Computer-aided detection of interstitial lung diseases: A texture approach

Tomas Plankis, Algimantas Juozapavičius, Eglė Stašienė, Vytautas Usonis

Faculty of Mathematics and Informatics, Vilnius University,
Naugarduko str. 24, LT-03225 Vilnius, Lithuania,
tomas.plankis@mif.vu.lt

Abstract. We have developed the flexible scheme for computer-aided detection (CAD) of interstitial lung diseases on chest radiographs. These schemes enable us to perform diagnostics in the broad circumstances of pneumonia and other interstitial lung diseases. It is applied in the case of children pneumonia when conditions are difficult to standardize. In the adults' case, the schemes of CAD are more adaptive since there are more characteristic interstitial lung tissue's changes to all kinds of pathological conditions. Even in the norm of drawing, there are more visible and more highlighted features leading to better results. The CAD scheme works as follows. For the first of all, we are using adopted algorithms of active contours to select the area of lungs and then divide this area into subareas – regions of interest (40 different ROI). Then ROIs were subjected to the 2-dimensional Daubechies wavelet transform, and only main transformation was used. For every transformation, 12 texture measures were calculated. Principal component analysis (PCA) was used to extract 2 main components for each ROI, and these components were compared to predictive component region.

Keywords: automatic lung disease recognition, image segmentation.

1 Introduction

The radiographs are the most common examination tools in medical practice. It plays an important role in the diagnosis of pulmonary diseases, in radiography formed by fluorescence (used chiefly in biomedical research), in radiographs clinical role in chiropractic, in diagnosing of tuberculosis, etc. [5, 6, 8, 16]. The chest radiography is also useful and popular routine because it is cost effective and produces low dose of radiation [6].

Chest radiography is used to diagnose pulmonary diseases such as lung cancer (nodules), viral pneumonia, bacterial pneumonia, bronchopneumonia, interstitial diseases, chronic obstructive pulmonary disease. Diseases are often indicated by the analysis of lung textures, searching for specific formations inside lungs. However, it is difficult for

radiologists to detect and characterize lesions on a chest radiograph when those lesions are low in contrast and/or overlap with ribs and large pulmonary vessels. Usually, these patterns are diffuse and typical for a large number of regions covering the entire unobscured lung fields or maybe peripheral areas. From the image processing point of view, this is usually considered a texture analysis and recognition problem [1].

Current research efforts have been focused on detection and classification of images of many different types of lesions in a number of organs obtained with various imaging modalities. It is likely that the present results of CAD are only at the tip of the iceberg. Although automated computer diagnosis is a concept based on computer algorithms only, CAD is a concept established by taking into account equally the roles of physicians and computers. The effect of CAD on differential diagnosis has already indicated that the performance level is high and that CAD would be ready for clinical trials and commercialization efforts. The presentation of images similar to those of an unknown case may be useful as a supplemental tool for CAD in the differential diagnosis. It is necessary to note that the concept of CAD made by a radiologist is different from that of an automated diagnosis made by a computer. The development of CAD methodology includes not only the development of computer algorithms for detection and classification of chest radiographs, but also the investigation of effects of the computer output on radiologists [2, 7, 13].

In order to implement CAD methodology, the appropriate computer system was developed. The task for such system was to analyze images of radiography in DICOM format and recognize lungs, as well as its texture patterns, calculating appropriate characteristics. Main steps of the system are: a) recognize and control radiograph images in order to analyze later lung textures, b) perform Daubechies wavelet transform, c) compute 12 measure texture parameters based on wavelet coefficients, d) produce supervised machine learning procedure in order to give recommendation for clinical decisions, e) evaluate statistical reliability of decisions having enough instances.

The article is organized as follows. Besides the introduction, where main principles of dealing with lung texture are given, Section 2 defines the correspondence of lung textures of interest to lesions. Section 3 is defining and calculating the region of interests of texture analysis and recognition (lung segmentation). Section 4 deals with automated disease recognition, while Section 5 are discussed on results and conclusions. References of the article include 16 items.

2 Texture recognition, segmentation and analysis

Interstitial lung disease is a general category that includes many different lung conditions, like of the types below:

- Bacteria, viruses, or fungi may infect the interstitium of the lung.
- Idiopathic pulmonary fibrosis: A chronic, progressive form of fibrosis (scarring) of the interstitium. Its cause is unknown.
- Nonspecific interstitial pneumonitis: Interstitial lung disease that is often present with autoimmune conditions (such as rheumatoid arthritis or scleroderma).

