Pulmonary Nodule Classification Based on Nodule Retrieval from 3-D Thoracic CT Image Database

Yoshiki Kawata¹, Noboru Niki¹, Hironobu Ohmatsu², Masahiko Kusumoto³, Ryutaro Kakinuma³, Kouzo Yamada⁴, Kiyoshi Mori⁵, Hiroyuki Nishiyama⁶, Kenji Eguchi⁷, Masahiro Kaneko³, and N. Moriyama³

> ¹ Dept. of Optical Science, Univ. of Tokushima, Tokushima {kawata, niki}@tokushima-u.ac.jp ² National Cancer Center Hospital East, ³ National Cancer Center Hospital ⁴ Kanagawa Cancer Center, ⁵ Tochigi Cancer Center, ⁶ The Social Health Insurance Medical Center, ⁷ Univ. of Tokai

Abstract. The purpose of this study is to develop an image-guided decision support system that assists decision-making in clinical differential diagnosis of pulmonary nodules. This approach retrieves and displays nodules that exhibit morphological and internal profiles consistent to the nodule in question. It uses a three-dimensional (3-D) CT image database of pulmonary nodules for which diagnosis is known. In order to build the system, there are following issues that should be solved, (1) to categorize the nodule database with respect to morphological and internal features, (2) to quickly search nodule images similar to an indeterminate nodule from a large database, and (3) to reveal malignancy likelihood computed by using similar nodule images. Especially, the first problem influences the design of other issues. The successful categorization of nodule pattern might lead physicians to find important cues that characterize benign and malignant nodules. This paper focuses on an approach to categorize the nodule database with respect to nodule database with respect to nodule database with respect to nodule database of nodule.

1 Introduction

Lung cancer is a leading cause of death among men in Japan [1]. Early detection of lung cancer by means of screening thoracic CT images is considered as an effective way to reduce the mortality rate of resulting lung cancer. Physicians' diagnosis, however, becomes more difficult in correctly classifying all lesions detected at threedimensional (3-D) thoracic CT images into benign or malignant cases as the detection rate of pulmonary lesions with smaller diameter increases. Computer-aided diagnosis (CAD) has been investigated to provide physicians with quantitative information, such as estimates of the malignancy likelihood, to aid in the classification of abnormalities detected at screening [2-5]. These conventional CAD techniques provide only potential malignancies. Physicians often compare an indeterminate nodule with nodules diagnosed in the past. In order to get interpretable CAD results, it is required to relate

C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3217, pp. 838-846, 2004.

[©] Springer-Verlag Berlin Heidelberg 2004

an indeterminate nodule to similar nodules stored in a nodule database. The purpose of this study is to explore an image-guided decision support system to assist physician's decision-making of the clinical differential diagnosis of pulmonary nodules.

Some works have been done to address the problem of lesion detection and classification using similar images [6-9]. The interpretation of the pulmonary nodule images often involves the matching features extracted from a database of the nodules with associated clinical information. When the matching procedure performs well, the database provides physicians with more diagnosis and prognosis information of the queried nodule. Moreover, the corresponding structures retrieved from the database might help to design the CAD scheme for the distinction between benign and malignant nodules.

In our own recent study [9], we developed an example-based assisting approach for the nodule classification. The approach formulates the classification problem of the indeterminate nodule as one of learning to recognize nodule patterns from examples of similar nodule images. The central module makes possible analysis of the query nodule image and extraction of the features of interest: shape, surrounding structure, and internal structure of the nodules. The malignant likelihood is estimated by the difference between the representation patterns of the query case and the retrieved lesions. In order to build a decision support system using similar nodule images, there are following issues that should be solved, categorization of the nodule database with respect to morphological and internal features, quick search of nodule images similar to an indeterminate nodule from a large database, and computation of malignancy likelihood based on similar nodule images. The first problem influences the design of other issues and the successful categorization of nodule pattern might lead physicians to find important cues that characterize benign and malignant nodules. In this paper, we focus on an approach to categorize the nodule database with respect to nodule shape and CT density patterns inside nodule, and we introduce the approach to our retrieval method.

