Computer-Aided Diagnostic System for Pulmonary Nodules Using Helical CT Images

K.Kanazawa[1], Y.Kawata[2], N.Niki[2], H.Satoh[3]
H.Ohmatsu[4], R.Kakinuma[4], M.Kaneko[5], K.Eguchi[6], N.Moriyama[4]

[1] Dept. of Information Engineering, Takuma National College of Technology
[2] Dept. of Optical Science, University of Tokushima, Japan
[3] Medical Engineering Laboratry, Japan
[4] National Cancer Center Hospital East, Japan
[5] National Cancer Center Hospital, Japan
[6] National Shikoku Cancer Center Hospital, Japan

Abstract. In this paper, we present a computer assisted automatic diagnostic system for lung cancer that detects nodule candidates at an early stage from helical CT images of the thorax. Our diagnostic system consists of analytic and diagnostic procedures. In the analysis procedure, we extract the lung and the blood vessel regions using the fuzzy clustering algorithm, then we analyze the features of these regions. In the diagnosis procedure, we define diagnostic rules utilizing the extracted features which support the determination of the candidate nodule locations.

1   Introduction

According to figures of cancer incidence in Japan, mortality from cancer is increasing annually with the present rate for cancer prevailing in a quarter of all deaths. Among cancerdeath causes, lung cancer is the most common one, accounting for 21.4% of all male cancer deaths cases in Japan. A possibility of opposing the lung cancer is its detection and treatment at an early stage of growth. As a conventional method for the mass screening process, chest X-ray films have been used for lung cancer diagnosis. Since the chest X-ray films are two-dimensional projection images, the overlapping of bone and organs shadows results in disturbing the detection of small lung cancer nodules at early stage.

Recently, chest CT images obtained by helical CT scanner which can make a wide range measurement of the lung region in a short time have drawn interest in the detection of suspicious regions[1]-[3]. However, mass screening based on helical CT images leads to a considerable number of images to be diagnosed which is time-consuming and makes its use difficult in the clinic. In order to increase the efficiency of the mass screening process, our group is working towards a computer-aided diagnosis system for pulmonary nodules detection based on helical CT images. We have already developed a prototype computer-aided diagnosis (CAD) system to automatically detect suspicious nodules from chest CT images and alert physicians attention. The ingredients of the system are the image analysis and the image diagnosis parts. The first part consists of the following procedures; extraction of the lung area, reduction of the partial volume

effect and beam hardening, extraction of the region of interest (ROI), and feature extraction from each ROI. The second part procedures detect suspicious nodules and alert physicians attention by applying their proposed diagnosis rules. In this paper, we describe our algorithm of computer-aided diagnosis for pulmonary nodules and we show the performance of the prototype CAD system by using image data of 450 cases.

2 Diagnostic Algorithm

2.1 Image Analysis Procedure

The prototype CAD system is designed to detect the candidates of lung cancer from helical CT images consists of two procedures; image analysis procedure and image diagnosis procedure. The input to this system is helical CT image acquired under the specified measurement conditions shown in Table.1. These images are reconstructed at 10mm increment by using an 180 linear interpolation algorithm and a matrix size of 512×512 pixels with 12 bits quantization. From every patient we collect 35 images with the above mentioned measurement conditions. The diagnostic algorithm is applied to each slice image from upper to lower lung sequentially.

The first procedure of the system deals with the extraction of lung area, the reduction of the partial volume and the beam hardening effects, the extraction of the region of interest(ROI), and finally the feature extraction of ROI.

[Step1] Extraction of the Lung Fields

The lung fields are mostly occupied by air, CT value of the pixels within the lung fields are lower than CT values of the lung wall. Firstly, the original image is transformed to a binary image using a gray-level thresholding algorithm with a threshold value t. Secondly, we eliminate the binary regions, external to the body, trachea and stomach. The tracheal regions are determined as isolated regions which locate at the central region of the body, and the stomachic region are determined as isolated regions which can not exist on the upper slice image. The edges of the remainder regions are defined as the lung contour, the regions within the lung contour are defined as the initial lung fields.

However, this thresholding technique excludes the lung boundary with high CT values from the real lung fields. In order to avoid this problem the lost parts are compensated by using curvature at the boundary of the initial lung fields contour. Therefore, we replace the contour with rapidly changing curvature with a straight line.

