Quantification of Emphysema Severity by Histogram Analysis of CT Scans

Paulo R.S. Mendonça[1], Dirk R. Padfield[1], James C. Ross[1], James V. Miller[1], Sandeep Dutta[2], and Sardar Mal Gautham[2]

[1] GE Research Center, Niskayuna NY 12309, USA
[2] GE Healthcare, Waukesha WI 53188, USA

Abstract. Emphysema is characterized by the destruction and over distension of lung tissue, which manifest on high resolution computer tomography (CT) images as regions of low attenuation. Typically, it is diagnosed by clinical symptoms, physical examination, pulmonary function tests, and X-ray and CT imaging. In this paper we discuss a quantitative imaging approach to analyze emphysema which employs low-level segmentations of CT images that partition the data into perceptually relevant regions. We constructed multi-dimensional histograms of feature values computed over the image segmentation. For each region in the segmentation, we derive a rich set of feature measurements. While we can use any combination of physical and geometric features, we found that limiting the scope to two features – the mean attenuation across a region and the region area – is effective. The subject histogram is compared to a set of canonical histograms representative of various stages of emphysema using the Earth Mover's Distance metric. Disease severity is assigned based on which canonical histogram is most similar to the subject histogram. Experimental results with 81 cases of emphysema at different stages of disease progression show good agreement against the reading of an expert radiologist.

1 Introduction

Emphysema is a widespread chronic respiratory disorder, characterized by the loss of lung recoil resulting from the deterioration of lung tissue, which leads to physical limitations. The destruction and over-distension of lung tissue in emphysema are manifest on high resolution computer tomography (CT) images as regions of low attenuation, as seen in Fig. 1. While physicans are trained to identify such regions, it is often difficult to accurately quantify the extent of the disease severity, and both intra- and inter-operator variability is common [1]. Thus a variety of computer aided methods have been developed to automatically detect and quantify the extent of disease severity.

An early technique for measuring the amount of tissue degradation in the lung parenchyma is the Müller density index [2]. This technique calculates the percentage of the pixels in the lung field that have low attenuation. A global threshold defines what is considered low attenuation. This threshold must be set in a manner that takes into account variations in attenuation due to differences in scanners, scanning protocols, patient size, and the patients' ability to inflate their lungs.

While the Müller density index provides insight into the percentage of the lung that may be degraded by the disease, the interpretation is based strictly on the percentage of
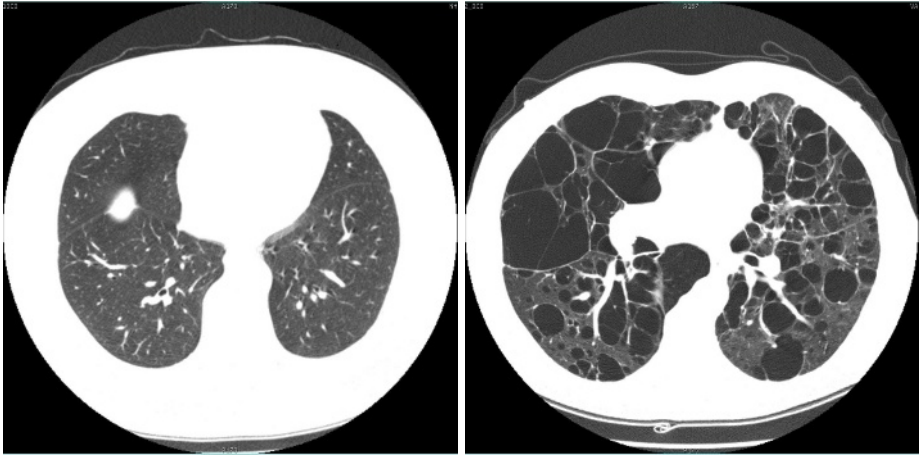
Fig. 1. The image on the left shows a CT scan of a healthy subject, which can be compared to the image of a severely emphysematous subject, shown on the right. The two diagnosis were established by and expert pulmonologyst.

low attenuation pixels and uses no information on the spatial distribution or grouping of the low attenuation pixels. Another method of analyzing the degree of lung degradation uses the fractal dimension of the low attenuation areas. The cumulative size distribution of the low attenuation areas has been shown to follow a power law relationship - with healthy subjects and emphysematous subjects having markedly different cumulative size distributions [3].

