

Potential Usefulness of Curvature Based Description for Differential Diagnosis of Pulmonary Nodules

Y. Kawata¹, N. Niki¹, H. Ohmatsu², M.Kusumoto³, R.Kakinuma², K.Mori⁴,
K.Eguchi⁵, M. Kaneko³, and N. Moriyama³

¹ Department of Optical Science, Tokushima University, Japan

² National Cancer Center Hospital East, Japan

³ National Cancer Center Hospital, Japan

⁴ Tochigi Cancer Center, Japan

⁵ National Shikoku Cancer Center Hospital, Japan

Abstract. This paper examines the problem of obtaining a representation of the three-dimensional(3-D) pulmonary nodule images, which is a key problem in discriminating benign and malignant nodules for differential diagnosis of the lung cancer using thin-section CT images. A curvature based approach is developed with the aim of characterizing internal intensity structures of benign and malignant nodules. This approach makes use of curvature indexes to represent locally each voxel in a three-dimensional (3-D) pulmonary nodule image. From the distribution of curvature indexes and CT value over the 3-D pulmonary nodule image a set of histogram features is computed for global characterization of benign and malignant nodules. Linear discriminant analysis is used for classification and leave-one-out method is used to evaluate the classification accuracy. Compared with the performance of experienced physicians the potential usefulness of the curvature based features in the computer-aided differential diagnosis is demonstrated by using receiver operating characteristic (ROC) curves as the performance measure.

1 Introduction

Recently the detection rate of small peripheral pulmonary nodules has increased due to advances in imaging technology such as helical CT scanner [1]. For small nodules, the differential diagnosis by means of transbronchial or percutaneous biopsies can be difficult. There has been a considerable amount of interest in the use of thin-section CT images to observe small pulmonary nodules for differential diagnosis without invasive operation [1,2]. In assessing the malignant potential of small pulmonary nodules in thin-section CT images, it is important to examine the condition of nodule interface, the nodule internal intensity, and the relationships between nodules and surrounding structures such as vessels, bronchi, and spiculation [1,2]. Several techniques have been developed to quantify the pulmonary nodules. Nodule density analysis from CT value was

employed in early investigations[3,4]. The pattern classification approach incorporating multiple features, including measures of density, density distribution, and texture, was proposed to classify suspicious areas into malignant and benign lesions using slice images[5]. To quantify the relationships between nodules and surrounding structures the 3-D concentration index was derived [7] and classification approach between pulmonary artery and vein was proposed [6]. To analyze nodule surfaces, curvatures were introduced [8,9] and fractal features were utilized [10].

In medical image analysis, several approaches have been investigated to represent local intensity structure directly computing the geometrical characteristics from gray-level 3-D medical images[16,17,18]. We have been investigating to analyze internal intensity structure of pulmonary by using curvature indexes and CT values [11]. The curvature indexes consist of the shape index and the curvedness [14]. In this paper we present histogram features of curvature indexes and CT values to characterize the internal intensity structure for the sake of classifying benign and malignant pulmonary nodules. Compared with the performance of experienced physicians the potential usefulness of the curvature based features in the computer-aided differential diagnosis is demonstrated by using a receiver operating characteristic (ROC) curves [22] as the performance measure.

2 Nodule Segmentation

The 3-D chest images used in this paper are reconstructed from thin section CT images obtained by the helical CT scanner (Toshiba TCT900S Superhelix). The thin-section CT images are measured under the following conditions; beam width: 2mm, table speed: 2mm/sec, tube voltage: 120kV, tube current :200mA and 250mA. For the scan duration, patients held their breath at full inspiration. Per patient, about 60 slices through the pulmonary nodule center at 1mm intervals are obtained. The range of pixel size in each square slice of 512 pixels is between $0.3 \times 0.3 \text{ mm}^2$ and $0.4 \times 0.4 \text{ mm}^2$, and the slice contains an extended region of the lung area. The 3-D chest image is reconstructed from the thin section CT images by a linear interpolation technique to make each voxel isotropic. Our segmentation algorithm is described in previous publication[8]. Briefly, the segmentation of the 3-D nodule image consists of three steps; 1) extraction of lung area, 2) region of interest(ROI) selection including the nodule region, 3) nodule segmentation based on the geometric approach. This lung area extraction step plays an essential role when part of a nodule in the peripheral lung area touches the chest wall [12]. The ROI including the nodule is selected interactively. The 3-D deformable surface model proposed by Casselles [13] is utilized.