- Hypersensitivity pneumonitis: Interstitial lung disease caused by ongoing inhalation of dust, mold, or other irritants.
- Cryptogenic organizing pneumonia (COP): A pneumonia-like interstitial lung disease, but without an infection present. COP is also called bronchiolitis obliterans with organizing pneumonia (BOOP).
- Acute interstitial pneumonitis: A sudden, severe interstitial lung disease often requiring life support.
- Desquamative interstitial pneumonitis: An interstitial lung disease that is partially caused by smoking.
- Sarcoidosis: A condition causing interstitial lung disease along with swollen lymph nodes, and sometimes heart, skin, nerve, or eye involvement.
- Asbestosis: Interstitial lung disease caused by asbestos exposure.

All interstitial lung diseases affect the interstitium, a part of the lungs' anatomic structure. Normally, the interstitium is so thin that it is not seen in details on chest X-rays. However, the interstitium caused usually a specific texture, which is subject to study.

Many approaches in CAD for chest radiology relay on texture analysis. Texture parameters, for the first of all, are expected to increase information obtainable from chest radiograph. Also, texture analysis is capable to present new forms of texture. The applications are ranging from segmentation of specific anatomical structures and detection of lesions, to differentiation between pathological and healthy tissue in lung and other organs.

3 Lung segmentation

There are various methods to segment lungs with similar results [4,9]. In this paper Active Shape Model (ASM) segmentation method was used. The description and notation will follow as in [3, 15]. The parameters of the method are listed in Table 1. All images were scaled to the same size of 700×600 for higher consistency. To simplify method rotation of the shape was omitted.

Table 1. ASM parameters.

Parameter	Description	Value
s	Number of training images	14
Shape model		
n	Number of landmarks points in profile	571
Alignment	To use alignment or not	yes
f_v	Variance to be explained	$> 90\%$
m	Bounds on eigenvalues λ_i	3
Appearance model		
k	Number of points on either side of the landmark point Perpendicular to contour	16
$L_{\max}$	Number of resolution levels	2
Search algorithm		
n_s	Number of new positions to evaluate for each iteration	6
$N_{\max}$	Number of iterations for resolution level	40

3.1 Algorithm

ASM algorithm was implemented in JAVA programming language. The basis for implementation was algorithm of Kroon [11] with some modifications to improve performance. The new landmark is put at the position with the lowest distance. Mahalanobis distance is commonly used to calculate edge gradient. In this work, both Mahalanobis distance and PCA parameters to calculate intensities were used. As shown in Table 2, PCA parameters show better results. Thus, it was decided to use it. Last but not least, lungs model was placed statically.

Table 2. Visual comparison of distance methods.

PCA	Mahalanobis	Comparison ($\|$PCA $-$ Mahanalobis$\|$)
acceptable	poor	1
acceptable	poor	1
good	poor	2
good	poor	2
acceptable	poor	1
good	poor	2
good	acceptable	1
good	acceptable	1
good	poor	2
acceptable	poor/acceptable	0
acceptable	poor	1
acceptable	acceptable	0
acceptable	acceptable	0
good	poor/acceptable	1

Good ($= 4$) – small variance (contour is almost in tact); acceptable ($= 3$) – higher variance, but still acceptable (some edge distortions, but in general, contour repeats lung shape); poor ($= 2$) – variance too high (lungs detected, but edge is out of place); bad ($= 1$) – not recognized.

3.2 Segmentation

We took 28 various real life images to test lung detection with ASM method. One problem with raw images was their contrast. Thus, we used adaptive histogram equalisation (Matlab version) to see if that makes some difference, but results are ambiguous.

It was noticed that there is a group of 7 problematic images – the ones of infants. Detection of lungs is almost always bad in such a case because infants cannot stand on their own, not to mention fulfilling other requirements for making good chest radiograph. Thus, their chest view is always irregular, making it hard for algorithm to adapt shape model. Also, there were 2 chest radiograph images of different kind, which also showed bad detection. The segmentation of 4 images from the 19 left ones was bad, leading to $\approx 79\%$ of detection rate. Most problems are related with recognizing top or bottom of the lungs. That is related to two cases:

- Model was placed too low. Which leads rib to be recognized as top of the lung.
- Model was placed too high. Which gives misleading information for ASM, and model do not converge there.

