

An Evaluation of Features Extraction from Lung CT Images for the Classification Stage of Malignancy

Santosh Singh¹, Yogesh Singh², Ritu Vijay³

1, 3(Department of Computer Science and Electronics, Banasthali University, Banasthali, INDIA)

2(Department of Physiology, SMS Medical College & Attached Hospitals, Jaipur, INDIA)

Email: Santosh_singh075@yahoo.co.in

ABSTRACT : Image processing has become an active research topic in recent years. Researchers have focused on developing an algorithm using image processing to detect the different types of cancer in its early stage. Various preprocessing steps using image processing have been proposed. But high accuracy, sensitivity and specificity are important. To achieve this, preprocessing of the acquired original image is needed. This study evaluates CT images of lung, which contains noise. The preprocessing of these obtained images are necessary that is done by using image processing technique such as histogram equalization, thresholding, filtering followed by feature extraction which helps to reduce the process complications as well as accuracy can be improve.

Keywords - Accuracy, Computed Tomography Image, Feature Extraction, Image Processing, Sensitivity

I. INTRODUCTION

When cells multiply and grows abnormally, it causes cancer or tumor. Type of cancer depends in which part of the body, abnormal cells appear. The most common cause of deaths among all other types of cancer is lung cancer. Khin Mya Mya Tun et al describe the structure and four stages of lung cancer. The study shows that the cancer is confined to the lung in stage I, to the chest in II and III stages. Cancer has spread from the chest to other parts of the body in stage IV [1].

There are several techniques available to diagnose the lung cancer such as X-ray (Chest Radiograph), CT (Computed Tomography), PET (Positron Emission Tomography), MRI (Magnetic Resonance Imaging) scan and Sputum Cytology. However, most of these techniques are expensive and time consuming. Therefore, there was a need to develop a new technique for the recognition of lung cancer in its early stage to increase the chance of survival among people. Due to better clarity, low noise and distortion, CT images are preferred as data input. Usually CT images contain some unwanted information like noise. Beside these, CT of a lung contains bones, liver, heart, bold vessels etc. Elimination of such type of information is necessary. For this purpose, image processing techniques are used, which provides good quality of the image. Now a day, in medical fields, image processing techniques are widely used in early detection of different types of cancers. By using this, manual analysis can be improved.

To eliminate some unwanted background information of the image, preprocessing is done which not only contains denoising but also removes other unnecessary parts of the body from the CT image. After applying preprocessing, lung nodule is obtained as Region of Interest (ROI) i.e. used for feature extraction. Features are calculated then for classification stage. For the classification stage, neural network is used [2].

II. REVIEW OF LITERATURE

The abnormality in the lung region that affects the corresponding organ is referred as lung diseases which can be mainly caused by genetic reason, infections and either by active or passive smoking. Thousands of people are affected from such type of diseases all around the world. The groups of common lung diseases are described by Nihad et al in their study in which they grouped different parts of lungs such as airways, alveoli (air sacs), interstitium, blood vessels, pleura and chest wall. A various kind of common lung diseases are listed

as acute bronchitis, asthma, cystic fibrosis, pneumonia, TB (tuberculosis), COPD, lung cancer, or different types of pulmonary edemas and so many breathing problems.

Shamala B Terdale et al successfully developed CAD system that detects lung nodules with diameter \geq 2mm. The developed system helps to identify the suspicious nodules and thus to increase the sensitivity of the diagnosis [...]. Mokhled S. AL-TARAWNEH proposed efficient technique for segmentation principles to be a ROI foundation for feature extraction obtaining. The comparison of other used techniques and the proposed technique gives extremely hopeful outcome [2].

Cox et al analyzed two types of statistical texture features i.e. Haralick's co-occurrence matrices and Lawis texture measures with different neural network classifiers [3].

A new method is proposed by Yang Song et al for lung tumor and lymph node abnormalities from PET-CT thoracic images based on the low-level intensity and neighborhood features and high-level contrast-type features, with a two-level SVM classification. Contrast-type features, with a two-level SVM classification. It showed high detection performance and capability in handling a wide variety of abnormal patterns [4].

A new implementation was done by Ye X et al, for detecting the nodules both solid and ground- glass opacity (GGO) by calculating 3-D local geometric and features of statistical intensity [5]. Raman Maini et al presented a comprehensive review of image enhancement techniques. The point processing methods are most primitive, yet essential image processing operations and are used primarily for contrast enhancement. They presented the effectiveness of each of the algorithms when applied separately and advised to use a combination of such methods to achieve more effective image enhancement [6].

III. METHODOLOGY

The preprocessing steps of the proposed system can identify the appropriate region affected by cancer. The aim of preprocessing is to get better quality of the acquired image. It involves image enhancement, noise removal, thresholding, segmentation. Every progression is successfully completed by using MATLAB as the software of choice. MATLAB provides multi-platform environment. This tool is used for prototyping, data analysis, and visualization with built-in support for matrices and matrix operations. Also, it is loaded with graphic capabilities, as well as a friendly programming language and development environment. In the software development cycle, it leads to precious time savings.