3 Curvature Based Representation

Each voxel in the region of interest(ROI) including the pulmonary nodule is locally represented by a vector description which relies on the CT value and two curvature indexes that decouples the shape attribute and the curvature

magnitude. By assuming that each voxel in the ROI lies on the surface which has the normal corresponding to the 3-D gradient at the voxel, we compute directly the curvatures on each voxel from the first and second derivatives of the gray level image of the ROI. To compute the partial derivatives of the ROI images, the ROI images are blurred by convolving with a 3-D Gaussian function of width σ . Herein, the width σ is represented as a scale. At each voxel the principal curvatures $\kappa_1(\mathbf{x}; \sigma)$ ($\kappa_1(\mathbf{x}; \sigma) \geq \kappa_2(\mathbf{x}; \sigma)$) are computed by using the approach proposed by Thirion and Gourdon [16].

The voxel \mathbf{x} in the ROI image is locally described by curvature indexes and CT value. The curvature indexes consists of the shape index and the curvedness [14,15]. The original shape index defined by Koenderink and van Doorn [14] gives a continuous distribution of surface types between -1 and 1. To introduce the shape spectral function of the object's surface patch, Dorai and Jain [15] have modified the original definition of the shape index so that the shape index maps the surface types on the interval between 0 and 1. The definition of shape index used here is based on the modified definition [15] and the shape index with scale σ at the voxel \mathbf{x} is given by

$$S_I(\mathbf{x}; \sigma) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{\kappa_1(\mathbf{x}; \sigma) + \kappa_2(\mathbf{x}; \sigma)}{\kappa_1(\mathbf{x}; \sigma) - \kappa_2(\mathbf{x}; \sigma)}. \quad (1)$$

The curvedness $R(\mathbf{x}; \sigma)$ [14] is given by

$$R(\mathbf{x}; \sigma) = \sqrt{\frac{\kappa_1(\mathbf{x}; \sigma)^2 + \kappa_2(\mathbf{x}; \sigma)^2}{2}}. \quad (2)$$

The curvedness quantifies how highly curved a surface is, and is inversely proportional to the size of the object.

4 Feature Extraction and Classification

In order to characterize globally the pulmonary nodule images through the local description, we use the shape spectrum which is originally proposed for object recognition in a range image by Dorai and Jain [15]. The shape spectrum measures the amount of the voxel which has a particular shape index value to characterize the 3-D pulmonary nodule image. The shape spectrum with scale σ is given by

$$H(h; \sigma) = \frac{1}{V} \iiint_O \delta(S_I(\mathbf{x}; \sigma) - h) dO \quad (3)$$

where V is the total volume of the specified region O , dO is a small region around \mathbf{x} , and δ is the Dirac delta function. In practice, a discrete definition of the shape spectrum is given by dividing the shape index range into B bins and counting the number of point falling in each bin k :

$$H\left(h = \frac{k}{B}; \sigma\right) = \frac{1}{N} \sum_{i=1}^N \chi_k(S_I(\mathbf{x}_i; \sigma) - h) \quad (4)$$

where \mathbf{x}_i is a point on the object's surface, N is the total number of points on object's surface, and χ_k is the characteristic function of the k th bin:

$$\chi_k(x) = \begin{cases} 1 & \frac{k-1}{B} \leq x < \frac{k}{B} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The objective in defining the shape spectrum is to measure the amount of the point which has a particular shape index value h . Thus, the shape spectrum is called the shape histogram [15]. For computational purposes, such as comparing spectra of different nodules, the shape histogram is normalized with respect to the volume of nodule. The normalized number of voxel falling in each bin represents the value of the shape histogram feature. The similar equations for the curvedness and CT value are obtained in the same manner. The domains of curvedness and CT value are specified $[0, C_{max}]$, $[I_{min}, I_{max}]$. A voxel in which the curvedness value is larger than C_{max} is considered as a voxel with curvedness value C_{max} . For the CT value the similar process is performed.

