Pre-Classification of Chest Radiographs for Improved Active Shape Model Segmentation of Ribs

Janakiramanan Ramachandran[1], Marios Pattichis[1], Peter Soliz[2]

[1]*Department of Electrical and Computer Engineering, University of New Mexico, NM*
[2] *Kestrel Corporation, Albuquerque,NM*

ramanan@eece.unm.edu, pattichis@eece.unm.edu, zilospx@rt66.com

Abstract

The parenchymal and skeletal structure as recorded on chest radiographs can vary significantly from person to person. The person's height, width, age, gender, and other factors will result in significant variations in the presentation of these structures. As a result, the application of an active shape model (ASM) for segmentation can be problematic. The segmentation task can be made easier if, before the creation of the model for the ASM, the chest x-rays are classified according to some measure of similarity. The ASM rib-parenchyma model is defined using the 14 chest radiographs from the International Labor Organization(ILO) set of standard chest x-rays. Groundtruth was established by manually segmenting the ribs for the standard x-rays. A "leave one out" procedure was used to perform the tests on the ILO set. The measure of success was defined by the sensitivity of the segmentation in correctly classifying a pixel as parenchyma, i.e. not rib (Sensitivity = True Positive / (True Positive + False Negative). In one experiment, the chest radiographs were first pre-classified according to the number of parenchymal regions visible in each lung, and consecutive ASM model training and testing was performed over each class. In a second, similar experiment, the x-rays were first pre-classified by the height of the lungs, e.g. "tall" and "short" x-rays. It was found that the sensitivity improved an average of 0.20 over the baseline test (no pre-classification), when pre-classifying the lung by the number of parenchymal regions (8 or 9). Sensitivity improved an average of 0.19 over the baseline, when pre-classifying the lung by height (tall or short).

Keywords: active shape model, chest radiographs, segmentation, ribs, parenchyma

1. Introduction

The long-term goal of this research is to develop a computer-aided diagnostic system for screening and classifying chest radiographs presenting with pneumoconioses. One of the critical steps in achieving this goal is the development of a robust and accurate rib/parenchyma segmentation system. A number of investigators have achieved some success in the application of active shape models for segmentation of medical images [1]. An active shape model segments an object based on the matching of local features. Cootes, et al. [2] applied the concept of local matching with global shape constraints to produce the active shape model adopted for this study.

In this study, the chest x-rays that are the object of the segmentation come from a cohort of miners who have been exposed to silica dust particles. The ASM model is created from the ILO standard chest radiographs. The standards were created in 1980 by the International Labor Organization (ILO) to implement a "reading" system [3] and to standardize the interpretation of chest radiographs presenting with pneumoconioses. The scheme was designed specifically to systematically record radiographic changes that reflect the inhalation of dust [1, 4] in order to assess the progression of the disease through the objective, though qualitative, evaluation of abnormalities in the chest x-rays. Chest radiographs have 8 or 9 parenchyma visible on each lung. If a training data was used which did not pre-classify the radiographs, the ASM model found segmentation to be tedious. A pre-classification system was therefore designed to aid the model to segment more easily based on separate training data which was based number of parenchymas.

2. Materials and Methods

2.1. Data Set

The ASM model for all the tests was derived from 14 "standard" chest x-rays. The ILO standard x-rays consist of 14 films that are representative of the three profusions (1, 2, and 3) and the six size and shapes of the opacities (p ,q, r for rounded and s, t, u for irregular). Archived posterior-anterior chest radiographs from the Miners' Colfax Medical Center (MCMC) Outreach Program were used to test the pre-segmentation procedures. A set of over 500 radiographs was selected from the Miners' Colfax Medical Center Outreach Program database and digitized. The radiographs that were selected did not reflect the distribution of the larger population, which was heavily

0-7695-1537-1/02 $17.00

biased toward the 0/0 (normal) profusion category. The research data set contained a greater percentage of "non-normal" radiographs (*i.e.*, profusion category of 0/1 or higher) than was contained in the overall population. In addition, a larger number of 0/1 and 1/0 films were selected than is reflected in the original population distribution. This distribution was decided because these low profusion categories are the most problematic for the readers in classifying films into these categories.

