diagnosis). Knowledge or Feature extraction in Image processing: involves using algorithms to detect and isolate various desired portions or shapes (features) of a digitized image or video (for example to be used in Optical Character Recognition, Knowledge Based Road Extraction from Multisensor Imagery)**.** Image features at various levels of complexity are extracted from the image data. Typical examples of such features are**:** Lines, edges and ridges. Localized interest points such as corners, blobs or points. More complex features may be related to texture, shape or motion. Feature extraction of medical images is used to collect effective models, relations, rules, abnormalities and patterns from large volume of data. This procedure can accelerate the diagnosis process and decision-making. Different methods of feature extraction have been used to detect and classify anomalies in medical images such as wavelets [22,23], statistical methods and most of them used feature extracted using image processing techniques [24].Some other methods are based on fuzzy theory [25] and neural networks [26]. Most of the Computer Aided Methods proved to be the powerful tool that assists the radiologist to speed up the treatment process.

D. Detection/Segmentation

Segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels) (Also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). However, an efficient way of automated segmentation is difficult to achieve. Image segmentation plays an important role in applications of image analysis.

Numerous segmentation schemes exist to address requirements in different areas such as biomedical image processing, fingerprint recognition and face recognition. Newer algorithms are constantly being developed to address various needs. At some point in the processing a decision is made about which image points or regions of the image are relevant for further processing. Examples are:

1. Selection of a specific set of interest points
2. Segmentation of one or multiple image regions which contain a specific object of interest.

E. High-level processing

At this step the input is typically a small set of data, for example a set of points or an image region which is assumed to contain a specific object. The remaining processing deals with, for example:

1. Verification that the data satisfy model-based and application specific assumptions.
2. Estimation of application specific parameters, such as object poses or objects size.
3. Classifying a detected object into different categories.