The linear discriminant analysis [19] is used to classify benign and malignant nodules based on the extracted features. A forward stepwise feature selection procedure [21] with the minimization of Wilks'lambda is used as an optimization criterion to select effective predictor variables. A leave-one-out method [20] is adapted to evaluate classification accuracy. In this method, one nodule image is left out from the classifier design group and a linear discriminant function is formed using the design group. The discriminant score is computed for the left-out case by using the resulting linear discriminant function. This process cycles through the data set until every nodule image is used. The discriminant scores are analyzed using receiver operating characteristic (ROC) method [22]. The discriminant scores of the malignant and benign nodules are used as the decision variable in the LABROC1 program which fit the ROC curve based on maximum likelihood estimation.

5 Results and Discussion

In this section, we present classification results with a set of histogram features derived from the curvature based representation of pulmonary nodules. The thin-section CT images of peripheral pulmonary nodules used in this study were provided by the National Cancer Center Hospital East. The data set included 128 nodule images from 128 patients. Of the 128 cases, 95 contained malignant nodules and 33 contained benign nodules. Lesions that showed no change or decreased in size over a 2-year period were considered benign. Other lesions were cytologically or histologically diagnosed. The performance of a set of

Table I. Classification results of histogram features based on CT values.

Feature space	HCT1	HCT2	HCT3
Num. of bins	25	50	100
Num. of features	2	3	3
Error rate (%)	36.7	35.2	28.9

Table II. Classification results of histogram features based on curvedness.

Feature space	HCV1	HCV2	HCV3
Num. of bins	25	50	100
Num. of features	1	1	3
Error rate (%)	38.3	36.7	26.2

Table III. Classification results of histogram features based on shape index.

Feature space	HSH1	HSH2	HSH3
Num. of bins	25	50	100
Num. of features	1	1	2
Error rate (%)	22.7	22.7	20.3

Table IV. Classification results for selected combination of feature spaces

Combined feature spaces	Num. of features	Error rate (%)
HCT3 HCV3 HSH3	12	12.5
HCT3 HSH3	7	13.3
HCV3 HSH3	4	15.6
HCT3 HCV3	6	26.6

histogram features was evaluated using our dataset. Table I, II, and III present the classification results for the histogram features based on CT value, curvedness, and shape index, respectively. In each table the histogram feature space name, the number of bins, error rate, and the number of features selected are presented. It is necessary to select appropriate scale for the effective description of nodules internal structure [23]. We selected a scale which provided the high classification rate over the specified discrete scales and assigned the value 2 to the scale in this study.

The benign/malignant classification results for the selected combinations of feature spaces are given in Tables IV. The results for only some of the interesting combinations tried were reported. The classification result of shape histogram features HSH3 achieved better accuracy than that of CT value and curvedness histogram features. For the selected combinations of feature spaces the combined feature spaces HCT3, HCV3, and HSH3 gave the best performance. The selected features provide an information which bins contribute to discriminate malignant cases from benign cases.

In order to comparison between physicians and computerized classification, the probability of malignancy of each pulmonary nodule in thin-section CT images which were printed on films was ranked by 11 physicians on a scale of 1 to 10, where a ranking of 1 corresponded to the nodules with the most benign cases.

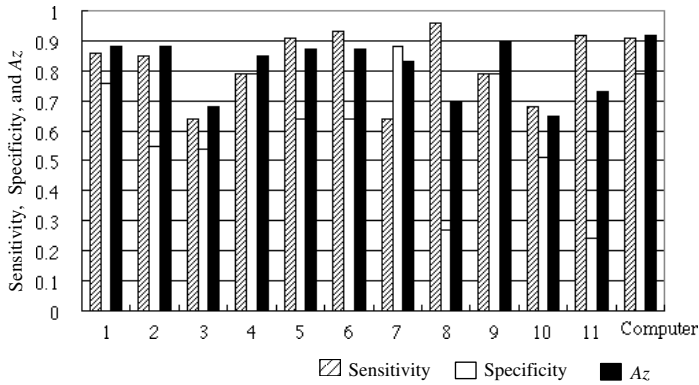


Fig. 1. Sensitivity, specificity, and A_z value of each physician and the computerized classification with the combined feature space HCT3, HCV3, and HSH3.

