# LETTER Differentiating Honeycombed Images from Normal HRCT Lung Images

## Aamir Saeed MALIK<sup>†</sup>, Nonmember and Tae-Sun CHOI<sup>††\*a)</sup>, Member

**SUMMARY** A classification method is presented for differentiating honeycombed High Resolution Computed Tomographic (HRCT) images from normal HRCT images. For successful classification of honeycombed HRCT images, a complete set of methods and algorithms is described from segmentation to extraction to feature selection to classification. Wavelet energy is selected as a feature for classification using K-means clustering. Test data of 20 patients are used to validate the method.

key words: segmentation, extraction, feature selection, wavelet energy, classification, K-means clustering

#### 1. Introduction

An automated Computer Aided Diagnosis (CAD) system requires unsupervised segmentation, segmented region extraction and labeling, feature selection and finally classification. Our aim is to address all these requirements. The objective of HRCT is to assess a variety of lung diseases, for example, pulmonary emphysema, nodules, interstitial lung disease. HRCT has been offering increasingly better anatomic resolution because of very thin image slices, and it is a noninvasive and non-traumatic diagnostic method. However, as a consequence of thin slices, we get a sequence of images consisting of large number of slices for a single patient. This fact results in time consuming visual assessment of all the images. According to [1], experienced observers typically make correct global diagnosis of parenchymal lung diseases in 40% to 70% of cases. Further, interpretation of HRCT images requires special training and experience from a radiologist. Hence, CAD can help to produce objective measures of abnormal patterns in lungs and assist radiologists in assessing diseases.

Today, HRCT is a widely used technique for the assessment of the lung diseases. In this letter, we deal with one of the disease patterns, i.e., Honeycombing. Honeycombing is a lesion belonging to the diffuse lung disease category. Among the entire diffuse lung disease family, honeycombing is the most difficult pattern to detect automatically as shown by Uppaluri et al. [2]. Pathologically, honeycombing is defined by the presence of small air-containing cystic spaces with thickened walls composed of dense fibrous tissue. As honeycomb cysts are usually clustered together, it

\*Corresponding author.

has the characteristic appearance of "honeycombing". Honeycombing is difficult to detect due to its textural and structural appearance, which changes with the progression of the disease.

In this letter, we present a method to differentiate the honeycombed images from the normal HRCT images. First, the optical filter described in Sect. 3 is used for the segmentation of the images. Then an algorithm is presented in Sect. 4 for extraction of the lungs. Section 5 describes features based on wavelet energy that are used for classification using K-means clustering in Sect. 6.

## 2. Previous Work

The related work can easily be divided into two categories; one dealing with the segmentation issue and the other with the classification. First we discuss the various segmentation methods published in literature. M.S. Brown et al. [3] used explicit anatomical knowledge (expected size, shape, relative position of objects in the same or adjacent slices, etc.) to generate an anatomical model. This model was developed with the guidance from experienced radiologists. S. Hu et al. [4] used an iterative searching method to compute an optimal threshold value for each CT case and use conditional morphological operations to segment lung regions. In addition, many other methods are published including using artificial neural networks to classify each pixel in the CT slice into different anatomical structure, a fully automated segmentation process based on segmenting the potential lung areas using an adaptive threshold based on pixel value distribution in each CT slice, using anatomy guided 3D watershed transform for lung lobe segmentation, etc.

Many papers have been published for assessing lung diseases. Generally they are based on textural features. Prasad et al. [5] presented an automated texture based unsupervised system for the classification of lung HRCT findings in emphysema, ground-glass opacity, honeycombing and bronchiectasis. Shamsheyeva et al. [6] interpreted lung patterns as textures and developed a texture classification technique for segmentation of lung patterns based on Support Vector Machines learning algorithm.

As mentioned above, in most of the previous research works, honeycombing was detected using texture. However, as the disease progresses, the appearance of honeycombed cysts changes to "reticular pattern", i.e., linear or curvilinear. Consequently, we have explored alternative methods for detecting honeycombing and our technique is different in

Manuscript received October 22, 2008.

Manuscript revised December 29, 2008.

<sup>&</sup>lt;sup>†</sup>The author is with Hanyang University, Korea.

<sup>&</sup>lt;sup>††</sup>The author is with Gwangju Institute of Science & Technology, Korea.

a) E-mail: tschoi@gist.ac.kr

DOI: 10.1587/transinf.E92.D.1218



Fig. 1 HRCT image, its DCT spectrum & filter designed with  $\sigma_1 = 0.01$  &  $\sigma_2 = 0.1$ .

respect that instead of exploiting various textural features, it uses energy information for classification of honeycombing.