2 Methods

The nodule database used here is provided by the National Cancer Center East in Japan. The malignant cases were diagnosed by histological diagnosis and benign nodules were confirmed by surgery or diagnostic follow-up examinations. Additionally, the malignant cases have information of the stage of lesion prognosis. The database contains nodule images with known diagnosis result and consists of two elements, such as text-based and image-based elements. The text-based elements of the database are as follows: ID number, measurement conditions, gender, age, and diagnosis result. The image-based elements consist of following seven elements; (a) thin-section CT images, (b) the 3-D ROI image with a nodule of interest, (c) the segmented nodule image by a deformable surface model, (d) shape features, (e) pattern of CT density histogram, (f) the histogram-based representation using CT density value inside nodule, and (g) the histogram-based representation using curvature indexes of internal nodule. Each step is described in the following sections.

2.1 3-D Thoracic CT Images

The 3-D thoracic images used in this paper were reconstructed from thin-section CT images obtained by a multislice helical CT scanner (Toshiba Aquilion; Toshiba Medical Systems, Japan). The thin-section CT images were measured under the following conditions; beam width: 0.5mm, table pitch: 3, scan time: 0.5 sec, tube voltage: 120kV, tube current: 300mA. For the scan duration, patients held their breath at full inspiration. The slice images at 0.2 mm intervals were obtained to observe whole nodule region and its surroundings. The number of slices per patient ranges from 200 to 300 slices. The range of pixel size in each square slice of 512 pixels was between 0.3x0.3 mm² and 0.5x0.5 mm², and the slice contains an extended region of the lung area. The 3-D thoracic image was reconstructed from the thin-section CT images by a cubic B-spline interpolation technique to make each voxel isotropic [10]. The data set in this study included 174 3-D thoracic images. Of the 174 cases, 98 contained malignant nodules, and 76 contained benign nodules. Whole malignant nodules were histologically diagnosed. In benign cases lesions showed no change or decreased in size over a 2-year period were considered benign nodules.

2.2 Extraction of Nodule Region

The segmentation of the 3D pulmonary nodule image consists of three steps [3]; 1)extraction of lung area, 2) selection of the region of interest (ROI) including the nodule region, and 3) nodule segmentation based on a geometric approach. The lung area extraction step plays an essential role when the part of a nodule in the peripheral lung area touches the chest wall. The ROI including the nodule was selected interactively. A pulmonary nodule was segmented from the selected ROI image by the geometric approach. Attachments of the nodule such as vessels were eliminated by morphological filtering approach.

2.3 Feature Extraction

2.3.1 Shape Feature

In order to characterize the shape of the 3-D nodule image, we compute the direction of principal axes of the 3-D nodule image and obtain the lengths along these directions [9]. Additionally, to measure the irregularity of the segmented nodule surface we compute the compactness that is obtained by volume and surface area of thodule. Using theses values, we classify nodule shapes into six categories as shown in Fig. 1.



Fig. 1. Shape types. (a) Round shape with smooth surface. (b) Elongated shape with smooth surface. (c) Flat shape with smooth surface. (d) Round shape with irregular surface. (e) Elongated shape with irregular surface. (f) Flat shape with irregular surface.

2.3.2 Pattern Classification of CT Density Histogram Inside Nodule

Recent advantages of CT technology enable analysis of internal density patterns of small pulmonary nodules. From visual assessment of the presence of ground-glass opacities within nodule, internal density patterns are classified into solid, part-solid, and nonsolid nodules [11]. Associating internal density patterns with malignant status attracts researchers' interests. Quantification of internal density patterns is required for assisting physicians in decision of adequate treatments. In our own recent study, we investigated numerical criteria for classifying nodule density patterns that provides information with respect to nodule status [12]. The pattern classification approach of CT density distribution inside nodule might provide a nodule categorization method to let physicians estimate malignancy status. Therefore, we introduced the pattern classification approach into image-guided decision support scheme.

The pattern classification approach categorizes nodule density patterns into five types that are named as Type A, B, C, D, and E. Slice images of typical types are shown in Fig.2. The approach consists of three steps. The nodule region is volumetrically separated into two parts such as core and marginal part using distance value from the nodule surface. Then, a CT density histogram for each region is computed. These CT density histograms are named as two layers histogram. Finally, nodule density patterns are classified by using the characteristics of two layers histogram. The characterization is computed by average of each layer histogram. The second characterization is performed by the standard deviation, skewness, and kurtoisis of each layer histogram. Observing those characteristics, classification rules were designed. The two layers histograms of the nodules shown in Fig. 2 are demonstrated in Fig. 3.