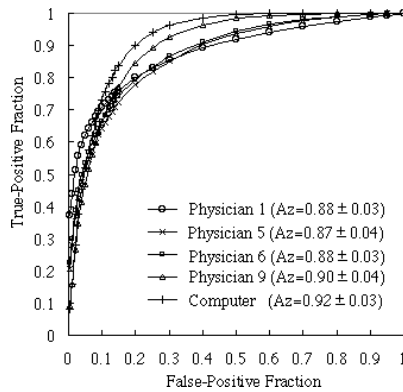


Fig. 2. ROC curves of four experienced physician and the computerized classification with the combined features space of HCT3, HCV3, and HSH3.

Based on these ranking, sensitivity, specificity, and A_z value, which denotes area under the ROC curve, were computed. Fig.1 compares the performance of the attending 11 physicians and computerized classification with the combined features space of HCT3, HCV3, and HSH3. Physicians 1, 5, 6, and 9 have 15, 17, 13, 12 years of experience in Chest radiology. The computerized classification performance is close to that of the experienced physicians. Fig.2 shows the ROC curves obtained by four experienced physicians and the computerized classification with the combined features space of HCT3, HCV3, and HSH3. The ROC curves and A_z values do not necessarily reflect the accuracy expected to be obtained under clinical conditions. Additionally, since there are several types of benign and malignant nodules a few feature parameters to characterize pulmonary nodule images are not enough to discriminate overall benign

and malignant cases automatically. Still, these results are promising that the features derived from the curvature based representation quantify one of the important clues for the differential diagnosis of the small pulmonary nodule images.

6 Conclusions

An important problem in the analysis of medical image data is the representation and characterization of normal and pathology variation in anatomical shape. In this paper, focusing on the small pulmonary nodule images, we have developed a curvature based representation of internal intensity structure of pulmonary nodules in thin-section CT images. The results of our classification study indicate that a set of histogram features derived from the curvature based representation are useful in differencing the benign and malignant pulmonary nodules. Although these results were obtained with a relatively small data set, they demonstrate the potential of using CAD techniques to analyze pulmonary nodules and to assist physicians in making diagnostic decisions. A simple prototype has been developed with good results, and ongoing work deals with characterizing the relationship with the surrounding structure of pulmonary nodules. Further studies need to be performed with a larger database to investigate the generalizability of our approach.

Acknowledgments

Authors are grateful to physicians cooperating to the reading test. The authors would like to thank Prof. Charles E. Metz for the LABROC1 program.

References

1. M. Kaneko, K. Eguchi, H. Ohmatsu, R. Kakinuma, T. Naruke, K. Suemasu, N. Moriyama, "Peripheral lung cancer: Screening and detection with low-dose spiral CT versus radiography," *Radiology*, vol.201, pp.798-802, 1996. 386
2. K. Mori, Y. Saitou, K. Tominaga, K. Yokoi, N. Miyazawa, A. Okuyama, M. Sasagawa, "Small nodular lesions in the lung periphery: New approach to diagnosis with CT," *Radiology*, vol.177, pp.843-849, 1990. 386
3. S. S. Siegelman, E. A. Zerhouni, F. P. Leo, N. F. Khouri, F. P. Stitik, "CT of the solitary pulmonary nodule," *AJR*, vol.135, pp.1-13, 1980. 387
4. A. V. Proto and S. R. Thomas, "Pulmonary nodules studied by computed tomography," *Radiology*, vol.156, pp.149-153, 1985. 387
5. M. F. McNitt-Gray, E. M. Hart, J. Goldin, C. W. Yao, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution computed tomography," *Proc. SPIE*, vol.2710, pp.1024-1034, 1996. 387
6. T. Tozaki, Y. Kawata, N. Niki, H. Ohmatsu, K. Eguchi, N. Moriyama, "Pulmonary organs analysis for differential diagnosis based on thoracic thin-section CT images," *IEEE Trans. Nuclear Science*, vol.45, pp.3075-3082, 1998. 387

