IFMBE Proceedings Series– A Hybrid Lung Nodule Detection Scheme on Chest X-ray Images

G. Orbán[1] and G. Horváth[1]

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

Abstract— **In the current study, we propose a computer aided detection (CADe) scheme targeting lung nodules on chest radiographs. Instead of using the common scheme of a nodule enhancement filter followed by a classifier, our novel approach utilizes separate filters targeting different types of lung nodules. Smaller and low-contrast nodules are enhanced based on the convergence of gradient vectors, while larger objects with higher contrast are to be found by a modified top-hat filter on a local contrast enhanced image. After segmentation of nodule candidates and calculation of features on them, a classifier network consisting of three Support Vector Machines (SVM) reduces the number of false positive findings and merges the results of the two enhancer filters. The CADe system is tested on a radiograph database containing images of 93 patients with validated lung nodules and 150 healthy cases. The results of a system using only one nodule enhancer filter and the hybrid system are compared using a Free-response Receiver Operating Characteristic (FROC) analysis. The hybrid solution turned out to be clearly superior to the other schemes. Considering the whole system, we experienced 72% sensitivity at a false-positive rate of 2 and 77% sensitivity at a false positive rate of 3.**

Keywords— **Nodule detection, chest radiograph, CAD, lung cancer.**

I. INTRODUCTION

Lung cancer is one of the most common causes of cancer death. Many cures are only effective in the early and symptomless stage of the disease. Screening can help early diagnosis, but a sensitive, cheap and side effect-free method has to be used to enable mass usage. Sensitivity means the fraction of correctly identified positive cases and all positive cases in this context. Standard chest radiography meets these requirements, except that current methods have moderate sensitivity. Efficiency can be improved by using a Computer Aided Detection (CADe) system. The most important problem of existing CADe systems is the low positive predictive value. In other words, high sensitivity can only be reached at the cost of many false detections. Recently published systems can detect 60-70% of cancerous tumors, while they also mark approximately four false positive regions on each image [1], which allows them to be used only as a second reader.

Usability of CADe systems can be improved either by reducing the number of false detections – to give the examiner less extra work –, or by finding more nodules – to increase sensitivity. The detections of CADe should be also complementary to the findings of radiologists to better improve sensitivity when radiologists and CADe work in cooperation. Although this is true from the performance point of view, we observed that radiologists lose their faith in the system and ignore its results if it fails to detect some obvious cases. This remains true even if we explicitly specify what type of nodules the system searches for.

To overcome this issue, the current study focuses on enhancing the capabilities of an existing CADe scheme to find special types of nodules that were missed previously, but were frequently found by the radiologists. Therefore we introduce the Large Nodule Filter (LNF) targeting nodules larger than 30 mm diameter, having high contrast and usually overlapped by large structures for example the shadow of the heart or the spine. Finding these nodules have little utility in everyday screening, but may help the acceptance of CADe amongst radiologists. As a side effect, we developed a framework that enables the simultaneous use of many nodule enhancing algorithms and the efficient integration of their results. This may help us in the near future to integrate more algorithms that complement our current algorithms by finding more subtle cases.

II. MATERIALS AND METHODS

For our solution we used the following three-step scheme. The first step, described in [2], segments the viewable area of the lung and frees the image from unnecessary objects and noise, making the nodule more visible. The next two steps can be seen in Figure 1. Nodule enhancement highlights round shaped objects like the target lung nodules by using image processing algorithms. After normalization and resizing, the processing splits up based on the targeted nodule type. The Constrained Sliding Band Filter (CSBF) – based on the Sliding Band Filter (SBF) described in [3] – is used for the enhancement of smaller and subtle nodules, while the recently developed LNF for large nodules with high contrast. The collection of nodule candidates utilizes the same algorithm for the two threads.