II. DIFFERENT METHODS BEING USED & CURRENT TREND

Neural networks and genetic algorithms machine learning techniques were applied in some studies for automated detection of pulmonary nodules. Suzuki et al developed a computer aided diagnostic (CAD) scheme that uses a massive training artificial neural network (MTANN) – which is a trainable nonlinear filter based on an artificial neural network (ANN) – for distinction between benign and malignant lung nodules in low-dose helical CT scans [16]. Six parallel arranged expert MTANNs were used to differentiate between malignant nodules and six different types of benign nodules. The MTANNs were trained with ten typical malignant and ten benign nodules for each of the six types. Training was done independently using input CT images and teaching images containing the estimate of the distribution for the "likelihood of malignancy", that is, the teaching image contains a 2-D Gaussian distribution of a malignant nodule whereas its peak is located at its centre, and that of the benign has a zero value. Then the six MTANNs outputs were combined using an integration ANN in order to provide a value for malignancy assessment, with high values relating to malignant nodules and vice versa to the benign nodules. Zhang et al applied a 3-D cellular neural network (CNN) to detect small pulmonary nodules in high resolution helical CT scans [1]. Relying on the local shape properties for the purpose of voxel classification, local shape differences between nodules and blood vessels were captured using a shape index feature. While classification by voxels would allow for coverage of neighbouring information, the 3-D CNN was trained using genetic algorithm (GA) to deal with the shape index variation pattern of nodules. Lee et al combined two template matching methods based on a GA and conventional template matching techniques for detection of lung nodules in helical pulmonary CT images [2]. Nodules were detected within the lung area by the GA after specifying the target position and selecting the appropriate template image from several reference patterns, while the conventional matching template method was used to determine lung nodules along the lung wall by rotating semicircular models – serving as reference patterns – according to the orientation of the target point on the contour of the lung wall. Then 13 texture features were extracted and employed for false-positive (FP) findings elimination. Li et al evaluated psychophysical measures' capability in distinction between benign and malignant lung cancers in low dose CT scans [3]. Subjective similarity ratings for benign and malignant nodules were recorded by 10 radiologists. Then after feature extraction, the performance of four different techniques for determination of similarity measures, namely, feature-based pixel value difference-base, cross correlation-based and neural network-based techniques were evaluated by correlation with subjective similarity ratings. Others focused on the geometric and/or morphological shape of the pulmonary nodules. Brown et al used a generic CAD system model where baseline scans were employed for detection of candidate lung nodules for previously unseen patients [4]. Nodule features such as position, shape and volume would serve as baseline results, to be used then for comparison in follow-up scans. Farag et al used four different types of deformable templates to describe typical geometry and grey level distribution of lung nodules [5]. The four types are: solid spherical models of large size classified and non-classified nodules; hollow spherical model of large lung cavity nodules; circular model of small nodules; and semicircular model of lung wall nodules. Then the normalised cross correlation template matching by genetic optimisation and Bayesian post-classification are combined for nodule detection. Ge et al developed a CAD system that detects nodules and reduces FP through extracting 3-D shape information features from VOIs [6]. 3-D gradient field features and ellipsoid fitting were designed to distinguish nodules – which have a spherical shape – from the elongated shape of blood vessels. Classification was performed using linear discriminant analysis with stepwise feature selection, and a receiver operating characteristic (ROC) analysis was used to evaluate the FP reduction performance. Paik et al developed a CAD technique using a surface normal overlap method for detection of lesions [7]. This technique assumes that lesions such as long nodules and colonic polyps tend to have some convex regions on their surface, and thus an intersection might occur between the inward pointing surface normal vectors of these features and the tissue. However, the type of nodule (benign or malignant) was not investigated. Armato et al applied a ROC analysis to evaluate the performance of a linear discriminant classifier to distinguish between benign and malignant nodules in low dose helical CT images [8]. Morphological and grey level features were computed from each lung nodule candidate after specifying their locations using grey level thresholding and then fed to the classifier. Some focused on nodule volumetric measurements as a mark of malignancy, such as Kostis et al who measured the volumetric growth of small pulmonary nodule over time using 3-D methods applied to HRCT images for the purpose of distinguishing malignant from benign nodules [9]. 3-D intensity and morphology-based segmentation algorithms were developed for four different morphologic classes of pulmonary nodules. They showed that 3-D methods for nodule growth estimation rate are more accurate that than those based on 2-D measurements. In an extension to Kostis et al work, Reeves et al determined the likelihood of malignancy of pulmonary nodules from CT images via measuring the growth rate (i.e. change in nodule size) from two successive CT image scans recorded at close but different times [10]. Benign nodules have usually a slow growing rate compared to malignant nodules, and quantitative volumetric measurements can serve as a predictor of nodule's possible malignancy. The growth rate measurement accuracy was improved by using methods that match two images according to density (adaptive thresholding), location (registration), and vessel removal consistency (rule-based segmentation). Clustering techniques were used by Tanino et al using principal component (PC) analysis clustering for classification of ground glass opacities – a radiological term to describe hazy opacities within the lungs [11]. Suspicious shadows are first classified according to size into two sub-clusters, and then further classified into two new sub-clusters according to PC scores, where the last step is iterated recursively. Finally the abnormality of suspicious shadows is determined via Mahalanobis discriminant functions. Kanazawa et al detected candidates of lung cancer from helical CT images through delineating lung and blood vessels regions using fuzzy clustering algorithm [12]. Then features related to shape, grey value and position is extracted from each region and certain diagnostic rules were applied for detecting lung cancer nodule candidates. Examples of model-based techniques include the employment of the fractal dimension by Al-Kadi et al for improving lung cancer staging prediction accuracy from conventional CT modality [1]. Tumour region of interests (ROIs) were extracted from contrast enhanced CT images and quantitative performance analysis was used for discriminating between early and late stage tumours. Also strong correlation was shown between extracted tumour ROIs FD values and corresponding tumour staging as determined by positron emission tomography scan. Takizawa et al used a 3-D Markov random field model (MRF) for lung nodule recognition from X-ray CT images [13]. A mathematical morphology filter was used for locating suspicious shadow candidates, then volume of interests (VOIs) containing the shadows were extracted. A 3-D MRF model is used to evaluate the relationship between the geometrical object models (i.e. nodules and blood vessels) after calculating the probabilities of the hypothesis that a certain VOI relates to a nodule or a blood vessel. Moreover, filtering techniques were applied by Arimura et al, a difference-image technique for lung nodule enhancement and suppression of normal background structure [14]. Using low dose lung cancer CT images, a ring average filter and a matched filter were applied to generate a nodule-suppressed image and a nodule-enhanced image, respectively. The difference-image would then represent the subtraction of the nodule-suppressed and enhanced images. FPs were reduced using rule-based schemes and MTANNs. Li et al developed a selective lung nodule enhancement filter for improvement of nodule detection and reduction of FP rate [15]. The aim was to simultaneously enhance nodules and suppress other normal anatomic structures such as blood vessels and airway walls. Classification was done via an automated rule-based classifier and a case-based four-fold cross validation for performance evaluation. At last I have done my work in the same area that will be elaborated in a separate paper. I have applied techniques that went beyond simply differentiating between normal and abnormal tissue to the process of tumour type and stage categorisation. The intention was to cover both kinds of textures that might be encountered in medical images: textures acquired at a molecular or *macro-scale* ; and at a tissue or *micro-scale*.