To solve this problem, we need to place model dynamically. The idea is to pre-calculate lungs position and size. Size will be needed to scale the model, and position will be used to place model at about right position. That should solve problems with lungs size and placement on image. One of the possible approaches is described in [12]. Though deformable models are used to finally segment lungs, but there is an algorithm to detect region of interest (the place of lungs) used.

4   Automatic disease recognition

In this section, we describe the idea for automatic disease recognition. According to paper [14], there are a lot of different methods to detect diseases. The idea of automatic disease recognition is somewhat similar to the one in paper [10].

Let us take K radiographs for training. First, we segment lungs. Then we divide lungs into $N \times M$ regions like in Fig. 1. For every region X, it is possible to calculate its features whose can be used for further classification of images. Also, it is possible to make few sets of features to represent different diseases.

In this research, 14 chest radiographs of healthy patients was used. Every lung was divided into $10 \times 4 = 40$ rectangles whose can be called ROIs (regions of interest). Thus, for the same ROI position, we get 14 sub-images.

To calculate features, the next method was used:

- ROI was processed with the two-dimensional Daubechies wavelets. Only the main transformation was used for further calculations. For this step, JWave (`https://github.com/cscheiblich/JWave`) library was used.
- For transformation, twelve texture measures shown in Table 3 were calculated. Let us say that x iterates rows and y iterates columns of ROI. $I(x, y)$ will mean pixels value at (x, y). Only nonzero value pixels are used for calculation. Texture measures form a 12-dimension data array, which leads to forming the data matrix of the size 14×12.
- For such data matrix, the standard PCA was used to extract 2 main components. These components were used to form the confidence region for ROI. Mathematical
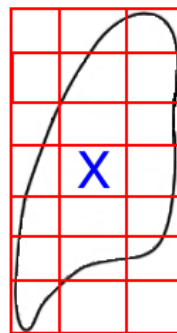

Figure 1. Lung's grid.

part was calculated using Efficient Java Matrix Library (`http://ejml.org`). As to how much variability these two components explained, it is hard to answer. Overall there is 80 ROIs, so there will be 80 values. Eigenvalues for some ROIs can be zeros as they are composed from only black pixels, but for others may vary. Evaluation of random ROIs gives at least 99.9% variability for each of them.

For new radiograph, two main components for each ROI can be calculated and compared to confidence region. In Fig. 2, there are shown only those ROIs whose components do not fall into confidence regions. According to doctors who checked processed radiographs, in some marked ROIs, there are irregularities, but not necessary a disease.

Table 3. Textures measures.

Energy	$\frac{1}{N}\sqrt{\sum_{x,y}(I(x,y)^2))}$
Contrast	$\frac{1}{N}\sum_{x,y}(I(x,y)\cdot(x-y)^2)$
Entropy	$-\frac{1}{N}\sum_{x,y}(I(x,y)\cdot\log I(x,y))$
Homogeneity	$\frac{1}{N}\sum_{x,y}(I(x,y)/(1+(x-y)^2)$
Homogeneity (Matlab)	$\frac{1}{N}\sum_{x,y}(I(x,y)/(1+|x-y|))$
Dissimilarity	$\frac{1}{N}\sum_{x,y}(|x-y|\cdot I(x,y))$
Max probability	$\max(I(x,y))$
Min probability	$\min(I(x,y))$
Mean	$\frac{1}{N}\sum_{x,y}I(x,y)$
Deviation	$\frac{1}{N}\sqrt{\sum_{x,y}((I(x,y)-\text{Mean})^2}$
Max row energy	$\max_x\left(\sum_y I(x,y)\right)$
Max col energy	$\max_y\left(\sum_x I(x,y)\right)$
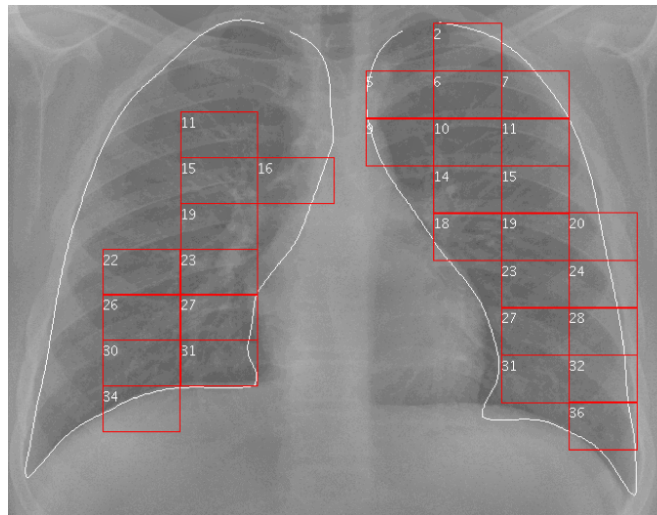
Figure 2. Examined lungs.