[Step2] Analysis of the Lung Fields

As shape of the lung fields and the structures in the lung are seen differently, according to the position of the lung in each slice, the diagnostic parameters have to be adjusted appropriately. Therefore, we classify the lung into four sections based on the area and the shape of the extracted lung regions as shown in Fig.1. In this process, the border between sections is detected through extracting the lung fields sequentially slice by slice.

[Step3] Reduction of the Artifacts

Structures in the lung, such as blood vessels and nodules, are individually defined by a segmentation process. However, the helical CT images involve artifacts by the beam hardening effect and the partial volume effect. These artifacts affect the ROI extraction procedure and present bad extraction result. These artifacts are reduced by a simple technique which used to smooth the lung field's image based on mathematical morphology operation. Then, we apply a smoothing method to the extracted lung images and we subtract this smoothed image from the original image in the lung area. These procedure is used as pre-segmentation process.

[Step4] Extraction of ROI

The interesting organs in the lung, such as blood vessels and nodules, are segmented within the lung fields obtained using fuzzy clustering method. The lung field images are segmented into two clusters, air cluster and other organs cluster. To obtain the binary regions of ROI, firstly, the gray-weighted distance transformation is applied to the segmented image, to which we apply the threshold algorithm in order to eliminate pixels lower than the threshold distance k. Secondly, we apply an inverse distance transformation to extract the ROI. The small size regions contained in the organs clusters are eliminated by this procedure.

The artifacts are reduced by the morphological method described in the previous step, however, the artifact near the circumference of the lung fields are not removed satisfactorily. Thus, to remove such artifacts, we use surface curvatures described in [3]. The ROI tend to be the convex and hyperbolic surface regions on the intensity surfaces. To extract convex and hyperbolic surface regions the calculated curvatures signs are utilized. The signs of these surface curvatures segment the intensity surface into eight basic viewpoint-independent surface types, such as peak surface ($K > 0$ and $H < 0$), flat surface ($K = 0$ and $H = 0$), pit surface ($K > 0$ and $H > 0$), minimal surface ($K < 0$ and $H = 0$), ridge surface ($K = 0$ and $H < 0$), saddle ridge ($K < 0$ and $H < 0$), valley surface ($K = 0$ and $H > 0$), and saddle valley ($K < 0$ and $H > 0$). Since it is difficult to decide whether K and H are zero or not, the four surface types including flat surface, minimal surface, ridge surface, and valley surface are not assigned in practical experiments. In this step, we consider the following surface types regions as ROI; pit surface, saddle ridge, and saddle valley surface types. We apply this extraction method using surface curvatures to the regions where artifacts tend to arise. To put it in the concrete, we extract two lung fields with different threshold values, this extraction method is applied to the different region between the two lung fields.

[Step5] ROI Feature Extraction

Most of the extracted ROI are blood vessel regions. We need to identify ROI as blood vessels or nodules. Thus, we give attention to shape, gray value and position of each ROI. This step extracts the features of each ROI. The following features are considered; F_a : Area, F_t : Thickness, F_c : Circularity, F_g : Gray-level, F_v : Variance of gray-level, F_l : Localization, F_h : Variance of gradient, F_p : Distance from the lung wall.

F_a is the number of the pixels of each ROI. F_t is the maximum gray-weighted distance in each ROI, F_c is the rate of occupation inside the circumscribed circle, F_g is the mean CT number of the pixels, F_v is the variance of the CT numbers, F_l is the rate of the isolation, F_h is the variance of the gradient on the boundary of each ROI, F_p is the minimum distance between the circumscribed circle's center of each ROI and the lung wall. The contour of the lung wall is defined as the section of lung contour which cuts out the mediastinal contour.

2.2 Image Diagnosis Procedure

The second procedure of our prototype system detects suspicious region and alerts physicians attention by applying their proposed diagnosis rules. Three diagnostic rules are constructed based on the following medical knowledge;

Point 1 : Lung cancer shape is generally spherical, and is seen as a circle on the cross section. Contrariwise the shape of blood vessels running parallel to the slice image are generally oblong.

Point 2 : The thickness of the blood vessel becomes smaller as its position approaches the lung wall, however the thickness of the lung cancer is generally larger than the ordinary thickness of the blood vessels at each position.

