Nodule Detection in Postero Anterior Chest Radiographs

Paola Campadelli[1] and Elena Casiraghi[1]

Universitá degli Studi di Milano, Dipartimento di Scienze dell'Informazione
http://www.dsi.unimi.it, {Campadelli,Casiraghi}@dsi.unimi.it

Abstract. The use of image processing techniques and Computer Aided Diagnosis (CAD) systems has proved to be effective for the improvement of radiologists' diagnosis, especially in the case of lung nodules detection. In this paper we describe a method for processing Postero Anterior chest radiographs which extracts a set of nodule candidate regions characterized by low cardinality and high sensitivity ratio.

1   Introduction

Chest radiography is by far the common type of procedure for the initial detection and diagnosis of lung cancer. This motivates the great deal of research aimed to the creation of CAD systems aimed to lung nodules detection in chest radiographs [1]; however the problem is still open due to the significant loss of true positives. The multiscale method described in this paper has been tested on a standard database acquired by the Japanese Society of Radiological Technology [5] and extracts a set of candidate regions with an high sensitivity ratio.

2   Method

We work on images down-sampled to 256×256 pixels and use the lung area obtained in [1] but extended in order to include the parts behind the heart, near the spinal column and behind the diaphragm. To enhance the conspicuity of nodules of different size and brightness we use a multiscale approach. We start creating 11 *Difference Images* by subtracting from the original one the result of its convolution with a gaussian filter with standard deviation $s (s \in [2, 12])$. The result of subtracting to a nodule sub-image its smoothed version is an image with a positive peak in the central part of the nodule, and negative values in the neighborhood. The histogram of each *Difference Image* always shows a peak on the set of the positive values; we then create 11 binary images by selecting for each *Difference Image* all the pixels with a value bigger than the one corresponding to this peak; they are the pixels corresponding to the details identified at the scale s. The sum of all the binary images is a *Sum Image*, where the nodules appear as circular regions of different sizes, characterized by the highest values and surrounded by a much darker ring. To detect these regions we repeat

the procedure described below for each possible radius value $R(R \in [2, 12])$. Having fixed the radius R, we calculate for each pixel $P = P(x, y)$ a coefficient $P_R = MEAN(Circle_R(P)) - MEAN(Ring_R(P))$, where $Circle_R(P)$ is the region composed by the pixels in the circle of radius R and centered in P, $Ring_R(P)$ is the region composed by the pixels in the 2-pixel-thick ring around the circle $Circle_R(P)$, $MEAN(X)$ is the mean of the gray values of the pixels in a region X. The thickness of the ring is fixed for every radius since what allows to identify a circular region is a darker ring surrounding it, no matter which is the thickness of the ring itself. To select the pixels which are potential nodule centers, we automatically find a threshold on the set of coefficients P_R with the algorithm described in [2]. For each connected region in the obtained *Binary Image*, we calculate its circularity, the biggest diagonal D of the minimum ellipse containing the region itself, and discard it either if the circularity is lower than 0.5 or D is bigger than $2R$. Repeating the procedure for all the 11 radius values we obtain a set of 11 *Binary Images* which are combined to extract a final set of candidates. All the regions appearing in only one of the *Binary Images* are taken as candidates. If some regions in different *Binary Images* overlap we choose as representative the one with the most circular shape.

The result is a set of about 24000 regions on all the 247 images of the database and only 7 true positives lost out of 153. We compared these results with those of other two extraction schemes tested on the same database. The first method, [4], extracts a set with 33000 candidates loosing 20 true positives; furthermore the authors apply a classification method that selects 5028 candidates, loosing other 15 true positives, for a total of 35 false negatives. We implemented the second method, [3] obtaining poor results in terms of true positive lost.

To prune the set of extracted candidates we calculated a set of 21 most representative features, based on the shape, the position, the gray level distribution in the original image, and the values of the coefficients P_R calculated during the candidates extraction. Applying to these features simple rules describing the relationships observed between pairs of features, we can reach a total number of candidates equal to 11500 loosing no true positive (sensitivity ratio= 0.95).

This high ratio demonstrates the efficiency of the method; further works will focus on the trimming of the false positive set to obtain a clinically useful system.

References

1. B. van Ginneken, A.: Computer-Aided Diagnosis in Chest Radiographs. P.h.D. dissertation, Utrecht Univ., Utrecht, The Nederlands. (2001)
2. Kapur,Sahoo,Woong, A.: A new method for gray level picture thresholding using the entropy of the histogram. Comp.Vis. Grap.and Im.Proc. **29** (1985) 273–285
3. Keserci,Bilgin and Yoshida, Hiroyuki, A.:Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model. Med. Image Analisys. **6** (2002) 431–447
4. A.Schilham, B.Van Ginneken, M.Loog, A.:Multi-scale nodule detection in chest radiographs. Proc. MICCAI. (2003)
5. Web address: http://www.macnet.or.jp/jsrt2/cdrom_nodules.html