5   Results and conclusions

In this work, the algorithm for lungs detection was implemented. Algorithm to detect if patient is healthy was constructed.

Few problems remain:

- How big the set of radiographs for algorithm's training should be?
- It is necessary to implement a part of algorithm to detect diseases. At the moment, it is still not clear how to do it for every ROI as diseases not necessarily covers all lungs.
- It is necessary to improve lungs segmentation. It is possible that improved segmentation also solve age problem.
- It is not clear how many ROIs there should be, so it is necessary to find optimal division of lungs.
- Is it necessary to process image with 2-dimensional Daubechies wavelet transform?
- How many textures measures are important? Maybe it is possible to use less of them and get the same or better results.
- Evaluate method's performance.

References

1. A.R. Aswar, K.P. Nagarikar, P.D. Pawar, Texture-based identification and characterization of pneumonia patterns in lung, *International Journal of Application or Innovation in Engineering and Management*, **3**(12):284–290, 2014.

2. B.J. Bartholmai, S. Raghunath, R.A. Karwoski, T. Moua, S. Rajagopalan, F. Maldonado, P.A. Decker, R.A. Robb, Quantitative computed tomography imaging of interstitial lung diseases, *J. Thorac. Imag.*, **28**(5):298–307, 2013.

3. T.F. Cootes, C.J. Taylor, Statistical models of appearance for computer vision, 2004, `http://www.face-rec.org/algorithms/AAM/app_models.pdf`.

4. H. Das, A. Nath, An efficient detection of tuberculosis from chest X-rays, *International Journal of Advance Research in Computer Science and Management Studies*, **3**(5):149–154, 2015.

5. K. Doi, Overview on research and development of computer-aided diagnostic scheme, *Semin. Ultrasound CT*, **25**(5):404–410, 2004.

6. K. Doi, Current status and future potential of computer-aided diagnosis in medical imaging, *The British Institute of Radiology*, **78**(1):S3–S19, 2005.

7. K. Doi, Computer-aided diagnosis in medical imaging: Historical review, current status and future potential, *Comput. Med. Imag. Grap.*, **31**(4-5):198–211, 2007.

8. K. Doi, H. MacMahon, S. Katsuragawa, R.M. Nishikawa, Y. Jiang, Computer-aided diagnosis in radiology: Potential and pitfalls, *Eur. J. Radiol.*, **31**(2):97–109, 1999.

9. S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R.K. Singh, S. Antani, G. Thoma, Y.X. Wang, P.X. Lu, C.J. McDonald, Automatic tuberculosis screening using chest radiographs, *IEEE Trans. Med. Imaging*, **33**(2):233–245, 2014.

10. E.F. Kao, Y.T. Kuo, J.S. Hsu, M.C. Chou, G.C. Liu, Zone-based analysis for automated detection of abnormalities in chest radiographs, *Med. Phys.*, **38**(7):4241–4250, 2011.

11. D. Kroon, Active shape model (ASM) and active appearance model (AAM), 2012, `http://www.mathworks.com/matlabcentral/fileexchange/26706-active-shape-model--asm--and-active-appearance-model--aam-.`

12. W.L. Lee, K. Chang, K.S. Hsieh, Unsupervised segmentation of lung fields in chest radiographs using multiresolution fractal feature vector and deformable models, *Med. Biol. Eng. Comput.*, **54**(9):1409–1422, 2016.

13. D.S. Morillo, A.L. Jiménez, S.A. Moreno, Computer-aided diagnosis of pneumonia in patients with chronic obstructive pulmonary disease, *J. Am. Med. Inform. Assn.*, **20**(1):111–117, 2013.

14. M.M. Theresa, V.S. Bharathu, A survey on CAD technique for various abnormality classification in chest radiography, *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, **7**(4):331–342, 2016.

15. B. van Ginneken, A.F. Frangi, J.J. Staal, B.M ter Haar Romeny, M.A. Viergever, Active shape model segmentation with optimal features, *IEEE Trans. Med. Imaging*, **21**(8):924–933, 2002.

16. B. van Ginneken, B.M. ter Haar Romeny, M.A. Viergever, Computer-aided diagnosis in chest radiography: A survey, *IEEE Trans. Med. Imaging*, **20**(12):1228–1241, 2001.