Coxson provides another extension to the Müller density index [4]. Here, the CT pixel values are converted to local measures of specific volume. This facilitates an interpretation of the volume of gas per mass of tissue in the lung parenchyma. Since emphysema manifests as a degradation of tissue (loss of mass) and a loss of elastic recoil (gain of volume), subjects with emphysema have higher volume of gas per unit mass ratios than healthy subjects.

Finally, emphysema has been analyzed using standard statistical pattern analysis approaches. The adaptive multiple feature method (AMFM) of Uppaluri [5] divides an image into a regular grid of regions and generates a series of derived measurements or features for each grid square. The features generated include local histograms, co-occurrence matrices, run length frequencies, and fractal attributes. Disease severity is determined using a Bayesian classifier on this large feature space. A main disadvantage of this approach is that the classifier must be manually trained with an expert assigning a classification to each grid square in each image in the training set. While the previous techniques looked at the percentage and distribution of low attenuation regions, the AMFM approach utilizes a richer feature set, in essence using more of the information present in the CT data.

The quantification approach presented here incorporates some of the previous technologies, extends others, and introduces new approaches for analyzing emphysema. In the following sections we will describe our classification technique in more detail and

provide quantitative results. Section 2 describes the segmentation and feature selection approach. Section 3 describes the metrics used for histogram comparison, and section 4 provides results computed on a dataset of CT volumes. Finally, in section 5 we will draw conclusions.

2   Feature Selection

In the problem of quantifying emphysema severity, a good deal is known about the data being analyzed, both in terms of anatomic characteristics of the lung and the physics of the scanner. This knowledge is incorporated in our feature extraction model, which includes both signal and shape models, to achieve the highest detection rates. These feature extraction methods are similar to those presented in [6].

In CT imagery, the signal models represent the blurring (point spread function) and noise characteristics inherent to the imaging system. These models permit the accurate detection of primitive events in the image such as intensity peaks and discontinuities. These primitives form the basis for our method's shape representation.

The first stage of shape representation consists of step-edge detection using the Canny edge detector [7]. To account for noise, each image slice is convolved with a two-dimensional Gaussian kernel in which the scale of the kernel can be optimally computed from the point spread function of the scanner [8].

After edge detection we seek to capture the appearance of features in the image in terms of their intensity and shape. This representation is formed by grouping the low-level features detected at the signal level. The edges from the preceding signal modeling stage are grouped into closed regions by tracing along the edges and following a path perpendicular to the edge gradient until a closed structure can be formed. The result is a connected network of vertices and edges delimiting adjacent image regions within
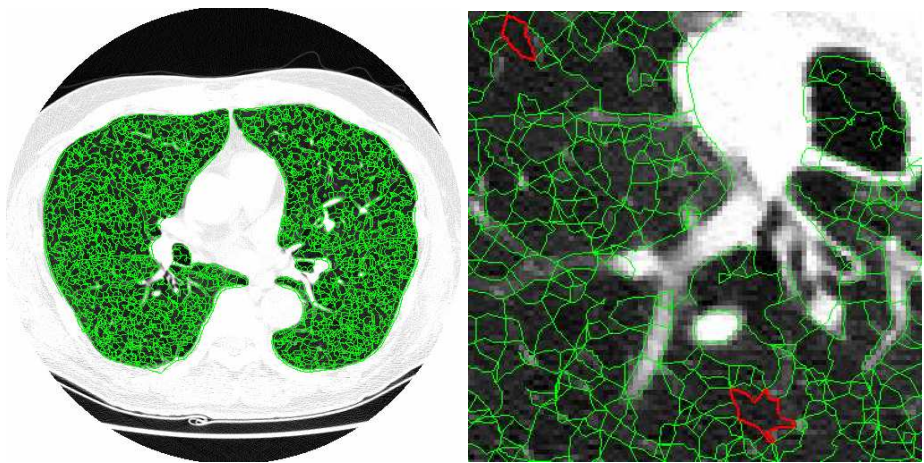
Fig. 2. Left: CT slice showing segmentation of the lung field. Right: close-up of the segmentation showing individual regions over which statistics can be computed.

which the voxel intensities are homogeneous [9]. The boundaries of these regions correspond to underlying tissue morphology, so that, for example, a large emphysematous area is represented by a single segmented region. Figure 2 illustrates a segmented CT slice and shows individual regions over which statistics can be computed.

In our application, we extract a variety of features for each region including intensity, area, number of edges, outer perimeter, diameter, intensity variance, min and max intensity, and many others. Of these, the first two, intensity and area, were chosen for histogram generation because they provide a logical characterization of disease severity. That is, as the severity of emphysema progresses, the lung tissue degrades, leading to larger and darker tissue patches.