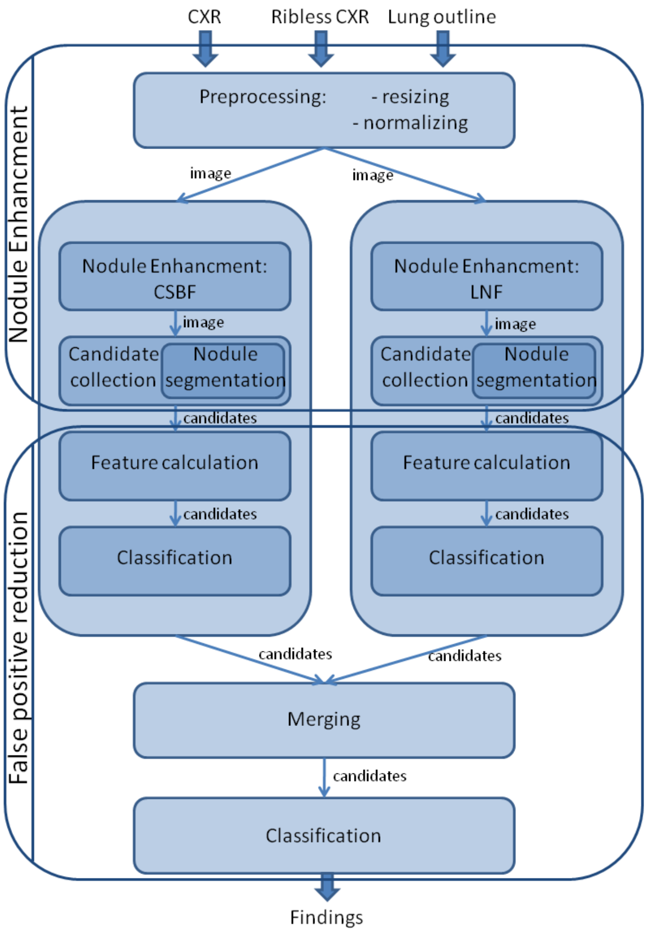

Fig. 1 The nodule detection process. Arrow labels reveal the type of data flow.

The last step reduces the number of false positive findings on the enhanced image with the help of a classifier. It begins with the calculation of features of candidates serving as an input for the classifier. Classification is done for each thread followed by a merging step that eliminates duplicate and highly overlapping results. A final classification is carried out on the merged set of candidates.

A. Nodule enhancement

According to our observations, different types of nodules may need completely different algorithms for efficient enhancement, as they have different characteristic properties. A smaller nodule in the early stage is usually dimmer, but has an approximately circular shape, thus the convergence of the gradient vectors can serve as an important clue. In large scales this can be computationally intensive. Furthermore, larger structures and intensity variations can alter gradient directions. On the other hand, large nodules tend to have high contrast that can help detection. Hereinafter we will introduce the new LNF algorithm specialized for large nodules, as the CSBF filter targeting smaller ones has been described in detail previously [4].

The LNF aims to enhance nodules with diameter between 30 mm and 75 mm and high contrast, but allows them to lie almost completely outside the viewable lung area. The basic idea behind the algorithm is a modified Local Contrast Enhancement (LCE) followed by a top-hat filter. The LCE output – $G(x,y)$ – is

$$G(x,y) = \frac{1}{1 - \exp(-D(x,y))},$$

$$D(x,y) = F(x,y) - \frac{1}{|R(x,y)|} \sum_{(u,v)\in R(x,y)} F(u,v),$$

$$R(x,y) = \begin{cases} C(x,y) \cap L & (x,y) \in L \\ C(x,y) \cap \overline{L} & otherwise \end{cases},$$

$$C(x,y) = \{(u,v) \mid (u-x)^2 + (u-y)^2 < 2r^2\},$$

where F is the original image, L is the viewable lung area, and r is the radius of the targeted nodule. The trimming of R with L ensures that we have a homogeneous area completely inside or outside the lung. The rationale behind the logistic function is to get a result in between local normalization and local thresholding. An example output can be seen in Figure 2 (top right).

Top-hat filtering is a simple convolution by a cylinder shaped kernel with radius r. The side of the cylinder is normally vertical, which maximizes filter output for perfectly circular shapes However, if the target shape is somewhat distorted, the filter output heavily decreases. To keep a high filter output for slightly distorted objects, the side of the cylinder is tilted. This method would also enhance – besides nodules – other dark structures, like remainders of rib shadows or areas filled with vessels. According to our observations, nodules tend to be more homogeneous than these misleading areas, so to suppress them, the filter output is weighted with the smoothness of the area. For smoothness, the standard deviation of smoothed nodule pixels inside the viewable lung is calculated. An example result is shown in Figure 2 (bottom left). The method uses a multi-scale framework to detect different sized nodules. At four different scales, r is set to 15 mm, 20 mm, 27 mm, and 38 mm.