Point 3 : Since the peripheral blood vessels are too small to be seen in the helical CT image and are difficult to recognize, shadows contacting the lung wall are generally nodules or partial volume artifacts.

Point 4 : The gray values of blood vessels running vertical with respect to the slice image are generally higher than the same sized cancer in helical CT images with the specified measurement conditions for mass screening.

Point 5 : The values of pixels of the cancer region are comparatively uniform.

We define the diagnostic rules combining features of ROI to detect suspicious nodules. The diagnostic rules are classified into following three types;

[RULE 1] : Regions for elimination

We eliminate the current ROI if the following conditions are satisfied.

(1) Our diagnostic algorithm aims to detect cancer with diameter over 4mm. A ROI with Thickness F_t under 2.0 (corresponding to approximately 4mm) is eliminated.

(2) A ROI for which the percentage of Circularity F_c is under 20% is eliminated as a blood vessel region according to Point 1.

(3) In case where bone exists in the adjacent slice images at the location for the current ROI, we eliminate this ROI as a shadow of a partial volume artifact arising from bone.

[RULE 2] : Detection of nodules in case of non-contact with the lung wall

In this rule, we quantify the likelihood of the lung cancer utilizing the extracted features for each ROI. Here we define three quantities as follows;

Roundness : We think that the likelihood of lung cancer for each ROI depends on Circularity F_c. Roundness is in proportion to Circularity F_c.

Size : The area generally becomes larger with distance from the lung wall. We assume the basic thickness which is varied in proportion to Position F_p, and we define Size as the ratio of F_t to the assumed thickness.

Intensity : The cross section of a blood vessel running vertical to the slice image displays high CT values according to Point 4. Thus, Intensity becomes smaller as the gray value F_g becomes large compared with the average value of lung cancer.

Furthermore, we define a likelihood index for lung cancer as follows;

Diagnosis index = Roundness × Size × Intensity

We classify the current ROI as a nodule candidate if this index is over a threshold.

[RULE 3] : Detection of nodules in case of contact with the lung wall

Shadows contacting the lung wall are ordinarily suspicious as described by Point 3, however these shadows can involve the artifact values which are caused by the partial volume effect. Here three characteristics can be observed, nodules have a convex shape inside the lung, artifacts are oblong along the lung wall and the gradient around nodules is higher than artifacts. Therefore to distinguish between nodules and artifacts, we define two indices of Convexness and Contrast as follows;

Convexness : For the three pixel dilation area outside the current ROI, we count the number of pixels which are higher than the average value F_g, and similarly count the number of pixels which are lower than the F_g. We define the Convexness as ratio of these numbers.

Contrast : We define the Contrast as the difference between the average values of the current ROI and the low intensity area outside this ROI.

2.3 Application of Diagnostic Algorithm

This section presents the application of our diagnostic algorithm to Helical CT images concretely width example of result.

Fig.2 presents the process of extraction of lung fields.

Fig.3 presents the extraction process of ROI. Fig.3(a) is the extracted lung fields. The result of the artifacts reduction shows in Fig.3(b), and the Fig.3(c) is the segmented image by using the fuzzy clustering method. Fig.3(d) shows the result of the surface segmentation by using surface curvatures. In Fig.3(e) the defined color-code corresponding to each surface type is as follows, peak surface is white, pit surface is green, saddle ridge is blue and saddle valley is red. Fig.3(e) is the finally extracted ROI. Then, we extract three type ROI as shown in Fig.3(f) using the threshold distance $k = 0.25(green), 1.5(orange), 3.0(red)$ (corresponding to approximately 0.5mm, 3mm, 6mm).

Fig.4 is an example of detection result. In the original image Fig.4(a) which involves four lung cancers where yellow-colored circle indicate. Fig.4(b) presents the result of the diagnosis, our results show nodule candidates as red-colored regions. Our system was able to detect all lung cancers nodules candidates.

3 Experimental Results

To evaluate the performance of the prototype CAD system, the system was applied to the helical CT images of 450 subjects (total: 15,750 slice images). These data were collected in trial mass screening which has been carried out since 1993 in Japan. All data are diagnosed by three expert physicians with the criterion as shown in Table.2.

