IFMBE Proceedings Series – Lung nodule detection on rib eliminated radiographs

G. Orbán[1], Á. Horváth[1] and G. Horváth[1]

[1] Budapest University of Technology and Economics/Department of Measurement and Information Systems, Budapest, Hungary

Abstract— **A lung nodule detection algorithm was developed and tested against a comprehensive set of radiographs. The aim of the study was to validate the usefulness of two new algorithms in the detection scheme. The first one is a preprocessing step eliminating ribs and collarbones on the image to enhance the visibility of nodules. The other new algorithm is the Constrained Sliding Band Filter (CSBF) that raises the intensity of round shaped objects while suppressing other areas. After consecutively using these algorithms the suspicious areas are processed by a Support Vector Machine (SVM) based on mostly textural features to reduce the number of false detections. The algorithms were tested on the public database created by the Japanese Society of Radiological Technology (JSRT) and a private database. The new methods showed promising results, while the overall performance, 61% sensitivity at 2.5 false positives per image is comparable with state-of-the art algorithms.**

Keywords— **Nodule detection, rib elimination, chest radiographs, CAD, sliding band filter**

I. Introduction

Lung cancer is one of the most common causes of cancer death. Many cures are known, but most of them are effective only in the early and symptomless stage of the disease. Screening can help early diagnosis, but an accurate, cheap and side effect free method has to be used to enable mass usage. Standard chest radiography mostly meets these requirements, except that current methods have a moderate accuracy. Efficiency can be improved by analyzing the radiographs using a Computer Aided Detection (CAD) system. Current CAD applications can only be used as a second reader as they mark several suspicious areas on the radiographs and the examiner has to determine the real nodules.

The most important problem of existing CAD systems is the high number of false detections. Although they can detect 60-70% of cancerous tumors, they also mark approximately four healthy regions on each image [1]. Despite that radiologists are able to filter false detections, observer studies show that the number of false positive diagnoses increases with CAD, showing that current methods have to be improved.

For our solution we used the following three step scheme. The first step frees the image from unnecessary objects and noise, thus making the nodule more visible. The next step enhances round shaped objects like the target lung nodules by using image processing algorithms. The last step selects suspicious areas on the enhanced image with the help of a classifier.

II. Materials and methods

A. Elimination of bone shadows

For eliminating ribs and clavicles they are first segmented based on a previously calculated rough model, afterwards the segmentation is refined and finally the objects are cleared form the image. The segmentation steps are not exposed here, a comprehensive description can be found in [2].

The segmentation data is used to remove the bone shadows from the images in order to enhance the structure of the lungs. The elimination is based on creating intensity profiles on vertically differentiated images, which are subtracted from the differentiated original image. An integration step returns to the original domain and produces the bone shadow free image [3]. An example result can be seen on figure 1.

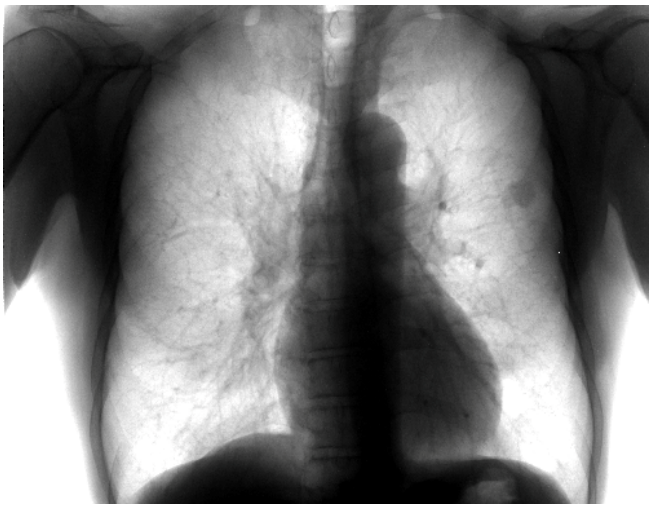

Fig. 1 The output of bone shadow removal

B. Enhancement of nodules

The aim of the next step in the scheme is the enhancement of nodules on chest radiographs. These objects are darker than the surroundings, mostly round shaped and have an approximate radius of 5-35mm. According to our experiences, shape information can provide a better clue for finding nodules, because of usual low contrast. A commonly used filter family called Convergence Index (CI) approximates object borders and enhances them if their shape is approximately rounded. A common property of round shaped objects is the radial direction of gradient vectors along their border. The filters consider the surroundings of each pixel. The output depends on the angles of the vectors connecting the center and the surrounding points and the gradient vector at the surrounding point. One of the most successful realizations is the Sliding Band Filter (SBF) using the following idea [4]. The most important parameters are illustrated on figure 2.