2.2 Ground Truth

A trained medical research assistant performed the manual segmentation of the ILO standards and 50 chest radiographs from the database. The manually segmented x-rays were reviewed by two B-certified radiologists. The ground truth was done manually, where in each parenchyma region 16 points were carefully chosen so that the locations can be identified in each lung. Example points include extremities in the parenchymal regions, points where posterior and anterior ribs met. The lungs were divided into four parenchymal regions, two for the left and two for the right lungs. Each parenchymal region included four parenchymas.

2.3 Experiment

Sensitivity was used as the benchmark to assess the results of the pre-classifications tests. Sensitivity was based on the number of pixels correctly segmented as parenchyma, true positive (TP) and the number of pixels missed as part of the parenchyma, false negative (FN). The sensitivity was computer using:

$$\text{Sensitivity} = TP / (TP + FN).$$

Expt 1. Segmentation of the ILO Chest Radiographs

The first experiment was conducted on all the 14 ILO standards. Every experiment consists of a training phase and a testing phase.

In this experiment the leave-one out method was used for training and testing. For training, 13 out of the 14 radiographs were used, and the trained system was tested on the remaining ILO chest radiographs. The experiment was repeated 14 times, once for each x-ray, and the test results were averaged.

Expt 2. Segmentation of the 8-9 parenchymal ILO Chest Radiographs

A second experiment was conducted whereby the 14 x-rays were manually classified by the number of parenchymal regions visible in the x-ray. Either 8 or 9 parenchymal regions per lung are seen in the typical chest radiographs. The purpose of the experiment was to calculate the improvement in sensitivity after pre-classification into 8 and 9 parenchmal regions.

Expt 2(a) Segmentation of chest radiographs with 8 parenchymal regions

The leave-one out method as described in Experiment 1 was used for seven ILO chest radiographs that had 8 parenchymal regions.

Expt 2(b) Segmentation of chest radiographs with 9 parenchymal regions

The leave-one out method was used for seven ILO chest radiographs that had 8 parenchymal regions.

Expt 3. Optimal Classification of Number of Parenchymal Regions based on lung height

The third experiment was conducted in order to develop a simple method of classifying chest radiographs into ones that have 8 versus 9 parenchymal regions. To this end, the height of the lung, defined as the vertical distance between the topmost and bottom most point of the lung was used.

For optimal classification based on height, the threshold that resulted in the minimum probability of error was computed for 200 chest radiographs.

Expt 4. Segmentation of the Tall/Short parenchymal ILO Standards

The fourth experiment divided the 14 x-rays into "tall" and "short" lungs according to the optimal threshold computed in experiment 3. The segmentation experiment was repeated.

The purpose of the experiment was to calculate the improvement in sensitivity after pre-classification based on the height of the lungs.

Expt 4(a) Segmentation of the short lung ILO Chest Radiographs

The leave-one out method was used for seven ILO chest radiographs that had short lungs.

Expt 4(b) Segmentation of the short lung ILO Chest Radiographs

The leave-one out method was used for seven ILO chest radiographs that had tall lungs.

Expt 5. Segmentation of the Tall/Short parenchymal 50 images from database

The fifth experiment was to segment 50 images from the database and determine their sensitivities based on the height-based classification.

Expt 5(a) Segmentation of 50 Short lung radiographs

Training phase: All 7 short ILO chest were used in the training phase.

Testing phase: 17 short images from a non-ILO database to segment the parenchyma.

Expt 5(b) Segmentation of 50 Tall lung radiographs
Training phase: All 7 ILO chest radiographs which were tall (above optimal threshold) were used in the training phase.
Testing phase: 33 short images from a non ILO database to segment the parenchyma.

Expt 5(c) Segmentation of all radiographs

Training phase: All 14 ILO chest radiographs were used in the training phase.

Testing phase: The training data was used on each of the 50 images from the database to segment the parenchyma.