Table.3 shows the comparative results between the physicians and the prototype CAD system. "definite" are definitive nodules which need further examinations, and the physicians considered these nodules highly suspicious for malignancy. "suspicious" are the nodules also need further examinations, however, the radiologists considered these nodules less likely malignancy compared to the nodules with "definite". A total of 230 nodules were detected by the three physicians. Here we separate each shadow into six groups according to the physician's judgment and the number of physicians who detected it. In this experiment, our system separates all detected nodule candidates into two clusters. The highly suspicious nodules are classified as "Class 1", and other detected shadows are classified as "Class 2". As a result, all the "definite" nodules and the "suspicious" nodules which were identified by the three physicians were detected perfectly as "Class 1". Then, 42 "suspicious" nodules were detected by two of the physicians, these nodules have been classified by CAD system as 36 of them in "Class 1" and 5 in "Class 2", but one nodule was not detected. Then, 167 "suspicious" nodules were detected by one of the physicians, these nodules have been classified by CAD system as 117 of them in "Class 1" and 28 in "Class 2", but 22 nodules were not detected. The sensibility for 230 nodules was 90% (including "Class 2").

Table.4 shows the comparative results between three physicians diagnosis by agreement and the CAD system. CAD system could not detect one nodule which was detected by two physicians also could not detect 5 nodules which have been detected by one of physician as "suspicious". In comparison with these results, the number of the nodules which have been detected by one of the physicians decreased considerably. This fact means that at least one physician changed his judgment, and these nodules are difficult to be detected. 5 cases of 6 false negative cases are involved width such nodules. The sensibility for 120 nodules was 95% (including "Class 2").

Table.5 shows the comparative result between one physician and CAD system. In Table.5, physician A diagnosed 102 of 120 nodules which were detected by agreement. He detected 80 nodules as "definite" or "suspicious", 13 nodules as "non malignant", and 9 nodules were not detected. CAD system detected 85 nodules as "Class 1"($\approx$"definite"), 12 nodules as "Class 2"($\approx$"suspicious"), and 5 nodules were not detected. The sensibility of the CAD system ranks with expert physician.

4 Conclusion

We have developed a computer assisted automatic diagnosis system for pulmonary nodules using helical CT images. We applied it to image data from 450 patients. Experimental results of our CAD system have indicated good performance when compared with physician diagnosis. Currently, we are carrying out the clinical field test program using the CAD system since '97 June. In the field test, first, a physician diagnoses the CT images without the results of CAD system, and after, the same physician diagnoses the same images referring the CAD results. Then, we compare the former and later physician's diagnostic results, and we evaluate the effectiveness of our CAD system for lung cancer screening.

References

1. M.L.Giger, K.T.Bae, and H.MacMahon: "Computerized Detection of Pulmonary Nodules in computed Tomography Images", Invest Radiol, vol.29, no.4, pp.459-465, 1994.
2. K.Kanazawa, K.Kubo, N.Niki, H.Satoh, H.Ohmatsu, K.Eguchi, N.Moriyama: "Computer Aided Screening System for Lung Cancer Based on Helical CT Images", Visualization in Biomedical Computing, Lecture Notes in Computer Science 1131, Springer, pp.223-228, 1996.
3. Y.Kawata, K.Kanazawa, S.Toshioka, N.Niki, H.Satoh, H.Omatsu, K.Eguchi, N.Moriyama: "Computer Aided Diagnosis System for Lung Cancer Based on Helical CT Images", Lecture Notes in Computer Science 1311, Image Analysis and Processing, Springer, pp.420-427, 1997.

Table.1 Measurement condition
of the helical CT images.

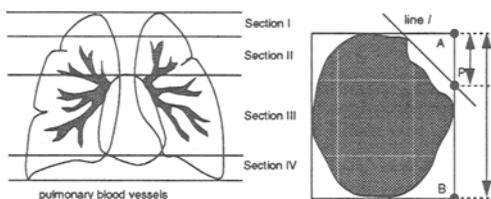| Beam Width	10mm
Table Speed	20mm/s
Tube Voltage	120kV
Tube Current	50mA
Scan Duration	15sec

Fig.1. Section classification of the lung
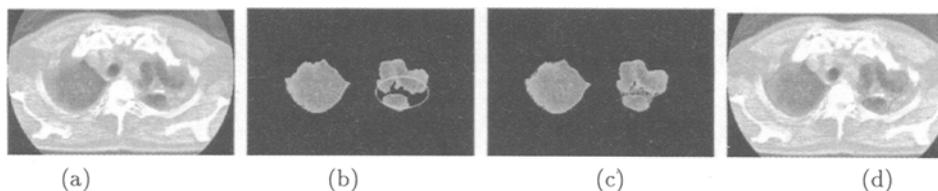