Fig. 2. Typical CT density patterns of pulmonary nodules. The nodule contours extracted automatically are superimposed nodule slice images. (a) Type *A*. (b) Type *B*. (c) Type *C*. (d) Type *D*. (e) Type *F*.



Fig. 3. Two layers histogram of nodule presented In Fig.2. (a) Type A. (b) Type B. (c) Type C. (d) Type D. (e) Type F.

2.3.3 Computation of Curvature Indexes

Each voxel in the region of interest (ROI) including the pulmonary nodule was locally represented by two curvature indexes that represented the shape index and the curvedness [5]. 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 computed directly the curvatures on each voxel from the first and the second derivatives of the gray level image of the ROI.

2.4 Joint Histogram-Based Representation

In order to characterize the distribution pattern of the CT density and the shape index inside nodule, we compute two joint histograms using the distance value from the nodule center [9]. Since the distance value depends on nodule size, the first type is directly derived from the distance value and used to search similar lesions. Once obtained similar lesions, the variation of internal structure provides more important in-

formation to classify nodule patterns rather than nodule size. Therefore, the second type is derived from the normalized distance value which ranges between zero and one value and used to evaluate the likelihood of the malignancy. Statistically, the normalized joint histogram denotes a joint probability of CT density value and distance value. It measures how the CT density value distributes inside nodule with respect to the distance from the nodule center. The similar equation of joint histogram of shape index and distance values was obtained.

2.5 Searching Similar Nodule Images

We classify nodules in databases into several categories with respect to the nodule shape, pattern of CT density histogram ,and histogram-based representations. The shape categories used here are round shape, elongated shape, flat shape, round shape with irregular surface, elongated shape with irregular surface, and flat shape with irregular surface. The patterns of CT density histogram are type A, B, C, D, and E as shown in Fig. 3. It is a possible way to directly apply a similarity measure to the 3-D nodule image. However, this approach requires solving the registration between two images. In this study, we apply a simple similarity measure, which is the correlation coefficient (CC) to the nodule representation based on the joint histogram. At the first glance of a given nodule image, it is thought of that the nodule size and nodule density are important indexes for the visual assessment. First, we select the shape category and the pattern of CT density histogram to which the query case belongs. Then, the features with respect to the local intensity structure are examined in detail to search similar patterns. In this study, we generate the list of similar nodule image by the searching process based on the similarity.

2.6 Classification of Benign and Malignant Nodules

We select respectively the *M* examples from each list of the malignant and benign similar pattern to construct local two malignant and benign classes that are similar to the indeterminate case concerning the CT density distribution pattern and the nodule size. The second type of joint-histogram is used to model the distributions of retrieved malignant and benign examples. Each malignant and benign example is represented by joint histogram of shape index. The Mahalanobis distance is used as a distance measure between the query nodule and each class. The query case is classified into the cluster with small Mahalanobis distance. The Mahalanobis distance from the retrieved benign and malignant classes are represented by MDB and MDM, respectively. A decision index (*DI*) is formulated by MDB/MDM. When *DI* is larger than 1, the query nodule is classified into malignant class. While the query case having smaller *DI* than one has less malignancy likelihood.



Fig. 4. Retrieval results. (a) Snapshot of searching similar nodule cases from our data set. (b) Expansion of window 2. (c) Expansion of window 3. (d) Expansion of window 4.

3 Experimental Results

Fig. 4 presents a snapshot of searching similar nodule cases from our data set. The window 1 shows a slice image of a query nodule. The window 2 and 3 show retrieval results of benign and malignant cases, respectively. These similar cases were selected from the same shape type and CT density histogram pattern as those of the query case. The retrieved cases are sorted in the order of the *CC* value. The diagnosis results of the similar cases are presented simultaneously. The window 4 presents slices of region of interests including the query nodule and a malignant case selected from the retrieval result. This window also presents CT density histograms of them. The shape index distribution and histogram can be presented in the same window 4.