3   Histogram Comparison

Using the low-level segmentation of the lung region, we construct two-dimensional histograms of area and mean intensity and compare these histograms to canonical histograms for various stages of the disease. These histograms are rich in information content. From these histograms, one can extract the percentage of the lung parenchyma occupied by low attenuation regions as in the Müller density index or analyze the size distribution of the low attenuation regions as in the fractal dimension approach. But the information in the histograms is richer than either of those approaches since it includes the size distributions for each attenuation value. Thus, in one representation you have the size distributions for the healthy parenchyma regions, the severely diseased regions, and the moderately diseased regions. The histograms are more relevant than the features extracted in the AMFM method because they are aligned with image events and not with an artificial grid.

To assess disease severity, we compare a subject's two-dimensional histogram to a set of canonical histograms. Each canonical histogram represents a different disease severity level, and each is derived from patient scans assessed by an expert pulmonologist, by averaging a set of histograms belonging to a given disease category as established by the expert. In order to compare the canonical histograms with the feature histogram of a new case several techniques can be employed, such as L_p metrics, chi-squre distance or Kullback-Leibler divergence. Motivated by the results in [10], we use the Earth's Movers Distance (EMD) as our metric. Intuitively, the EMD between two histograms is the minimum amount work needed to "reshape" one histogram into the other, normalized by the total amount of mass moved to perform the reshaping. The EMD depends both on the amount of mass that has to be moved as well as on the distance over which the transport occurs. In the situation when the distance between histogram bins is a metric [11] and the histograms have the same mass, the EMD is itself a metric. This is an important property in the application tackled here, since the triangular inequality prevents a given histogram from being simultaneously close to two others if those are themselves far apart. This should avoid the idiosyncratic situation in which the disease level of a subject is simultaneously deemed as close to high and low, as long as there is enough separation between these two classes themselves. In the current work the distance between bins was chosen to be the Euclidean distance between the bin locations.

4   Experimental Results

We generated two-dimensional histograms of region intensity and area for 81 CT data sets. Each data set was acquired with a GE scanner with 120 kVp, 1 mm slice thickness, 20 mm interslice distance, 200 mA tube current, 0.5 sec rotation, and x and y pixel spacing ranging from 0.547 mm to 0.813 mm. An expert pulmonologist examined and classified each case as one of five emphysema levels ranging from 0 (no signs of emphysema) to 4 (severe emphysema). The data sets were subgrouped as follows: 16 level 0 cases, 25 level 1 cases, 19 level 2 cases, 9 level 3 cases, and 12 level 4 cases.

For each case we constructed 400×400 two-dimensional histograms as described in section 2. Intensity ranged from -1024 HU to -220 HU with a bin size of 2.01 HU. Area ranged from 0 mm^2 to 338 mm^2 with a bin size of 0.8452 mm^2. Figure 3 shows area and intensity histogram marginals for a level 0 case and a level 4 case. Though the area marginals are quite similar to one another, the difference in the intensity marginals is clear.

We employed a bootstrapping technique to classify each of the 81 cases. That is, we generated canonical histograms for each severity level while withholding the case to be tested. Table 1 is a confusion matrix showing classification results. The rows indicate the subject cases, and the columns indicate the level at which the case was classified. If considering perfect classification, randomly choosing emphysema levels for each case yields an accuracy of 20%. Table 1 indicates that 62% of the level 0 cases, 20% of the level 1 cases, 21% of the level 2 cases, 33% of the level 3 cases, and 67% of the level 4 cases were perfectly classified, with a weighted kappa statistic of 0.468 (standard error of 0.074) [12, 13].

It is interesting to observe that although the classification rates for levels 1, 2, and 3 are somewhat low, they tend to be misclassified as a similar emphysema level. Therefore, if we instead consider a correct classification to be exact agreement or off-by-one agreement, results improve substantially. This is not an unreasonable approach given that cases with similar levels of severity have similar intensity and spatial characteris-
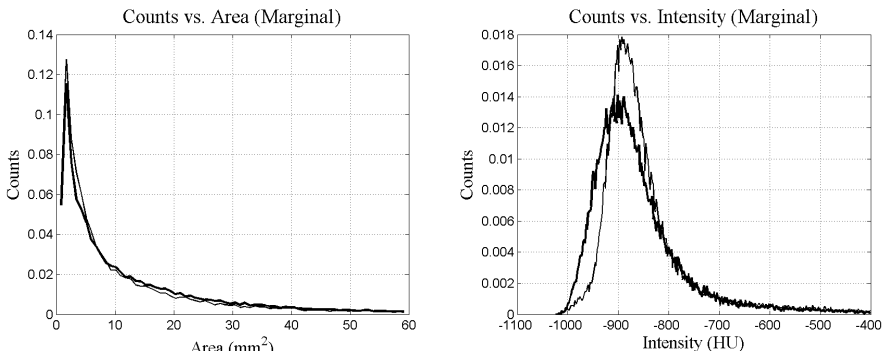
Fig. 3. Left: Area marginal histograms for a level 0 case and a level 4 case (bold). Right: Intensity marginal histograms for a level 0 case and a level 4 case (bold).