1.1. Data acquisition

First of all, normal and abnormal CT images are collected from the available database. On these dataset images, preprocessing is applied to get an enhanced image, which involves denoising, enhancement, segmentation etc. The preprocessing steps of the proposed system can identify the appropriate region affected by cancer. The aim of preprocessing is to get better quality of the acquired image. Next step involves feature extraction which is then calculated mathematically.

1.2 Preprocessing

First of all, the image was converted to gray scale [7] image that contain only brightness information. These images are determined as a 2-D array of pixels which is using 8 bits/pixels. Here, a pixel value of 0 is black and value of 255 is white, with intermediate values corresponding to varying shades of gray. The advantage of converting an image to gray scale is to reduce the processing time and to produce a faster algorithm [8].

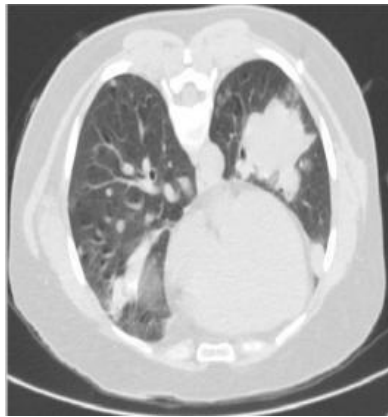


Fig. (a) Original image



Fig. (b) Grayscale image

Some factors such as blurriness, unnatural colors, noise, and artifacts often affect the quality of the acquired image, which is not adequate for further processing. Therefore, it is necessary to pre-process the image so that the irrelevant information or noise is removed. This process is called Image Enhancement i.e. it makes the image more valuable for further processing. The reasons for doing this consist of highlighting interesting details, removing noise and making images more visually appearing. To improve the quality of an image, it is necessary to assess its quality. This requires some quality quantification & assessment tools. Image enhancement techniques can be grouped into two broad categories as spatial domain and frequency domain. Spatial domain includes direct manipulation of image pixels whereas frequency domain includes manipulation of Fourier transform or wavelet transform of an image.

For the moment, this study concentrates on techniques that operate in the spatial domain i.e. Histogram Equalization.

1.3 Histogram Equalization

The histogram of a monochrome image is a graphical illustration of the frequency of occurrence of each gray level in the image. It reassigns the brightness values of pixels based on the image histogram. Histogram- modeling techniques are used to modify the image so that its histogram has a desired shape. This is helpful in stretching the low-contrast levels of images with narrow histograms. Histogram modeling has been found to be a powerful technique because they are simple, fast and with them acceptable results for some applications can be achieved. The aim of histogram equalization is to obtain a uniform histogram for the output image. Histogram equalization provides more visually pleasing results across a wider range of images. Mathematically, each individual histogram can be expressed as

$$h(k) = n_k = \text{card} \{ (x, y) | f(x, y) = k \} \quad (1)$$

Here, $k = 0, 1, \dots, L-1$, where L is the number of gray levels of the digitized image, and $\text{card}(\dots)$ denotes the cardinality of a set, that is, the number of elements in that set (n_k).

Mathematically, a normalized histogram can be defined as

$$p(r_k) = \frac{n_k}{n} \quad (2)$$

Where n is the total number of pixels in the image and $p(r_k)$ is the probability (percentage) of the k^{th} gray level (r_k). Figure (c) represents the histogram of the original image with one bar per gray level in which the height of the bar is proportional to the number (or percentage) of pixels that correspond to that particular gray level. Figure (d) signify the uniform histogram after equalization using `imhist()` function.

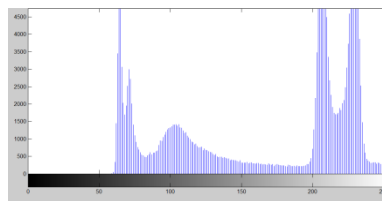


Fig. (c) Histogram of original image

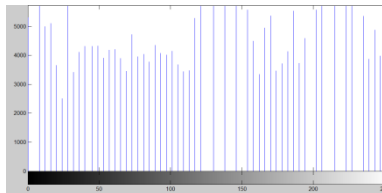


Fig. (d) Histogram Equalization

1.4 Noise Removal

Apart from contrast and intensity manipulation image enhancement includes noise removal, filtering etc which make the image suitable for further processing [9]. Image smoothing is useful for giving a softer visual effect and for removing noise. Noise can be defined as any unwanted information that contaminates an image. Noise can add up in an image while analog-to-digital conversion, communication channels, camera sensors etc. resulting in different types of noises such as

- Gaussian
- Impulse (Salt and pepper)
- Rayleigh
- Speckle
- Uniform
- Exponential
- Gamma (Erlang)

Enhancement of the acquired image focuses on eliminating noise that is the product of low-level data errors. Therefore, it is vital to remove noise to enhance the quality of the acquired image and to facilitate the further processing such as edge detection, segmentation and analysis etc. Filtering an image to smoothen noise, while maintaining the details of the image preserved is one of the most important issues.