The algorithm considers each pixel of the image as a potential center of an object. For each center it slides a band in different directions within given bounds, while the band has a fixed width. For each band position the algorithm takes the points inside the band and sums the cosine of the angles of radial vectors (the vector connecting the center and the given point) and gradient vectors at the points. The final position of the band for a direction will be the one with the highest sum. Note that a high sum is caused by the convergence of negative gradient vectors towards the center and can sign an object border. For a round shaped dark object the negative gradients are convergent in every direction, if the starting point was the center of the object. Using this idea, the algorithm sums the maximal band values in each direction, and a high final sum indicates a nodule.

A weakness of the algorithm is the independence of the bands in each direction, enhancing very spiculated and distorted objects. An intuitive solution is to apply a constraint on the bands in different directions. Our proposed algorithm the Constrained Sliding Band Filter (CSBF) links the position of the bands allowing smaller distortion. It ensures that the final band positions satisfy a circularity constraint controlled by a coefficient. The coefficient forces an upper bound to the ratio of the distance of the farthest and the closest band from the center. The enhanced pixel values can be calculated with the following formula.

$$CSBF(x,y) = \max_{R_{\min} \le r \le \frac{R_{\max}-d}{c}} \frac{1}{N} \sum_{i=1}^{N} Cmax_{ir},$$

$$Cmax_{ir} = \max_{r \le n \le r*c} \frac{1}{d} \sum_{m=n}^{n+d} \cos\theta_{im},$$

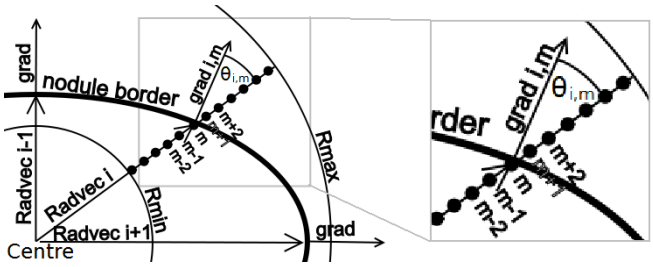

Fig. 2 Illustration of SBF algorithm

where R_{min}, R_{max} are the bounds of the target object radius, c is the shape constraint coefficient, N is the number of directions concerned, d is the width of the band and θ_{im} is the angle of the m^{th} gradient vector along the i^{th} radial direction and the corresponding radial vector. An output identical to CSBF can be achieved by running several SBF filters with different bounds (R_{min}, R_{max}) and taking the minimum for each center, however the execution times would be much greater. Thanks for the implementation the CSBF take only slightly longer than one SBF run. For very large c values the CSBF works as a standard SBF, thus it is a more general algorithm. Furthermore for c=1 the CSBF is identical to the Iris filter [4], another realization of the CI family.

Parameter selection for CSBF was made heuristically. R_{min} and R_{max} were set to match the smallest and larges nodules to be found. The parameter d affects the sensitivity to noise and was set to 5.6mm. N was set to 16 as a good compromise between precision and speed. The circularity parameter was optimal around 1.2.

C. False positive reduction

The last step of the CAD scheme concerns the areas with high CSBF value. A good practice is to collect many areas and select the suspicious ones with the help of a classifier. Our solution uses a Support Vector Machine (SVM) [5], due to its good generalization capability. The training sample set is extracted from a radiograph database with validated nodules. For the SVM to work efficiently the dimensionality of the input has to be reduced. This is done by calculating various features that describe the shape, texture and symmetry of the area to be classified. This way the raw image of the area is reduced to a vector of 140 dimensions. Afterwards it is further reduced to 12 dimensions by relevance based dimension reduction techniques.