In order to compare with retrieval results based on combination with shape type, CT density histogram pattern, and *CC* value, retrieval results were obtained by other combinations. From visual assessment, it is found that the lack of shape type or CT density histogram pattern from similarity criterion results in retrieving the dissimilar. To evaluate the categorization of data set quantitatively, it is though that an evaluation of *DI* from similar cases obtained from each categorization may be one of approaches.

We computed DI values for the following five categorizations: shape type alone, pattern of CT density histogram alone, CC value alone, combination with shape type and CC value, combination with pattern of CT density histogram and CC value, and combination with whole categories. These results of DI computation of the query cases presented in Fig. 4 were 0.87, 1.1, 0.91, 2.0, 2.3, and 2.3, respectively. Since the query case is a malignant nodule, the correct classification is that the query case has DI value larger than 1. In this case, it is found that the combination of categorizations provides better classification performance. More researches using a large data set will be required to conclude in detail. Still, the preliminary experimental result encourages us to develop an image-guided decision support scheme combining the nodule categorization techniques.

4 Conclusion

We have explored a nodule categorization approach to develop an image-guided decision support system of pulmonary nodules. The main idea is to give representations of nodule internal density patterns by introducing a two layers histogram. We presented application results of searching similar images for a query nodule. We believe that the proposed approach might be considered as one of important technique for that imageguided decision support system. The proposed approach could be expanded to include prognosis of lung cancer and build a model showing a likelihood of patient's diseasefree survival.

References

- Kaneko M., Eguchi K., Ohmatsu H., Kakinuma R., Naruke T., Suemasu K., Moriyama N.: Peripheral lung cancer: Screening and detection with low-dose spiral CT versus radiography. Radiology. 201 (1996) 798-802
- Cavouras D., Prassopoulos P. and Pantelidis N.: Image analysis methods for solitary pulmonary nodule characterization by computed tomography. European Journal of Radiology, 14. (1992) 169-172
- Kawata Y., Niki N., Ohmatsu H., Kakinuma R., Eguchi K., Kaneko M., Moriyama N.: Quantitative surface characterization of pulmonary nodules based on thin-section CT images. IEEE Trans. Nuclear Science. 45 (1998) 2132-2138
- 4. McNitt-Gray M. F., Hart E. M., Wyckoff N. Sayre J. W., Goldin J. G., and Aberle D. R. : A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results. Medical Physics. 26 (1999) 880-888
- 5. Kawata Y., Niki N., Ohmatsu H.: Curvature based internal structure analysis of pulmonary nodules using thoracic 3-D CT images. Trans IEICE, J83-DII (2000) 209-218
- Giger M. L., Huo Z., Lan L., Vyborny C. J.: Intelligent search workstation for computeraided diagnosis. Proc. Computer Assisted Radiology and Surgery. (2000) 822-827

- Kawata Y., Niki N., Ohmatsu H., Kakinuma R., Mori K., Eguchi K., Kaneko M., Moriyama N. : Three-dimensional CT image retrieval in a database of pulmonary nodules. Proc. IEEE International Conference on Image Processing. vol. III (2002) 149-152
- Li Q., Aoyama M., Li F., Sone S., MacMahon H. M., Doi K.: Potential clinical usefulness of an intelligent computer-aided diagnostic scheme for distinction between benign and malignant pulmonary nodules in low-dose CT scans. RSNA 2002 Scientific Program (2002) 534-535
- Kawata Y., Niki N., Ohmatsu H., Moriyama N.: Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images. Academic Radiology. 10 (2003) 1402-1415
- Unser M., Aldroubi A., and Eden M.: B-spline signal processing: Part II: Efficient design and applications. IEEE Trans. Signal Processing. 41 (1993) 834-848
- Henschke C. I., Yankelevitz D. F., Mirtcheva R., McGuinness G., McCauley D., Miettinen O. S.: CT screening for lung cancer: Frequency and significance of part-solid and nonsolid nodules. American Journal Roentgenology. 178 (2002) 1053-1057
- 12. Kawata Y., Niki N., Ohmatsu H., Kakinuma R., Kaneko M., Moriyama N. : A classification method of pulmonary nodules using 3-D CT images. 89th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA). (2003) 648