III. CONCLUSIONS

Different available medical image feature extraction techniques had been studied in this paper. There is the need for development of medical image processing method that will be less time consuming and still effective. These results demonstrate that the developed systems could help the radiologists for a true diagnosis and decrease the number of the missing cancerous regions or unnecessary biopsies. Computer Aided Diagnosis System has to be developed, which acts as a secondary tool for the radiologists for diagnosing the cancer. Also, public medical database should be developed where categorized medical images can be made available to test system being developed by researchers. Future work involves working with radiologists to study the effect of CT image reconstruction algorithms and imaging protocols of various scanners on the FD for lung tumour stage prediction, and for other texture measures as well. This will allow for standardising lung tissue texture analysis procedures which would maintain texture feature quality consistency. Also, maintaining clinicians' diagnostic accuracy is an important issue when reducing the radiation dose in CT images.

REFERENCES

[1] X. W. Zhang, G. McLennan, E. A. Hoffman, and M. Sonka, "A complete CAD system for pulmonary nodule detection in high resolution CT images," in *Medical Imaging 2005 Conference*, San Diego, CA, 2005, pp. 85-96.

[2] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on Medical Imaging,* vol. 20, pp. 595-604, 2001.

[3] Q. Li, F. Li, J. Shiraishi, S. Katsuragawa, S. Sone, and K. Doi, "Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules," *Medical Physics,* vol. 30, pp. 2584-2593, Oct 2003.

[4] M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, "Patient-specific models for lung nodule detection and surveillance in CT images," *IEEE Transactions on Medical Imaging,* vol. 20, pp. 1242-1250, 2001.

[5] A. Farag, A. El-Baz, G. G. Gimel'farb, R. Falk, and S. G. Hushek, "Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates," *Medical Image Computing and Computer-Assisted Intervention - Miccai 2004, Pt 2, Proceedings,* vol. 3217, pp. 856-864, 2004.