Table 1. Classification Confustion Matrix

	Level 0	Level 1	Level 2	Level 3	Level 4
Level 0	10	2	2	0	2
Level 1	9	5	9	2	0
Level 2	3	4	4	7	1
Level 3	1	1	1	3	3
Level 4	0	0	0	4	8

tics. In this case, random classification is 52%. Table 1 indicates that 75% of the level 0 cases, 92% of the level 1 cases, 79% of the level 2 cases, 78% of the level 3 cases, and 100% of the level 4 cases were classified correctly.

5 Conclusion

Emphysema is a serious disease leading to extensive physical restrictions for many individuals. In this paper we demonstrated a completely automated method of measuring emphysema severity. It is based on the observation that emphysema manifests in CT as large regions of low attenuation, properties captured on a two-dimensional joint histogram of the area and intensity of segmented regions in the lung parenchyma. We showed that our algorithm yields results in good agreement with manual classifications of an expert pulmonologist.

The automation of severity quantification addresses the issue of human intra- and inter-observer variability. Inter-observer agreement on emphysema scores for radiologists with varying degrees of expertise was reported in [1], with kappa values in the range of 0.431 to 0.589. The weighted kappa value for our algorithm (0.468) compares well with the range of values given in that work.

Although the results reported here are encouraging, improvements can be gained by employing more sophisticated classification schemes. In particular, we should consider a method that takes the variance of the histograms into account. Additionally, we chose features (mean intensity and area) based on clinical significance. However, because we compute a rich set of features for each segmented region, it would be interesting to validate this choice using a statistical feature selection technique. In principle three-dimensional segmentations would provide a more faithful representation of the true anatomical scenario. However, the imaging protocol may preclude accurate 3D segmentations owing to large interslice distances.

References

1. Bankier, A.A., De Maertelaer, V., Keyzer, C., Gevenois, P.A.: Pulmonary emphysema: Subjective visual grading versus objective quantification with macroscopic morphometry and thin-section ct densitometry. Radiology **211** (1999) 851–858
2. Muller, Staples, Miller, Abboud: Density mask: An objective method to quantitate emphysema using computed tomography. Chest **94** (1988) 782–787

3. Mishima, M.: Fractal analysis of emphysema in X-ray, CT and simulation. Medical Imaging Technology **18** (2000) 179–186
4. Coxson, Rogers, Whittal, D'Yachkova, Pare, Sciurba, Hogg: A quantification of the lung surface area in emphysema using computed tomography. American Journal of Respiratory and Critical Care in Medicine **159** (1999) 851–856
5. Uppaluri, R., Hoffman, E.A., Sonka, M., Hartley, P.G., Hunninghake, G.W., McLennan, G.: Computer recognition of regional lung disease patterns. American Journal Respiratory and Critical Care in Medicine **160** (1999) 648–654
6. McCulloch, C.C., Kaucik, R.A., Mendonça, P.R.S., Walter, D.J., Avila, R.S.: Model-based detection of lung nodules in computed tomography exams. Academic Radiology **11** (2004) 258–266
7. Canny, J.F.: A computational approach to edge detection. IEEE Trans. Pattern Analysis and Machine Intell. **8** (1986) 679–698
8. Elder, J.H., Zucker, S.W.: Local scale control for edge detection and blur estimation. IEEE Trans. Pattern Analysis and Machine Intell. **20** (1998) 699–716
9. Rothwell, C., Mundy, J., Hoffman, W., Nguyen, V.D.: Driving vision by topology. In: IEEE International Symposium on Computer Vision. (1995) 395–400
10. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. Journal of Computer Vision **40** (2000) 99–121
11. Sutherland, W.A.: Introduction to Metric and Topological Spaces. Clarendon Press, Oxford, UK (1975) 1998 reprint.
12. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20** (1960) 37–46
13. Cohen, J.: Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin **70** (1968) 213–220