3. Results

Figure 1 shows the results of the ASM segmentation using 13 of the ILO standard x-rays for the model and testing the segmentation of standard 1/1 t/t (expt 1). Figures 2 and 3 show the segmentation of the same image using the pre-classified x-rays, by number of parenchyma (8 or 9, expt 2) and by height (short or tall, expt 4), respectively. This case was selected because it represents one where the significant improvement as measured by the sensitivity was seen by pre-classifying the x-rays before creating the model for the ASM.

Table 1 summarizes the calculated sensitivities for the three segmentation approaches as applied to the 14 x-rays using the leave one out technique for the tests. The average improvement of the segmentation based on number of parenchyma (expt 2) over the baseline approach is 0.20% (expt 1). The average improvement of the segmentation based on the height is 0.19% (expt 4).

The third experiment aimed at finding the optimal height threshold for estimating the number of parenchymal regions. The optimal threshold was selected to be the height value were the minimum probability of classification

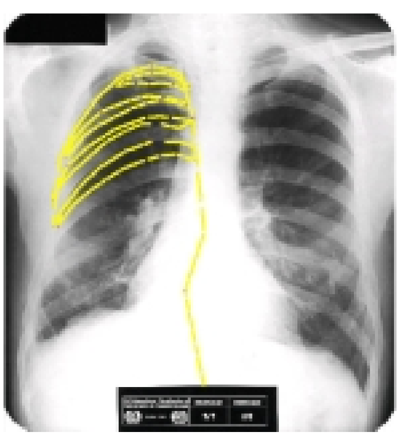

Figure 1. Baseline segmentation of ILO standard x-ray 1/1 t/t. Sensitivity of segmentation was 0.41 (expt 1).

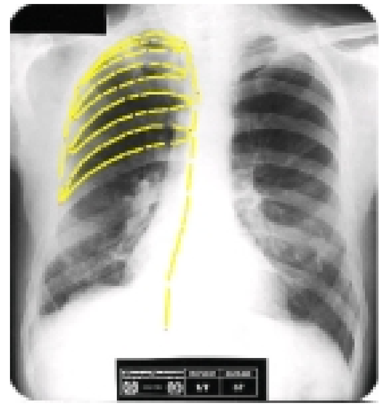

Figure 2. Number of paren-chyma-based segmentation of ILO standard x-ray 1/1 t/t. Sensitivity of segmentation was 0.93 (expt 2).

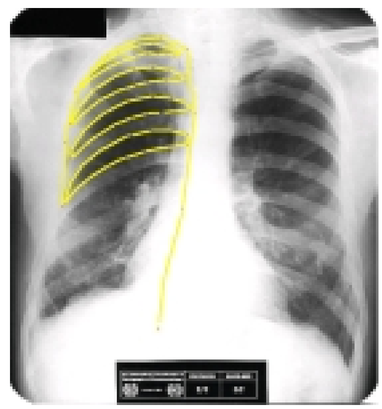

Figure 3. Height-based segmentation of ILO standard x-ray 1/1 t/t. Sensitivity of segmentation was 0.92 (expt 4).

error was measured. The procedure is summarized in Figure 4. In 4(a), the height distribution of all 8-parenchyma radiographs is shown. In 4(b), the height distribution of all 9-parenchyma is shown. The probability of incorrect parenchyma detection based on the height of the lung is shown in Figure 4(d).

At 11 inches, the probability error was the lowest and that was then used as the optimal threshold for pre-classifying chest radiographs into tall and short (assumed to contain 9 and 8 parenchymal regions respectively). Setting the threshold at 27.5cm (11 inches) for the 200 chest radiographs which were used for this experiment, 192 of them had height – number of parenchymal regions coincidence. This means that the short radiographs had 8 parenchymal ones and the tall radiographs presented 9

Fifth IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'02)
0-7695-1537-1/02 $17.00 © 2002 IEEE

parenchymal regions. So 96% of the radiographs had height that can be used to predict the number of parenchymal regions correctly. This is an extremely useful result because it is easier to automate the height and width calculation than attempting to develop automated methods for counting the number of parenchyma.