#### 3. Segmentation

We introduce a segmentation method based on bipolar incoherent image processing. T.C. Poon and P.P. Banerjee [7] has discussed bipolar incoherent image processing in detail. The sharpness of pixel values in the image is found by convolving the spectrum of the intensity image  $(|I_0(x, y)|^2)$  with the transfer function which in our case is Optical Transfer Function (OTF). The computed image  $[i_c(x, y)]$  is given as:

$$i_c(x, y) = \text{Re}[\text{DCT}^{-1}\{\text{DCT}\{|I_0(x, y)|^2\}\text{OTF}_{\Omega}(k_x, k_y)\}]$$
(1)

Where:

$$OTF_{\Omega}(k_x, k_y) = \exp[-\sigma_1(k_x^2 + k_y^2)] - \exp[-\sigma_2(k_x^2 + k_y^2)]$$
(2)

The processing in the frequency domain is particularly useful for noise reduction as the noise frequencies are easily filtered out. The high frequency component of an image area can be determined by processing with Discrete Cosine Transform (DCT) and analyzing the frequency distribution. So, the OTF is a filtering operation that provides the sharpness at pixel points in an image. The filtering operation depends upon  $\sigma_1$  and  $\sigma_2$ . Figure 1 shows an HRCT image, its corresponding spectrum obtained using DCT and the filter with  $\sigma_1 = 0.01$  and  $\sigma_2 = 0.1$ .

Once the OTF is applied to the HRCT images, we can select the boundary points for the segmented regions in the HRCT segmented image. So the selection at a point (i, j) can be computed in a small window around (i, j) and the value at (i, j) can be replaced by the sum of computed values of all pixels in that window only after enhancing the extracted edges and removing noise, if any. Therefore, the segmented point can also be given by the following equation where OS<sub>O</sub> stands for Optical Segmentation technique:

$$OS_{O}(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} i_{c}(x, y)$$
(3)

Figure 2 shows the segmentation results with various different values of  $\sigma_1$  and  $\sigma_2$ . We have used heuristic approach to select  $\sigma_1$  and  $\sigma_2$  for segmentation.

#### 4. Extraction of Lungs

Lets first consider the extraction of lungs from the image.



Fig. 2 Segmentation results with different values of sigmas.

The algorithm for extraction of lungs is as follows:

1. Process original HRCT slice using OTF in DCT domain ( $\sigma_1 = 0.05$  and  $\sigma_2 = 1$ ) and invert it (hence removing thorax). The image obtained after OTF processing is inverted so as to highlight the edges detected by OTF. The inverted image  $I_V(x, y)$  is obtained from the computed image in Eq. (1) as:

$$I_V(x, y) = \max(i_c(x, y)) - i_c(x, y)$$
(4)

2. Convert the image to a binary image  $(I_B)$  using thresholding.

$$I_B = \begin{cases} 0, & I_V(x, y) < 0.99\\ 1, & I_V(x, y) \ge 0.99 \end{cases}$$
(5)

3. Label the regions as separate independent objects and remove small objects from the image based on the area of each object. Let  $R_L$  be region with label *L* then:

$$R_{L} = \begin{cases} 0, & Area(R_{L}) \le 8000 pixels \\ 1, & Area(R_{L}) > 8000 pixels \\ L = 1, 2, 3, 4. \end{cases}$$
(6)

4. Remove the regions that connect to any of the boundary in the image.

$$R_L = \begin{cases} 0, & boundary(R_L) = B & pixels \\ 1, & boundary(R_L) \neq B & pixels \\ L = 1, 2, 3, 4 \end{cases}$$
(7)

*boundary*  $(R_L)$  = Any boundary pixel of the region LB = Any boundary pixel of the image slice

- 5. Fill the holes within the remaining regions (the holes are filled with the boundary pixel value of the region).
- 6. Extract lungs based on the area calculated in step 3.

$$R_{Lung} = \begin{cases} \max(area(R_L)), & Lung = 1\\ (\max - 1)(area(R_L)), & Lung = 2 \end{cases} (8)$$
  
Where  $Lung = 1, 2; L = 1, 2, 3, 4 \dots$ 

Figure 3 shows some of the steps of above mentioned algorithm.

#### 5. Wavelet Energy

We base our theory on the premise that the various anatomical structures exhibit different and unique energy. Hence, the normal HRCT lung image will exhibit different energy as compared to the honeycombed HRCT image. Therefore, the energy becomes the unique attribute that will differentiate a honeycombed image from a normal HRCT image.



**Fig. 3** Some of the steps of the algorithm for extraction of lungs. (a) Inverted (step 1) (b) Binary (step 2) (c) Removing small objects (step 3) (d) Lungs extraction (step 6).

 Table 1
 Results of energy components of 3 patients.

Patient #	Shannon	Log Energy	Level 1			
			Approx	Horizontal	Vertical	Diagonal
1	-5.73E+10	32469.3	42.0847	16.91027	15.0102	26.03083
2	-5.10E+10	491028	39.79653	16.77827	17.25227	26.1729
3	-6.43E+10	533710.3	44.99213	15.28007	14.89673	24.83103

It is similar to the concept that every body part has unique absorption characteristic when exposed to X-Rays. X-ray attenuation can be used to distinguish between various body parts. Alternatively other information can be also used, for example, the energy information of a body part.

