Computer Assisted Detection systems for lung nodule identification in chest Computed Tomography images

R. Cataldo[1 2], I. De Mitri[3 2], G. De Nunzio[1 2], L. Martina[3 2], A. Massafra[3 2], M. Peccarisi[3 2], M. Quarta[4 2], M. Torsello[5], I. Zecca[5], and the MAGIC-5 Collaboration [6]

[1]Dipartimento di Scienza dei Materiali, Università del Salento, Lecce, Italy

[2]Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, Italy

[3]Dipartimento di Fisica, Università del Salento, Lecce, Italy

[4]Dipartimento di Matematica, Università del Salento, Lecce, Italy

[5]Unità Operativa di Radiodiologia, Presidio Ospedaliero "V. Fazzi", Lecce, Italy

[6]Universities and INFN units of Bari, Genova, Lecce, Napoli, Pisa, Torino

Introduction. Screening programs based on low-dose Computed Tomography (CT) have been shown to be a useful tool for the early detection of lung nodules and the reduction of the number of lung cancer deaths. A key instrument for this purpose is the use of Computer Aided Detection (CAD) systems for automatic identification of pathological structures in CT images. CAD systems are able to support the radiologist's diagnosis by improving his performance and saving time. The basic steps of a CAD system can be summarized as follows: (1) lung parenchyma segmentation, (2) localization of the Regions Of Interest (ROIs), (3) statistical and morphological feature extraction from the ROIs and (4) classification of nodule candidates. The setting up of CAD algorithms for an automated lung nodule identification is one of the main scientifc goal of the MAGIC-5 [1] collaboration, involving several italian Universities, INFN units and hospitals. The CAD system so far developed by the collaboration already gave satisfactory results[1]. Anyway, among several types of nodules those in contact with pleura, called "juxta-pleural" nodules, are often difficult to be detected due to their location and high density. Hereafter we will focus on the description of the tools that we are currently building for their detection.

Lung segmentation algorithm. The segmentation algorithm consists of several steps [2]. First, we find an appropriate gray-value threshold θ_0 of the respiratory apparatus by analyzing the image Hounsfield values. Using the θ_0 value we apply to the CT volume a simple-threshold 3D Region Growing (RG) to obtain a binary mask of the respiratory system. Voxels are included in the grown region if their Hounsfield value is smaller than θ_0. The external airways are subsequently extracted and removed by a wavefront simulation model under appropriate stop conditions, in order to obtain a mask for lungs only. A RG algorithm is then applied to grow separately the left and the right lung mask after fixing the cases of lung "fusion" by identifying the fusion region and by inserting a separation surface into the binary mask. As a result of the whole process, we have a pair of binary masks for the two lungs that do not contain nodules or vessels. While the cavities due to internal nodules and dense vessels can be easily filled up in the binary mask, more difficult is the treatment of concavities due to juxta-pleural nodules.

Detection of juxta-pleural candidate nodules. In order to include juxta-pleural nodules in the segmented volume we have to close the concavities by using a concavity-patching method that returns a smoothed lung border. The difference between the original border and the "closed" one gives the set of concavities. Among various concavity-patching methods, we use the *morphological closing* and the *alpha-hull*. The morphological closing is the application of two consecutive morphological operators called dilation and erosion, both using the same structural element (SE) (i.e. a disk with varying radius r). Moving this SE onto the image, the morphological closing returns a lung with all the concavities being closed. The α-hull [3] is a convex hull generalization, able to detect concavities, whose shape depends on a curvature parameter α. Given a set S of points in the plane, and a positive number α, the α-hull of S is defined as the intersection of all closed discs of radius $1/\alpha$ that contain all the points of S. By definition, the α-hull when α=0 is the convex hull. The effect of calculating the α-hull of a closed and dense spatial distribu-

[1]Medical Applications on a Grid Infrastructure Connection, http://www.magic5.unisalento.it

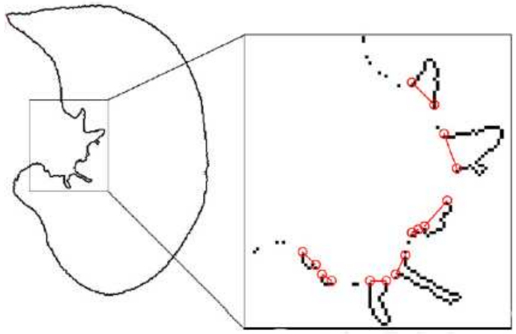

Figure 1. The automatically segmented lung border in a chest CT slice after concavity detection at an arbitrary α value (see text).

tion of points (such as the segmentation mask of a lung CT slice) is the gradual closing of concavities, depending on the value of α. In particular, if the border of the binary mask has nested concavities, the application of α-hull with increasing values of α allows the identification of concavity for different values of this parameter. This creates a natural hierarchy of the concavity that is ordered with respect to α (see Fig.1).

Feature extraction. Now that a list of concavities is available, we want to reduce the number of False Positives (FPs) by rejecting the concavities that are not related to juxta-pleural nodules (natural concavities). This task can be obtained by introducing a set of proper threshold values acting on several features. We can distinguish among two big classes of features: the geometrical and the texture ones. The most important geometrical features are: *span* that is the length of a segment that joins the concavity extremal points; *depth* that is the length of the longest perpendicular segment defining the span; *boundary length* that is the number of boundary points composing the concavity profile; *area* that is the number of pixels between the span segment and the concavity boundary. Other geometrical features are *depth over span*, *radius* and *circularity*. As for the features based on texture, the most obvious is the average "gray-value" *gray mean*. Other first-order features are related to the gray-value histogram, namely *grayStd*, *graySkew*, *grayKurt* and *grayEntropy*, referring to the standard deviation, skewness, kurtosis and entropy respectively.

Classification. A supervised two-layer, 13-input, 20-hidden-neurons, 1-output feed-forward neural network, trained with gradient descent learning rule with momentum, was chosen as the classifier system. To calculate classification efficiency for each of the concavity-patching methods, we define as true positives (TPs) the candidate nodules that meet the radiologist's diagnosis (see next paragraph) according to the following condition: the Euclidean distance between the centroid of the concavity, and the centroid of a diagnosed nodule, is lower than $1.5r_R$, where r_R is the nodule radius according to the radiologists. All other candidates are considered to be false positives. The 1.5 factor takes into account the radius measurement uncertainty. The Artificial Neural Network output, calculated on the concavity list, is distributed in the range [0,1]. By varying a decision threshold, and assigning target t=1 to candidates above threshold (probably positive), and t=0 to candidates below threshold (probably negative), sensitivity and specifcity referred to the known diagnosis can be calculated. The Receiver Operating Characteristic (ROC) curve is then obtained by plotting (for each threshold value) sensitivity versus [1-specificity], thus giving the overall algorithm figure of merit.

Lung CT datasets and the annotation protocol. Ground truths for clinical evaluation of imaging systems, to be used as a "gold standard" for CAD system, are of crucial importance. The Lung Image Database Consortium in America, the Cancer Action Project and the ANODE09 [2] initiative in Europe, are involved in the challenge of creating standard databases. MAGIC-5 is the first Italian project that foresees a close collaboration with expert radiologists, both for choosing and annotating lung cases, together with a statistical management of this data towards the formation of a CT lung data bank. Till now we collected about 300 scans, annotated by expert radiologists using the LUNA (LUng Nodule Annotation) automated tool, developed within the collaboration, and a common annotation protocol. It is then possible to test algorithms on a large database and results can be quantified.

REFERENCES

1. G.De Nunzio et al., *Approaches to juxta-pleural nodule detection in CT images within the MAGIC-5 Collaboration*, Nuclear Instruments and Methods in Physics Research A (2011),doi:10.1016/j.nima.2010.12.082
2. G. De Nunzio et al., *Automatic Lung Segmentation in CT Images with Accurate Handling of the Hilar Region*, Journal of Digital Imaging (2009) 10-20. doi:10.1007/s10278-009-9229-11 ISSN0897-1889 (print) 1618-727X (online).
3. W.F.Sensakovic et al., *A general method for the identification and repair of concavities in segmented medical images*, in: Proceedings of Medical Image Conference 2008, Dresden.

[2] http://anode09.isi.uu.nl/