To find nodules lying outside the viewable lung area, the top-hat filter is run for the entire image. This would also enhance structures like the vertebra, so the filter output is kept only where both negative and positive parts of the cylinder overlap with the viewable lung, and the area of intersection for both is greater than 15% of the filter part area. As a post processing step the areas where the predicted nodule would lie outside the whole lung are suppressed. This requires an overestimation of the whole lung area, for which we use the following algorithm. The binary masks of the left and right viewable lung parts are dilated towards the centre and then eroded with the estimated nodule radius. Then, the union is taken for the two parts. The resulting mask is consistent with our assumption that nodules can reside under the shadow of the heart, aortic arth or hemi-diaphragm, but cannot hang out towards the side of the body. An example is shown in Figure 3.

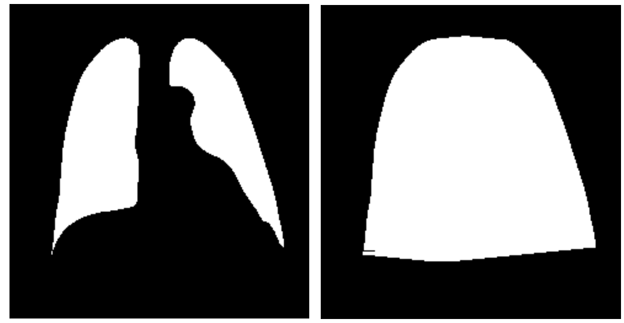

Fig. 3 The masks of the viewable lung area (left) and the estimated area of the whole lung (right)

B. Reduction of false positives

The candidate collector finds approximately 20-30 candidates on the enhanced image. For the reduction of false positive findings, we use a Support Vector Machine (SVM) classifier. The training data comes from validated findings of pulmonologists. The capabilities of an SVM largely depend on the choice of the kernel function, for which we use the widespread isotropic Gaussian kernel. The input vector of the kernel consists of various features describing texture, geometry and location.

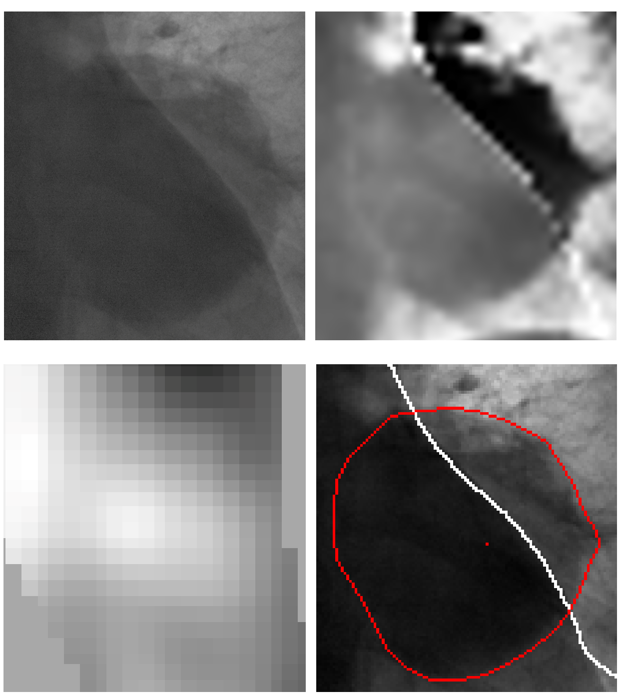

Fig. 2 A large nodule partly overlapped by the heart (top left), the LCE output (top right), the final LNF output (bottom left) and the segmented nodule candidate (bottom right).

To eliminate irrelevant features, we ran a simple feature selection algorithm which observes the performance of the SVM while adding various features in a forward selection manner. As the relevant features turned out to be different for the output of the two nodule enhancing algorithms, we decided to use separate classifier for each result set. Using one classifier requires the union of relevant features which would increase the number of dimensions, thus reducing overall performance. Furthermore, tests showed that multiple classifiers can save run time.

The SVM with isotropic Gaussian kernel requires two hyperparameters to be chosen by the user. For this, we use an iteratively refining grid search. Using many, well placed grid points in a single iteration makes the algorithm robust against local maxima, while iterative refinement cuts computational complexity. For performance measurement, cross validation is utilized.

III. Results

We tested the system on a private chest X-ray database containing images of 243 patients where 93 of the cases contained at least one malignant lung nodule. Nodule diameter ranged from 2 mm to 98 mm, the average was 24 mm. The malignant cases were validated by various clinical tests, most of them by CT. The images came from a digital X-ray machine working in daily practice at a Hungarian clinic. We labeled a CADe marker as true positive if its center of gravity fell inside a physician's marker.