The 12 features that turned to be the most useful are the following. The coordinates of the nodule serve as an important clue, due to the uneven distribution of tumors on the image. Concerning the texture, several statistical features based on the distribution of directed pixel pairs provided

useful results. These were the contrast, angular second moment and various entropy related measurements described in [6]. Other important features are related to the output of a Laplacian of Gaussian filter and a so-called Average Fraction Under the Minimum (AFUM) filter. The former is a well known filter for edge detection, and can also be used for nodule finding. The value of the pixel at the center, the average and the entropy of filter output was used as a classifier input. The AFUM algorithm is another filter for finding round shaped objects described in [7]. The filter output at the center of the area proved to help classification.

For the SVM itself, we use a general purpose radial kernel function. It showed good performance on low dimensional input compared to polynomial kernels. Parameter selection was made by a simple algorithm, which searches the parameter space on a logarithmic scale and iteratively refines the search near the found optimal solution.

III. Results

First we compared the new CSBF with an existing SBF solution, and then we analyzed the effect of rib removal and finally measured the overall performance.

For testing we used two separate databases. A set of 247 images widely used for benchmarking created by the Japanese Society of Radiological Technology (JSRT) [8] and a private database of 150 images originating from a Hungarian clinic. They contain 154 and 100 nodules respectively of various subtleties. While the JSRT radiographs come from an analogue X-ray machine and digitized by a scanner, the ones in the private database are directly made by a digital detector.

For the SBF versus CSBF comparison we disabled both the rib removal and the classifying phase to get more accurate results of nodule enhancing capability. Nodules are selected from the enhanced image by a simple adaptive thresholding. The results on the JSRT database can be seen from table 1. The new algorithm succeeded in five more cases at finding the real nodule than original SBF, meaning a 3% increase in sensitivity.

We ran the complete three step algorithm on the JSRT database to test the effect of rib removal. This way we're able to get a clear view how this preprocessing step can help the latter algorithms.

Table 1 Nodule detection performance of SBF and CSBF

Method	Sensitivity	Avg. no. of false positives
SBF	71.4	19.5
CSBF	74.7	19.5

As the removal algorithm failed to detect ribs on a few images, but a future improvement will likely to fix the issue, we used only the images where rib removal succeeded. As a test method we've chosen 4-fold cross-validation.

After optimizing the parameters and using the original images, 57% of the real nodules were found while producing on average 4 false detections per image. With utilizing the rib removal algorithm the sensitivity increased to 61% for the same number of false positives, which means a clear improvement. The low absolute performance is caused by the incomplete database.

To approximate overall performance, we ran experiments on both databases. Because of the mentioned problem we couldn't run the three step solution on the complete databases. To provide comparable results with the literature, we needed to process all the images, so we've chosen to omit the bone shadow removal algorithm when calculating overall performance.

For the JSRT database we plotted the results on a Free-Response Receiver Operating Curve (FROC). The results can be seen on figure 3. Here the sensitivity is shown as the function of the average number of false positives per image. An appropriate working point can be the 61% accuracy with 2.5 false positives on average.

The performance on the private database turned to be somewhat worse. For example the sensitivity was 60% at a false positive rate of 4.

The execution time of the original SBF based system without preprocessing was 10 seconds on a 2.6GHz Intel Pentium Dual Core processor. With the CSBF implementation execution time remained 10 seconds while with rib removal increased to 30 seconds.

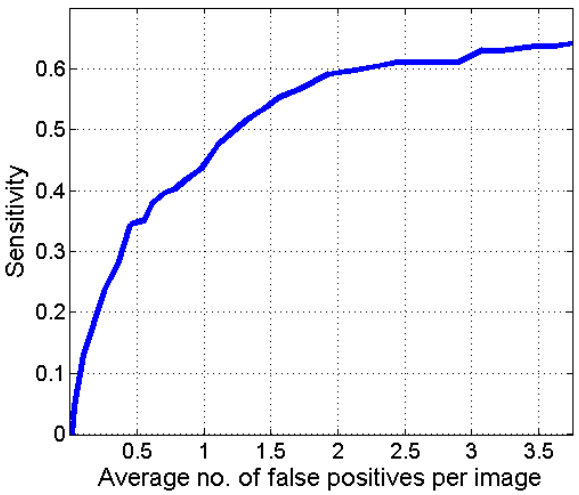

Fig. 3 Overall performance on the JSRT database

IV. Discussion

The main reason behind the improvement, when using CSBF is the insensitivity of the algorithm to distorted or spiculated but nodule sized objects. When these objects for example the parts of the bronchia are less enhanced, the used adaptive threshold gets lower, causing more nodules to become over the threshold. We've found the optimum of the circularity parameter around 1.2. It means that objects whose largest radius is more than 1.2 times larger than the smallest radius are usually not nodules. If we set the constraint to a lower thus stricter value, we start to lose a considerable amount of real nodules.