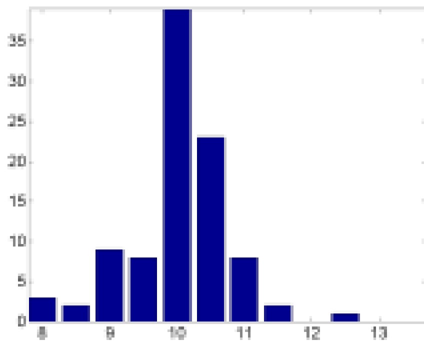

Figure 4(a) Height distribution of chest x-rays with 8 parenchyma (expt 3)

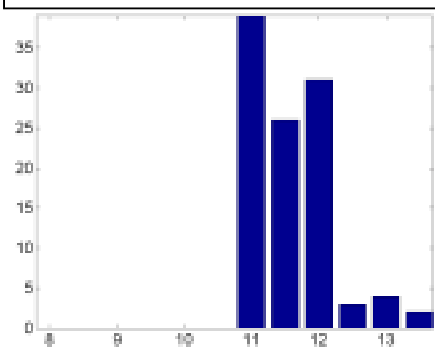

Figure 4(b) Height distribution of chest x-rays with 9 parenchyma (expt 3)

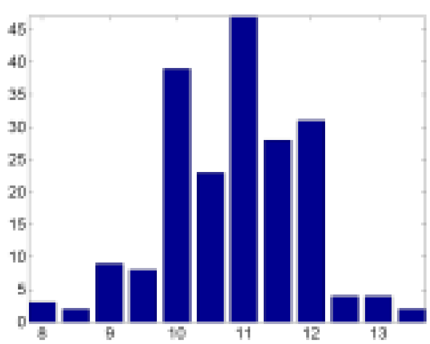

Figure 4(c) Height distribution of chest x-rays with all parenchyma (expt 3)

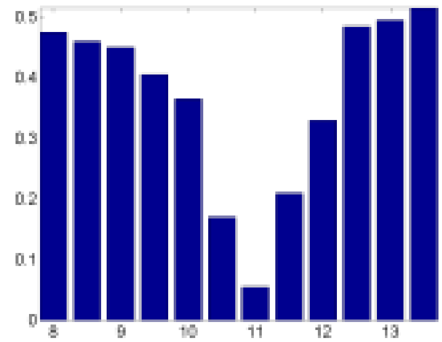

Figure 4(d) Histogram of the probability error (expt 3)

In experiment 4, the height based pre-classification greatly aided segmentation, as it is evident from the results, shown in table 1. The mean sensitivity was 0.86 for the upper right parenchymal region. The same protocol was applied for the other three regions of the lungs. The mean sensitivities for the upper left region was 0.82, for the lower right region, it was 0.90 and for the lower left region, it was 0.87.

In experiment 5, the segmentation was performed on 50 chest radiographs from a non-ILO database. The new sensitivities after pre-classification were far better than the old sensitivities before pre-classification. For the entire set of 50, there was an average increase of sensitivity of 0.15.

Table 1. Sensitivity of segmentation for 14 ILO standard chest radiographs. The baseline value sensitivity is given. Change in sensitivity from the baseline is given for the segmentation procedure: The results are reported for the Upper right parenchymal region(see section 2.2)

X-ray ID	Baseline Expt 1	# parenchyma Expt 2	Height Expt 4
00	**0.73**	**+0.11**	**+0.12**
11 pp	**0.83**	**+0.02**	**0.00**
11 qq	**0.49**	**+0.37**	**+0.37**
11 tt	**0.41**	**+0.52**	**+0.51**
11 st	**0.16**	**+0.69**	**+0.68**
22 pp	**0.94**	**+0.02**	**-0.01**
22 qq	**0.72**	**+0.10**	**+0.09**
22 rr	**0.71**	**+0.17**	**+0.18**
22 ss	**0.66**	**+0.18**	**+0.15**
22 tt	**0.56**	**+0.24**	**+0.24**
33 pp	**0.75**	**+0.05**	**+0.05**
33 qq	**0.93**	**-0.01**	**-0.01**
33 ss	**0.74**	**+0.15**	**+0.11**
33 tt	**0.68**	**+0.19**	**+0.18**