We use the energy information to separate honeycombed image from a normal HRCT image. That energy information is obtained using wavelet transform. We have used Reversible Biorthogonal wavelet and we show that energy obtained from this wavelet transform is quite distinct for differentiation between a honeycombed and a normal HRCT image.

We performed 3-level wavelet decomposition. Hence the coefficient matrices for all the three levels were generated. The coefficient matrices were generated for the approximation, horizontal, vertical and diagonal details. Once the coefficient matrices were obtained, we calculated energy corresponding to each of the coefficient matrix, i.e., for a two-dimensional wavelet decomposition, the percentage of energy corresponding to the approximation, horizontal, vertical, and diagonal details. Hence a total of 12 decimal values per wavelet are calculated for each image slice. We also calculated Shannon and Log energy entropy values for each image slice. All these energy values are calculated for the lungs extracted in previous section.

Table 1 gives a summary of these values for 3 patients. These values are given as a percentage of energy calculated. Table 1 provides energy values for the Shannon, Log energy and approximation, horizontal, vertical and diagonal energy percentage values for decomposition level 1. Similarly the approximation, horizontal, vertical and diagonal energy percentage values for decomposition levels 2 and 3 were calculated. We performed this experiment for 20 patients: 18 with normal HRCT images and 2 with honeycombed images.

## 6. Classification

In Hierarchical method, as clusters grow in size, the actual expression patterns become less relevant. Therefore, as the number of patients increase, the hierarchical method might fail. Hence, we have used the K-means clustering for clas-



sification which performs well for large cluster sizes. Kmeans clustering method partitions the observations in the data into K mutually exclusive clusters. We used cosine distance method for this calculation. Next, we used the member objects and the centroid to define each cluster. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized.

We performed the experiment with 20 patients. The patients with honeycombed HRCT images are patient #1 and patient #15. The rest have normal HRCT images. The results of this experiment are shown in Fig. 4. It can be seen that cluster #2 has patient #1 as well as patient #15 (both honeycombed patients in one cluster) while cluster #1 have Patient #2 to #14 and #16 to #20 (all the normal patients). From the silhouette plot in Fig. 4, it can be seen that almost all points in cluster #1 have a large silhouette value, greater than 0.7. Also, both the points in cluster #2 have a large silhouette value too, greater than 0.7. Hence, this large silhouette value indicates that the cluster #1 is well separated from neighboring cluster #2.

## 7. Conclusions

The objective of this letter is to present a classification method that can successfully differentiate the honeycombed HRCT images from the normal HRCT images. For this purpose, first the images are segmented using an optical filter followed by our proposed extraction algorithm that successfully extracts the lung regions. Then, percentage of wavelet energy is selected as a feature that is used by K-means clustering for classification of honeycombed and normal HRCT images. For testing, a set of 20 patients were selected: 18 of them had normal HRCT images and 2 of them had honeycombed HRCT images. It was found that the wavelet energies for the normal HRCT differed from those of honeycombed HRCT images. Hence the wavelet energies served as a differentiating feature for classification of normal and honeycombed HRCT images.

#### 8. Future Work

We plan to include data of more patients as well as comparison with other methods for validation of the proposed algorithm. Additionally, we will provide method for adaptive selection of  $\sigma_1$  and  $\sigma_2$ .

## Acknowledgments

This work was supported by the Bio Imaging Research Center at GIST.

#### References

- K. Nishimura, T. Izumi, M. Kitaichi, S. Nagai, and H. Itoh, "The diagnostic accuracy of high resolution computed tomography in diffuse infiltrative lung diseases," Chest, vol.104, pp.1149–1155, 1993.
- [2] R. Uppaluri, E.A. Hoffman, M. Sonka, P.G. Hartley, G.W.

Hunninghake, and G. MClennan, "Computer recognition of regional lung disease patterns," American Journal of Respiratory and Critical Care Medicine, vol.160, no.2, pp.648–654, 1999.

- [3] M.S. Brown, M.F. McNitt-Gray, N.J. Mankovich, J.G. Goldin, J. Hiller, L.S. Wilson, and D.R. Aberie, "Method for segmentation chest CT image data using an anatomical model: Preliminary results," IEEE Trans. Med. Imaging, vol.16, no.6, pp.828–839, 1997.
- [4] S. Hu, E.A. Hoffman, and J.M. Reinhardt, "Automatic segmentation of accurate quantitation of volumetric X-ray CT images," IEEE Trans. Med. Imaging, vol.20, no.6, pp.490–498, 2001.
- [5] M.N. Prasad and A. Sowmya, "Multi-class unsupervised classification with label correction of HRCT lung images," IEEE International Conference on Intelligent Sensing & Information Processing, pp.51– 56, 2004.
- [6] A. Shamsheyeva and A. Sowmya, "The anisotropic Gaussian kernel for SVM classification of HRCT images of the lung," IEEE Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp.439–444, 2004.
- [7] T.C. Poon and P.P. Banerjee, Contemporary Optical Image Processing, 1st ed., Elsevier Science, New York, 2001.