(a) (b) (c) (d)

Fig.2. Results of the extraction of the lung fields: (a) Original image; (b) Application of thresholding algorithm; (c) Correction of the lung contour; (d) Extracted lung fields.
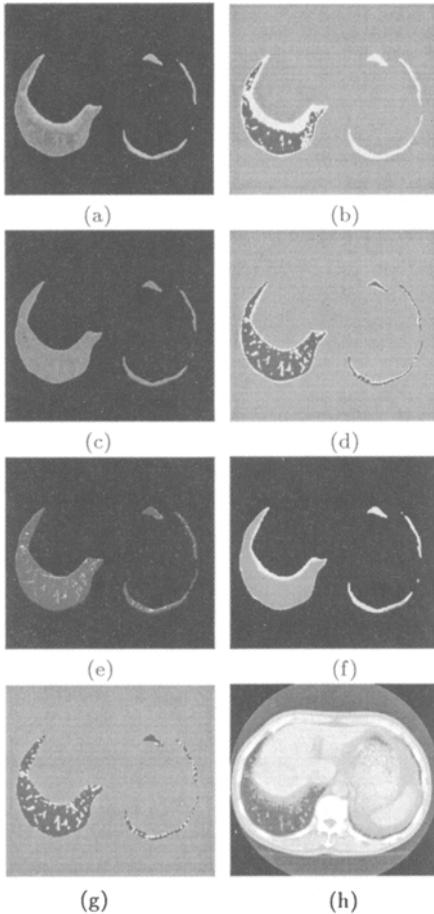
(a)   (b)

Fig.4. An example of detection result.

Table.2 Criterion of judgment.

Judgment E	definite malignant lesion
Judgment D	suspicious malignant lesion
Judgment C	non malignant lesion
Judgment B	normal case

Table.3 Comparison between physicians and the CAD system's results.

		Physician	CADsystem		
	Drs.	Nodules	Class1	Class2	FN
E	by 3 Drs.	1	1	0	0
	by 2 Drs.	2	2	0	0
	by 1 Dr.	8	8	0	0
D	by 3 Drs.	10	10	0	0
	by 2 Drs.	42	36	5	1
	by 1 Dr.	167	117	28	22
	Total	230	174	33	23

Table.4 Comparison between agreemental results and the CAD system

		Physician	CADsystem		
	Drs.	Nodules	Class1	Class2	FN
E	by 3 Drs.	1	1	0	0
	by 2 Drs.	2	2	0	0
	by 1 Dr.	8	8	0	0
D	by 3 Drs.	10	10	0	0
	by 2 Drs.	36	30	5	1
	by 1 Dr.	63	49	9	5
	Total	120	100	14	6

Table.5 Comparison between physician A and the CAD system.

			Physician A			CADsystem		
	Drs.	Nds	E,D	C	FN	C1	C2	FN
E	by 3 Drs.	0	1	0	0	1	0	0
	by 2 Drs.	1	2	0	0	2	0	0
	by 1 Dr.	7	6	1	0	7	0	0
D	by 3 Drs.	9	9	0	0	9	0	0
	by 2 Drs.	32	30	2	0	26	5	1
	by 1 Dr.	51	32	10	9	40	7	4
	Total	102	80	13	9	85	12	5

Fig.3. Results of the extraction of ROI: (a) Extracted lung fields image; (b) Thresholding result; (c) Artifact reduction results; (d) Segmentation result using fuzzy clustering method; (e) Surface segmentation result by using surface curvatures; (f) Extracted lung fields by using different two threshold values; (g) Extracted ROI; (f) Three types ROI using different threshold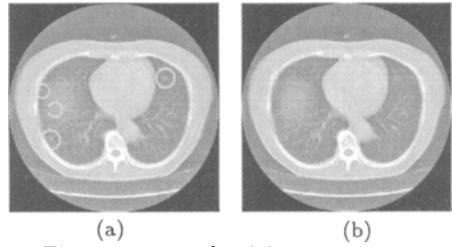 distance k=0.25(green), k=1.5(orange) and k=3.0(red).