Fifth IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'02)
0-7695-1537-1/02 $17.00 © 2002 **IEEE**

Table 2: Statistics for the 50 radiographs

	Old Sensitivity-expt 5(c)	New sensitivity-expt5(a),(b)
Minimum	0.12	0.70
Maximum	0.92	0.95
Mean	0.64	0.79
Median	0.68	0.80
Standard-Deviation	0.1925	0.0599

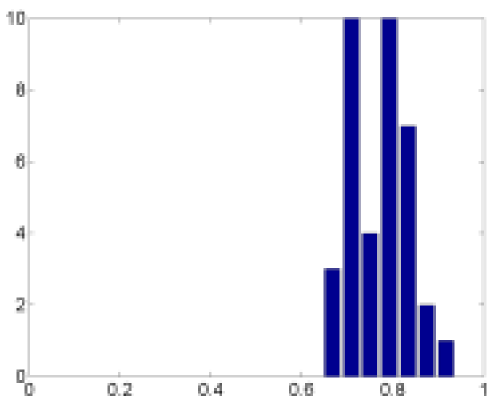

Figure 5a: Histograms of new sensitivity. The x axis is the sensitivity and the y axis is the frequency(expt 5(c))

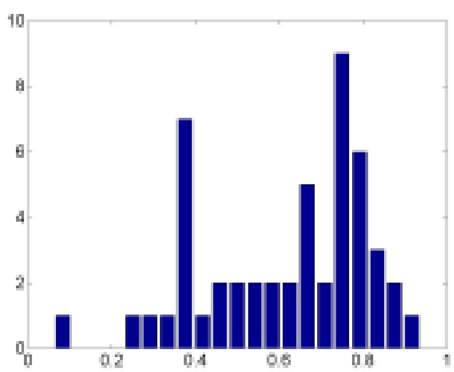

Figure 5b: Histograms of old sensitivity. The x axis is the sensitivity and the y axis is the frequency(expt(5(a)+(b))

Figures 5(a)-(c) show the results of experiment 5 in the form of histograms of old (non pre-classified) and new (pre-classified) sensitivities of the 50 radiographs. Table 2 displays the statistics between the old and new sensitivities of the 50 radiographs

4. Discussion

Clearly, ASM segmentation is improved significantly by pre-classifying the x-rays. Without extensive or complex processing, the segmentation can be improved significantly. The improvement was especially noticeable in the x-rays like 11 tt and 11 st where the baseline ASM segmentation was off by one parenchymal region. A hypothesis was posed that giving the ASM model greater global information, such as the number of parenchyma in a lung would improve the local fitting. Experimental results verified that this was indeed the case. Because it is easier to implement a height measurement algorithm instead of a method for counting parenchymal regions, the fourth experiment was performed. The results of the ASM segmentation using height-based model are very close to those using the number of parenchyma-based model. A simple technique is being developed to automate the pre-classification based on height. A third experiment was used to estimate an optimal threshold for using the height parameter associated with the number of parenchymal regions.

Acknowledgements

The authors wish to acknowledge the Miners' Colfax Medical Center (MCMC) Outreach Program for providing the chest radiograph database. Ms Ana Edwards painstakingly provided the ground truth analysis. Dr. David James (MD) and Dr. Loren Ketai (MD) provided the medical diagnosis and advice on the research. Funding for the study was through a grant from the National Institute for Occupational Safety and Health (NIOSH), Grant 2R44 OHRRGM 03595.

References

[1] U. Grenander, Y. Chow, D. Keenan, HANDS. *A pattern theoretical study of biological shapes*, Springer Verlag, New York, 1991. 96

[2] T.F. Cootes, C.J. Taylor, D. Cooper, J. Graham, "*Active shape models their training and application*", Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38.

[3]*Guidelines for the use of ILO international classification of radiographs of pneumoconioses*, 1980, International Labour Office: Geneva.

[4]Epler, G.R., *Chest films: "Underused tool in interstitial disease?"*. J Respir Dis, 1997(18): p. 168-180.

0-7695-1537-1/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY