SUPERVISED TEXTURE CLASSIFICATION FOR SEGMENTATION OF ABNORMAL LUNG IN CT

P. Korfiatis1, C. Kalogeropoulou2, A. Karahaliou1 and L. Costaridou1

Departments of 1Medical Physics and 2Radiology, School of Medicine, University of Patras, 265 00 Patras, Greece
E-mail: costarid@upatras.gr

Abstract: Delineation of lung fields in presence of diffuse lung diseases, such as interstitial pneumonias (IP), challenges conventional gray level based segmentation algorithms. To deal with IP patterns affecting lung borders, a texture classification scheme for lung segmentation is proposed. The proposed method is based on supervised texture classification to distinguish surrounding tissue (ST) from lung parenchyma (LP), both normal and IP affected, since IP patterns are primarily manifested as texture alterations. A support vector machine classifier was trained to distinguish the ST from LP classes based on second order statistics textural features extracted from a 9x9 sliding region of interest across a CT image. A case sample of 120 HRCT images depicting abnormal lung boundary, corresponding to 17 patients diagnosed with IP secondary to connective tissue diseases, was analyzed. The segmentation accuracy of the method was assessed by comparing automatically derived lung borders to manually traced ones, by an experienced radiologist, using area-overlap and border shape differentiation metrics. Average segmentation accuracy in terms of area-overlap was 0.940±0.028, while border shape differentiation with respect to mean, root mean square and maximum distance were 1.161±0.401, 1.520±0.832 and 5.200±3.752 mm, respectively. The method is envisioned as an initial step of a computer aided quantification scheme for IPs.

Introduction

High Resolution Computed Tomography (HRCT) is the modality of choice for visualization of Diffuse Parenchymal Lung Diseases (DPLDs) related patterns [1]. The interpretation of DPLDs on HRCT scans is characterized by high inter and intra-observer variability due to their complex and variable morphological appearance and the lack of standardized interpretation criteria, further complicated by the large amount of image data to be reviewed [2].

Computer-Aided Diagnosis (CAD) schemes that automatically detect and quantify radiologic patterns of DPLDs in HRCT images have been proposed to provide a second opinion to radiologists to improve follow-up management decisions [3-8]. These systems generally consist of two stages. The first stage is the segmentation of lung parenchyma (LP), left and right lung field region, while in the second stage characterization of DPLD areas is achieved by exploiting two-dimensional [3-7] (2D) and recently three-dimensional (3D) texture [8]. As highlighted by Armato and Sensakovic [9] the performance of lung CAD schemes is influenced by the accuracy of LP segmentation algorithms.

LP segmentation methods available in the literature are based on: low level image processing routines guided by a semantic network [10], 3D iterative thresholding [11] followed by anatomically guided smoothing of the mediastinum area [12], adaptive 3D region growing followed by 3D morphological closing [13] and automated 3D gray level thresholding applied on wavelet edge highlighted images followed by morphological closing [14], to correct under-segmentation in the mediastinum area.

Segmentation methods proposed either for normal or pathological lung are highly challenged by the presence of pathologies affecting lung borders such as DPLDs [15]. This is due to the insufficiency of gray level information in distinguishing diseased tissue from surrounding tissue (ST) corresponding to bone, fat and muscle.

The aim of this work is the development of a LP segmentation algorithm dealing with lung borders affected by Interstitial Pneumonia (IP) patterns, a subset of DPLDs. Since IP is primarily manifested as texture alterations of the lung parenchyma tissue, also affecting lung border, the proposed algorithm exploits supervised classification by means of a Support Vector Machine (SVM) and local texture analysis. The SVM classifier assigns a pixel into LP or ST based on its local texture properties (second order gray level statistics). The method is tested on a dataset of 17 HRCT cases spanning a range of intermediate (mild density) IP patterns such as ground glass, reticular and honeycomb. The accuracy of the method is assessed using quantitative metrics, by comparing automatically derived lung borders to manually traced ones.

Materials and methods

Dataset

Clinical cases were acquired from 120 HRCT images of 17 patients diagnosed with IP secondary to connective tissue diseases, radiologically manifested
with ground glass, reticular and honeycomb patterns. All HRCT images obtained with a Multislice (16x) CT (LightSpeed, GE), in the Department of Radiology at the University Hospital of Patras, Greece. All images were selected by an expert radiologist and care was taken to choose those with pathology affecting lung border. Acquisition parameters of tube voltage, tube current and slice thickness were 140 kVp, 300 mA and 0.625 mm, respectively. The image matrix size was 512x512 with average pixel size of 0.62 mm.

**SVM based segmentation of the abnormal lung**

An SVM classifier was employed to assign a label of LP or ST to each pixel based on its local texture properties. The classifier uses as inputs texture features extracted from a 9x9 pixel Region Of Interest (ROI) sliding along the original image. The classifier output represents/provides the class of the central pixel on the sliding ROI.

**Texture feature extraction**

In the current study second order gray level features, extracted from the 9x9 pixel ROI centered at the pixel to be classified, were employed to capture the inherent tissue texture differentiation between LP and ST.

The Gray Level Co-occurrence Matrix [16] (GLCM) is a well established tool for characterizing the spatial distribution of gray levels in an image (second order statistics). An element at location \((i,j)\) of the co-occurrence matrix signifies the joint probability density of the occurrence of gray levels \(i\) and \(j\) in a specified direction \(\theta\) and specified distance \(d\) from each other. Thus, for different \(\theta\) and \(d\) values, different matrices are generated. In this study, four GLCMs corresponding to four different directions \((0^\circ, 45^\circ, 90^\circ \text{ and } 135^\circ)\) and one distance \((d=1\ \text{pixel})\), were computed from the 9x9 pixel ROI. Thirteen features were derived from each GLCM: angular second moment, contrast, correlation, variance, inverse different moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measure of correlation 1 and information measure of correlation 2. The mean and range of each feature over the four GLCMs were calculated, comprising a total of 26 GLCM-based features.

**Support Vector Machine classifier**

Pixel classification was achieved by an SVM classifier [17]. The basic principles of SVM are the maximal margin of separation and the kernel trick. Considering a two-class pattern classification problem, the SVM first performs a non linear mapping \((\Phi)\) from a low-dimensional input space to a higher dimensional feature space via a kernel function \(K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)\). This mapping allows the SVM to achieve better class separation. An SVM can be trained to construct a hyperplane for which the margin of separation is maximized. Maximally separated margins parallel to the hyperplane divide the new feature space into class-specific sub-spaces based on labelled training patterns.

In this work, the radial basis kernel [18] function was utilized, given by:

\[
K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right)
\]

The parameter \(C\), which controls the trade off between margin maximization and error minimization, and \(\sigma\) were automatically derived according to radius-margin bound [19].

The training sample was consisted of 400 (9x9 pixels) ROIs originating from 26 HRCT slices, distinct from the 120 HRCT slices used for evaluating segmentation accuracy of the proposed method. 200 ROIs were extracted from the LP region, both normal and IP affected (ground glass, reticular and honeycombing), and 200 ROIs were extracted from ST (fat, muscle, and bone).

**Feature selection**

An exhaustive search procedure was employed to select a best feature set among the 26 extracted GLCMs features. Specifically, combinations of 2-6 features were investigated and the combination of minimum number of features that provided the highest classification performance was selected. For each feature combination fed into the SVM classifier, classification performance was evaluated by means of the area under Receiver Operating Characteristic (ROC) curve [20] and utilizing the ‘leave-one-out’ strategy. For each feature combination, SVM parameters \((C\ \text{and } \sigma)\) were automatically estimated according to the radius-margin bound [19].

**Segmentation accuracy assessment**

Segmentation accuracy of the proposed method was assessed by means of area overlap and shape differentiation metrics [11,14] by comparing automatically derived lung borders to manually delineated ones. A radiologist experienced in CT image interpretation, not participating in the training sample selection, generated manual outlines of lung borders (left and right lung fields). For manual delineation, a tablet (Wacom Intuos3 Tokyo, Japan) was used with an active area of 305x305 mm with 5.080 lpi and accuracy of ±0.25 mm. For delineation of lung borders in the mediastinum, bifurcation of the main bronchi was considered as the lung border.

**Results**

Among the 26 GLCMs features, a subset of 5 features was selected yielding an area under ROC curve \((A_c)\) of 0.958±0.026. The features selected were: Mean of Angular Second Moment, Mean of Sum Variance, Mean of Sum Entropy, Mean of Entropy and Mean of Difference Variance. The SVM parameters corresponding to the selected features set were \(C=0.891\) and \(\sigma=3.638\). Classification performance without
considering the feature selection step (26 features) was 0.897±0.036.

Figures 1 and 2 provide application examples of the proposed method on two HRCT images corresponding to ground glass pattern and combined reticular and ground glass pattern, respectively. Automatically derived borders are indicated by red line while manual segmentations provided by the Radiologist (ground truth) are indicated by blue line. The proposed method segmented accurately lung borders, including all the IP affected areas. Normal LP areas were further accurately segmented in both cases.

Figure 1: Application example of the proposed method on HRCT image corresponding to ground glass pattern (black arrows). Red and blue lines indicate automatically and manually derived lung borders, respectively.

Figure 2: Application example of the proposed method on HRCT image corresponding to combined pattern of both ground glass and reticular (black arrow). Red and blue lines indicate automatically and manually derived lung borders, respectively.

Segmentation accuracy of the method was assessed by calculating area overlap and three border differentiation metrics (mean distance, root mean square distance and maximum distance) between the automatically derived LP borders and the manually traced ones.

Average segmentation accuracy in terms of area-overlap was 0.940±0.028, while border shape differentiation with respect to mean, root mean square and maximum distance were 1.161±0.401, 1.520±0.832 and 5.200±3.752 mm, respectively.

Discussion

A lung segmentation algorithm in HRCT is proposed in this study, to deal with lung parenchyma affected by IP patterns. The algorithm performs SVM pixel classification based on local texture analysis.

Second order grey level statistics were used to capture texture pattern differentiations between lung parenchyma (normal and IP affected) and ST. A five-feature combination was selected demonstrating the highest ability (A_z=0.958±0.026) in discriminating LP (normal and IP affected) from ST.

Lung segmentation was achieved by an SVM classifier that assigned each pixel into LP or ST using as inputs the five texture features extracted from its local neighborhood (9x9 pixel ROI centered at the pixel to be classified).

Segmentation accuracy of the proposed method was quantitatively assessed against ground truth, as provided by one experienced radiologist. The proposed method demonstrated high segmentation accuracy which can be attributed to the high discrimination of LP from ST patterns provided by the selected feature set.

A relatively low performance of the proposed method with respect to maximum distance metric (5.200±3.752 mm), being the most sensitive one used in this study, probably due to under-segmentation in the mediastinum area in a few slices of the dataset analyzed. This effect may be attributed to the size of analyzing ROI that was not capable of capturing texture differentiations between large vessels and ST. ROI size selection by anatomical guidance could potentially deal with under-segmentation in the mediastinum area [12].

A direct comparison of the proposed method performance with the one recently proposed for the segmentation of the pathological lung [15] is not applicable, due to the heterogeneous data sets utilized. However, considering reported performance [15] for the one IP case analyzed, the current study seems to perform equally with respect to the Overlap metric.

Since IP patterns correspond to mild density abnormalities, the performance of the method should be investigated with respect to high density abnormalities such as consolidation. Future efforts should focus on investigating additional feature sets and on adapting the proposed segmentation algorithm on 3D HRCT datasets.

The proposed algorithm was implemented using MATLAB programming environment. The development and test platform was a PC equipped with an Intel Core 2 Duo processor with 2 GB of RAM.
memory. The processing time was approximately 10 minutes for each slice.

Conclusion

In this study, a lung segmentation algorithm dealing with interstitial pneumonia patterns in HRCT is proposed. A support vector machine classifier was designed to distinguish lung parenchyma from surrounding tissue classes based on texture features extracted from a 9x9 ROI sliding across a CT image. In the current study, second order grey level statistics features were investigated to capture the inherent tissue texture differentiation between lung parenchyma and surrounding tissue. The proposed method for segmentation of the abnormal lung has demonstrated high performance and could be used as a first stage of a computer aided diagnosis scheme for interstitial pneumonia patterns.

Acknowledgements

This work is supported by the Carathéodory Programme (C.180) of the University of Patras, Greece.

References