The bone shadow removal algorithm has a complex effect on nodule finding performance. When the nodule is not overlapped by a bone shadow, it does not change anything on the original image. However, the output of the CSBF often increases if the nodule is close to an eliminated shadow. This is because a neighboring edge modifies gradient vectors of the nodule border usually to the opposite direction, reducing the CSBF value. In this case, bone shadow removal helps nodule finding. If the nodule is overlapped by a bone shadow, after removal it becomes slightly dimmer on the radiograph. The CSBF algorithm enhances somewhat the bony structures, which causes lower output for the nodule on bone shadow free images, so this effect reduces the performance. In most cases lung nodules aren't overlapped by bone shadows, as they cover less than 50% of lung area, thus the first effect is more dominant and recognition efficiency increases. However, it would be still important not to suppress nodules overlapped by bone shadows but for this problem further improvements are needed.

Concerning the overall performance, it shows comparable results to the best solution in my scope using the JSRT database [1]. 2.5 false positives mean acceptable extra work for the examiner, but the over 60% sensitivity ensures that the CAD system will have some results that the examiner would've missed otherwise.

On the other hand the results on the private database showed that we can sometimes face worse results in real-world scenarios, and the JSRT may not be fully representative. The main reason behind the worse overall results on the private database is the greater variety of lungs in it and most of the patients having other disorders affecting their lung. This is because the included radiographs come from a lung clinic and not a screening station, so most of the cases are not like the average healthy lung with a lung nodule. The disorders can cause contrast changes in some parts of the lung creating false detections and decreasing detection performance.

The execution time without preprocessing is appropriate for on-line usage, because the time while the radiologist examines the radiograph without markings is enough for the algorithm to run in the background. Unfortunately the approximately half minute run time when using rib removal can be frustrating for some radiologists, however it still seems to be useful for most of the examiners.

V. Conclusions

In conclusion the proposed algorithms can improve lung nodule detection accuracy. Both the CSBF and the bone shadow removal caused a small but obvious increase in performance making them useful for latter nodule detection systems. The achieved performance should enable our system to help radiologists; however the current number of false positives leaves the need for further improvements. Furthermore execution times have to be shortened for convenient usage. To gain experiences of live operation and usefulness, the system is built in the software of an X-ray machine and used experimentally at a clinic.

VI. Acknowledgment

This work was partly supported by the National Development Agency under contract KMOP-1.1.1-07/1-2008-0035.

References

1. R. C. Hardie, S. K. Rogers, T. Wilson, A. Rogers (2008) Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med. Image Anal. 12/3:240–258
2. S. Juhász, Á. Horváth, L. Nikházy, G. Horváth, Á. Horváth (2010) Segmentation of anatomical structures on chest radiographs. Submitted.
3. G. Simkó, G. Orbán, P. Máday, G. Horváth (2008) Elimination of clavicle shadows to help automatic lung nodule detection on chest radiographs. IFMBE Proc. Vol. 22, 4th Eur. Conference of the International Federation for Medical and Biological Eng., Antwerp, Belgium, 2008, pp 488–491
4. Carlos S. Pereira et al. (2007) Evaluation of Contrast Enhancement Filters for Lung Nodule Detection. ICIAR Proc. vol. 1, International Conference on Image Analysis and Recognition, Montreal, Canada, 2007, pp 878–888
5. M. Altrichter, G. Horváth, B. Pataki, Gy. Strausz, G. Takács, J. Valyon (2006) Neurális Hálózatok. Panem, Budapest
6. R. Haralick, K. Shanmugam, I. Dinstein (1973) Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-3, no. 6: 610–621
7. M. D. Heath, K. W. Bowyer (2000) Mass Detection by Relative Image Intensity. IWDM Proc., International Workshop on Digital Mammography, Toronto, Canada, 2000, pp 219–225
8. J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, K. Doi (2000) Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule. AJR Am J Roentgenol 174: 71–74