The most admired and useful of the different types of filters is Median Filter. It is a nonlinear common enhancement digital filter [10]. It works by selecting the middle pixel value from the ordered set of values within the $m \times n$ neighborhood (W) around the reference pixel. In comparison with the smoothing filters, median filters offer the following advantages:

- No reduction in contrast across steps
- It does not shift boundaries (no disturbance in edges) as well as does not degrade edges,
- It smoothen additive white noise and effective in removing impulses.



Fig. (e) Filtered image

1.5 Thresholding

The purpose of the lung segmentation requisite for the CAD from CT scan images is to essentially separate the voxels corresponding to the lung cavity in the axial CT scan slices from the surrounding lung anatomy. Optimal thresholding selects the threshold based on the object and background pixel mean.

The simplest method within the intensity based methods is image thresholding. It is very popular due to its intuitive properties, simplicity and ease of implementation. In machine visual systems, thresholding is a common preprocessing step in which there are relatively few objects of interest. The fundamental problem of thresholding is the alteration of the image with several gray levels into another image with fewer gray levels, usually only two.

The most popular approach under the optimal thresholding was anticipated by Otsu and implemented by `graythresh` which compute global threshold. This method is based on threshold selection criteria. Based on this method, threshold value will be between 0 and 1. After achieving this value we can segment an image based on it [11].



Fig. (f) Threshold Image



Fig. (g) Removal of the background

1.6 Morphological Operations

The main morphological operations are dilation, erosion, closing, opening and the hit-or-miss transform. Among these apply morphology closing (dilation and erosion) on the image. It fills the indentation caused by the pulmonary vessels [12]. The effect of dilation is to “grow” or “thicken” objects in a binary image whereas, in erosion, outcome is to “shrink” or “thin” objects. The extent and direction of the thickening and thinning are controlled by the shape and size of the structuring element. In the present study structuring element is disk of radius is one. Dilation is implemented by using function `imdilate()` and erosion by `imerode()`.



Fig. (h) Image obtained after morphological operations



Fig. (i) Segmented nodule

1.7 Feature Extraction

One of the most important tasks is feature extraction. After the segmentation the next step is feature extraction that is performed on segmented lung region obtained from above step that distinguishes one region of interest from another. A feature contains some specific information which is extracted from the image to understand the details of the image. Features extracted from the image are as follows:

TABLE I VALUES OF EXTRACTED FEATURES FROM THE IMAGE

S. No.	Feature Extracted	Values
1.	Area	4970
2.	Perimeter	338.0488
3.	Shape Complexity	1.3530
4.	Mean	0.0190
5.	Standard Deviation	63.3872
6.	Circularity	0.5462

IV. CONCLUSION

Feature extraction is a challenging but important task in image processing which reduces the process complications as well as isolates various desired shapes of a given image. By using feature extraction one can improve the accuracy of classification stage. The features extracted from the segmented image acts as the foundation of the next stage. Further work is in progress to improve the accuracy.

REFERENCES

Journal Papers:

- [1] Khin Mya Mya Tun, Khaing, Implementation of Lung Cancer Nodule Feature Extraction using Digital Image Processing, *International Journal of Scientific Engineering and Technology Research*, vol 3 (9), May 2014, 2204-2210
- [2] Shamala B. T, Kulhalli K V CAD System for Lung Cancer Detection using ANN, *IOSR Journal of Electronics & Communication Engineering*, 11-15.
- [3] Cox G S, Hoare F J, Jager G de, Experiments in lung cancer nodule detection using texture analysis and neural network classifiers
- [4] Song Y, Cai W, Kim J and Feng D D, A Multistage Discriminative Model for Tumor and Lymph Node Detection in Thoracic Images, *IEEE Transactions on Medical Imaging*, May 2012, vol 31(5), pp 1061-1075
- [5] Ye X, Lin X, Dehmeshki J, Slabaugh G and Beddoe G, Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images, July 2009, *IEEE Transactions on Biomedical Engineering*, vol. 56(7), pp 1810-1820
- [6] Maini R and Aggarwal H A Comprehensive Review of Image Enhancement Techniques, *JOURNAL OF COMPUTING*, vol 2(3), MARCH 2010, ISSN 2151-9617
- [7] Sharma N, Aggrawal L M, Automated medical image segmentation techniques, *Journal of Medical Physics*, Jan-Mar 2010, vol 35(1), pp 3-14.
- [8] Chaudhary A, Singh S S, Lung Cancer Detection Using Digital Image Processing, *International Journal of Research in Engineering & Applied Sciences*, vol 2(2), February 2012, pp 1351-1359
- [9] Mesanovic N, Grgic M, Huseinagic H, Males M, Skejic E, Smajlovic M, Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm, *Proc. 18th International Conference on Systems, Signals and Image Processing*, IWSSIP 2011, pp 395-400.
- [10] Kumar S, Kumar A, Lung Segmentation Using Region Growing Algorithm, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4(3), March 2014, pp 184-187.
- [11] Hashemi A, Pilevar A H, Rafeh R, Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network, *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, 2013, 6, pp 16-24.
- [12] Kumar V, Saini A, Detection system for lung cancer based on neural network: X-Ray validation performance, *International Journal of Enhanced Research in Management & Computer Applications*, vol. 2(9), Nov - Dec 2013, pp 40-47.