7. Y. Hirano, Y. Mekada, J. Hasegawa, J. Toriwaki, H. Ohmatsu, and K. Eguchi, "Quantification of vessels convergence in three-dimensional chest X-ray CT images with three-dimensional concentration index," *Medical Imaging Technology*, vol.15, pp.228-235, 1997. **387**
8. Y.Kawata, N.Niki, H.Ohmatsu, R.Kakinuma, K.Eguchi, M.Kaneko, N.Moriyama, "Quantitative surface characterization pulmonary nodules based on thin-section CT images" *IEEE Trans. Nuclear Science*, vol. 45, pp.2132-2138, 1998. **387**
9. H. Kitaoka, R. Takaki, K. Itho, H. Kobatake, H. Ohmatsu, N. Moriyama, and K. Eguchi, "Shape analysis of pulmonary nodules in 3D-CT images with a new method of curvature estimation," *Proc. Computer Assisted Radiology and Surgery (CAR'98)*, H.U.Lemke, M.W.Vannier, and K.Inamura, Eds., pp.51-56, 1998. **387**
10. J. Bejer, T. Liebig, R. C. Bittner, P. Wust, E. Fleck, R. Felix, "Surface analysis of pulmonary lesions using fractal features," *Proc. Computer Assisted Radiology and Surgery (CAR'97)*, H.U.Lemke, M.W.Vannier, and K.Inamura, Eds., pp.228-233, 1997. **387**
11. Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Mori, K. Eguchi, M. Kaneko, N.Moriyama, "Curvature base analysis of internal structure of pulmonary nodules using thin-section CT images" *Proc. IEEE Int. Conf. Image Processing*, vol.III, pp.851-855, 1998. **387**
12. J. Hasegawa, K. Mori, J. Toriwaki, H. Anno, K. Katada, "Automated extraction of lung cancer lesions from multi-slice chest CT images by using tree-dimensional image processing," *IEICE Trans.*, vol.J76-D-II, pp.1587-1594, 1993. **387**
13. V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert, "Minimal surfaces: A three-dimensional segmentation approach," Technion Technical Report, 973, Israel, 1995. **387**
14. J. J. Koenderink and A. J. Van Doorn, "Surface shape and curvature scales," *Image and Vision Computing*, vol.10, pp.557-565,1992. **387, 388**
15. C. Dorai and A. K. Jain: "COSMOS-A Representation scheme for 3D free-form objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.19, pp.1115-1130,1997. **388, 389**
16. J. -P. Thirion and A. Gourdon, "Computing the differential characteristics of isointensity surfaces," *Comput. Vision and Image Understanding*, vol.61, pp.190-202, 1995. **387, 388**
17. S. Aylward, S. Pizer, E. Bullitt, D. Eberly, "Intensity ridge and widths for tubular object segmentation and description," *Proc. Mathematical Methods in Biomedical Image Analysis*, pp.131-138, 1996. **387**
18. Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, and R. Kikinis "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol.2, pp.143-163, 1998. **387**
19. R.O. Duda and P.E. Hart "Pattern classification and scene analysis" *John Wiley & Sons*, 1973. **389**
20. K. Fukunaga, "Introduction to statistical pattern recognition," *Academic Press, Inc., Second Edition*, 1990. **389**
21. M.C. Costanza and A.A. Afifi, "Comparison of stopping rules in forward step wise discriminant analysis," *J. of the American Statistical Association*,vol.74, pp.777-785,1979. **389**

22. C.E. Metz "ROC methodology in radiologic imaging," *Investigative Radiology*, vol.21, pp.720-733, 1986. 387, 389
23. T.Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *Int. J. Computer Vision*, vol.11, pp.283-318, 1993. 390