As a testing method, we used 4-fold cross validation with 100 iterations for each setup, to reduce the variance introduced by the random permutation of images. The results can be seen on an FROC curve in Figure 4. The plot shows the fraction of malignant cases that can be found as a function of average number of false positives produced for each image. We ran the FROC analysis for a system using only the CSBF or the LNF enhancers and for the complete version using both algorithms.

We expected the poor performance of the LNF on its own. Despite finding almost all large lesions, we measured a very low sensitivity, as the radiographs contained mostly small nodules; however, integrating its results with the CSBF improves sensitivity without adding many false positives. The increase compared to the standard CSBF solution is obvious. At constant 70% sensitivity the number of false positives can be reduced from 2.3 to 1.7 per image. Alternatively, with a false positive rate of 3 the sensitivity can be increased from 73% to 77%.

The final results are good enough if we consider everyday applicability. 72% sensitivity with 2 false positives or 77% with 3 requires acceptable extra work from the examiner while it marks malignant areas, the examiner may overlook otherwise. Of course, the number of false positives should be reduced to further improve the usability of the system. However, it is not yet examined whether the used image database can be considered representative, thus comparison of the results with other systems should be made carefully. Although we cannot be sure that system performance remains the same in everyday practice, we are optimistic as an in progress clinical study reported similar results based on the first 800 cases.

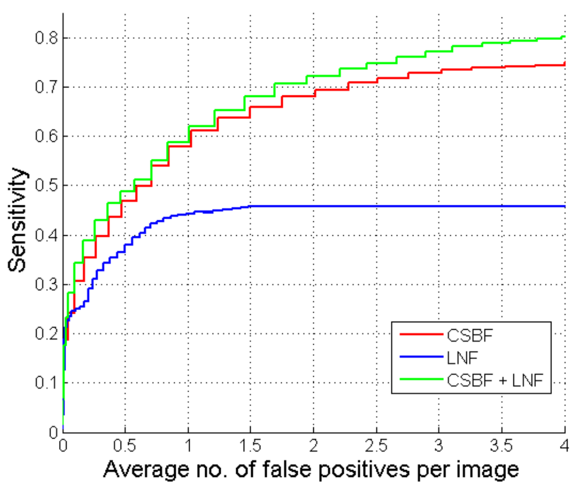

Fig. 4 Comparison of systems using only CSBF or LNF and the complete (CSBF + LNF) version.

IV. Conclusions

In the current work a new filter able to find larger nodules was developed and integrated into our existing CADe scheme. The main contribution to the CADe community is twofold. First, the introduced LNF algorithm can be a useful tool if the automated detection of mature tumors is needed. Second, the current case shows that a hybrid system involving specialized filters and proper synthesis can be more efficient than a system using one general-purpose filter. Furthermore, the resulting CADe scheme turned out to be efficient for everyday clinical use. Future improvements should focus on the further reduction of false positive rate to ensure that CADe increases only the number of true positive diagnoses and not the number of unnecessary examinations.

Acknowledgment

The authors express their thanks to all physicians of Pulmonological Clinic, Semmelweis University, who helped us with validated images and to the development staff of ImSoft ltd. 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. Medical Image Analysis, 12(3):240–258
2. S. Juhász, Á . Horváth, L. Nikházy, G. Horváth, Á . Horváth (2010) Segmentation of Anatomical Structures on Chest Radiographs, IFMBE Proc. vol. 29, XII Mediterranean Conference on Medical and Biological Engineering and Computing, Chalkidiki, Greece, 2010, pp 359–362
3. Carlos Pereira, Ana Mendonça, Aurélio Campilho (2007) Evaluation of Contrast Enhancement Filters for Lung Nodule Detection. Image Analysis and Recognition, 4633: 878-888
4. G. Orbán, Á . Horváth, and G. Horváth (2010) Lung Nodule Detection on Rib Eliminated Radiographs, IFMBE Proc. vol. 29, XII Mediterranean Conference on Medical and Biological Engineering and Computing, Chalkidiki, Greece, 2010, pp 363–366

Author: Gergely Orbán
Institute: Budapest University of Technology and Economics
Street: Magyar tudósok körútja 2.
City: Budapest
Country: Hungary
Email: orbang2001@yahoo.com