[6] Z. Y. Ge, B. Sahiner, H. P. Chan, L. M. Hadjiiski, P. N. Cascade, N. Bogot, E. A. Kazerooni, J. Wei, and C. A. Zhou, "Computer-aided, detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting," *Medical Physics*, vol. 32, pp. 2443-2454, Aug 2005.
[7] D. S. Paik, C. F. Beaulieu, G. D. Rubin, B. Acar, R. B. Jeffrey, J. Yee, J. Dey, and S. Napel, "Surface normal overlap: A computer-aided detection algorithm, with application to colonic polyps and lung nodules in helical CT," *IEEE Transactions on Medical Imaging,* vol. 23, pp. 661-675, June 2004.
[8] S. G. Armato, M. B. Altman, J. Wilkie, S. Sone, F. Li, K. Doi, and A. S. Roy, "Automated lung nodule classification following automated nodule detection on CT: A serial approach," *Medical Physics,* vol. 30, pp. 1188-1197, Jun 2003.
[9] W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," *IEEE Transactions on Medical Imaging,* vol. 22, pp. 1259-1274, Oct 2003.
[10] A. P. Reeves, A. B. Chan, D. F. Yankelevitz, C. I. Henschke, B. Kressler, and W. J. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging,* vol. 25, pp. 435-450, 2006.
[11] M. Tanino, H. Takizawa, S. Yamamoto, T. Matsumoto, Y. Tateno, and T. Iinuma, "A detection method of ground glass opacities in chest x-ray CT images using automatic clustering techniques," *Medical Imaging 2003: Image Processing, Pts 1-3,* vol. 5032, pp. 1728-1737, 2003.
[12] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama, and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on helical CT images," *Computerized Medical Imaging and Graphics,* vol. 22, pp. 157-167, Mar-Apr 1998.
[13] H. Takizawa, S. Yamamoto, T. Nakagawa, T. Matsumoto, Y. Tateno, T. Iinuma, and M. Matsumoto, "Recognition of lung nodule shadows from chest X-ray CT images using 3D Markov random field models," *Systems and Computers in Japan,* vol. 35, pp. 82-95, 2004.
[14] H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, J. Shiraishi, S. Sone, and K. Doi, "Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening," *Academic Radiology,* vol. 11, pp. 617-629, 2004.
[15] Q. Li, F. Li, and K. Doi, "Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier," *Academic Radiology,* vol. 15, pp. 165-175, Feb 2008.
[16] K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Transactions on Medical Imaging,* vol. 24, pp. 1138-1150, 2005.
[17] Sathish kumar.M, Dinesh.E, MohanRaj.T, “Feature Extraction Method for the Discovery of Breast Cancer Lacerations by Using Mixture Model and EM Algorithm”, International Journal of Scientific & Engineering Research Volume 3, Issue 12, December-2012 1 ISSN 2229-5518
[18] Abhijit Smnathe, Ujjwal Harode, “A Novel Approach of Image Fusion based on Wavelet Transform and Curvelet Transform”, International Conference & Workshop on Recent Trends in Technology, (TCET) 2012 Proceedings published in International Journal of Computer Applications® (IJCA)
[19] A.P. James, B V. Dasarathy, 'Medical Image Fusion: A survey of state of the art', Information Fusion, 2014;http://dx.doi.org/10.1016/j.inffus.2013.12.00
[20] Hashemi, R.H., and Bradley, W.G., (1996), MRI: the Basic, Williams & Williams Baltimore.
[21] Signal Angenent, Eric Pichon, and Allen Tannenbaum, (2005). Mathematical methods in medical image processing, Bulletin (New Series) of the American Mathematical Society Volume 10. Pages 719-725.
[22] C.Chen and G.Lee, “Image segmentation using multitiresolution wavelet analysis and Expectation Maximum (EM) algorithm for mammography”, International Journal of Imaging System and Technology, 8(5):491-504, 1997.
[23] T.Wang and N.Karayaiannis, “Detection of microcalcification in digital mammograms using wavelets”, IEEE Trans. Medical Imaging, 17(4):498-509, 1998.
[24] S.Lai, X.Li and W.Bischof “On techniques for detecting circumscribed masses in mammograms”, IEEE Trans on Medical Imaging, 8(4):377-386, 1989.
[25] D.Brazokovic and M.Nescovic, “Mammogram screening using multisolution based image segmentation”, International journal of pattern recognition and Artificial Intelligence, 7(6):1437-1460,1993.
[26] I.Christiyanni et al., “Fast detection of masses in computer aided mammography”, IEEE Signal processing Magazine, Pages: 54-64, 2000.
[27] Manavalan Radhakrishnan1 and Thangavel Kuttiannan, “Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012 ISSN (Online): 1694-0814

