**World Scientific**
www.worldscientific.com

# CURVATURE-BASED CORRECTION ALGORITHM FOR AUTOMATIC LUNG SEGMENTATION ON CHEST CT IMAGES

SHICHENG HU[*],[¶], KESEN BI[§], QUANXU GE[§],
MINGCHAO LI[†], XIN XIE[‡] and XIN XIANG[†]

[*]School of Economics and Management

[†]Department of Mathematics

[‡]School of Computer Science and Technology
Harbin Institute of Technology
No. 2 West Wenhua Road, Weihai 264209, P. R. China

[§]Department of CT, Weihai Municipal Hospital
No. 70 Heping Road, Weihai 264200, P. R. China
[¶]hu_shicheng@aliyun.com

In order to ameliorate the lung defects caused by missed juxtapleural nodules in lung segmentation on chest computed tomography (CT) images, we develop a Newton–Cotes-based smoothing algorithm (NCBS) which is used as a preliminary step to remove noises as many as possible. Next considering the crescent outline features of the lung, we propose a curvature-based correction algorithm (CBC) for the determination of the correction threshold. The application of the proposed algorithms is demonstrated in the process of lung segmentation and the experimental results on 25 real datasets are illustrated. Furthermore, some experiments are conducted to investigate the effects of the key parameters in CBC on the performances of lung segmentation so as to decide their optimal values. In addition, the CBC is compared with other methods analytically and experimentally. The overall results show that our proposed algorithm in lung segmentation excels the related methods on the capability of automatic selection of the correction threshold, as well as the performances of accuracy, efficiency and feasibility.

*Keywords*: Lung Segmentation; Computer-Aided Detection; Medical Image Processing; Lung Nodules.

## 1. Introduction

According to statistics from the American Cancer Society, lung cancer is the primary cause of cancer-related deaths worldwide.[1] Early detection of potentially cancerous pulmonary nodules may be a way to improve a patient's chances for

---

[¶]Corresponding author.

survival.[2] Computed tomography (CT)-based computer-aided detection (CAD) is the most commonly used diagnosis technique due to its high sensitivity of small pulmonary nodules and flexible representation of the human thorax.[3] A CAD system is composed of several successive processes, such as candidate nodule detection, feature computation, classification, etc. In recent years, a number of CAD systems have been proposed and have achieved many outstanding progresses: thresholding,[4–6] mathematical morphology,[7] template matching,[8] clustering,[9–12] model-based method[13,14] are widely used for candidate nodule detection. More and more features such as gray-level values[15–18] and morphological values[19–22] are found to be very effective for nodule differentiation. And many classifiers, e.g., linear discriminant analysis (LDA),[15,16,23,24] artificial neural network (ANN),[20,21,25,26] genetic algorithm (GA)-based method,[8] Bayesian classifier,[27,28] clustering,[29,30] Markov random field analysis[31] and support vector machine (SVM),[32,33] etc. are designed for the distinction between malignant and benign nodules. As a preprocessing step prior to these analyses in a CAD system, lung segmentation is often used to reduce computation and limit extraneous results. Since 5–17% of the lung nodules in test data may be missed in a nodule detection setting due to the preprocessing,[34] automatic lung segmentation has been regularly incorporated into a CAD system as a necessary first step for computer analysis.

Thresholding methods combined with region growing[4,5,7–11] are commonly used to extract lung parenchyma, by which those tissues with higher gray level than the selected threshold are excluded from the thoracic region. Because the juxtapleural lung nodules have similar densities with their surrounding tissues, they will be removed from the lung area with their surrounding tissues and will never be recognized in the later stages. As the number of juxtapleural nodules accounts for as much as 17% of all the nodules in lung parenchyma,[34] the accuracy of lung segmentation can have a significant effect on the detection sensitivity of a CAD system. Small inaccuracies in inclusion of juxtapleural nodules could lead to inaccurate volumetric measurement and estimation of malignancy.[35] But so far, no evident threshold is reported to distinguish juxtapleural nodules from their surrounding tissues, ameliorating the under-segmentation of juxtapleural lung nodules has to be solved by image analysis algorithms. Designing an efficient and robust segmentation algorithm is difficult because of the highly varied properties of juxtapleural nodules such as shape, size, intensity and location. Additionally, a good solution should not only include juxtapleural nodules properly, but also minimize under-segmentation of lung structures and over-segmentation of non-lung structures.[36]

In earlier research,[37] manual editing approaches are used to ameliorate the under-segmentation of juxtapleural nodules, but that is extremely laborious and prone to intra- and inter-observer variability. Therefore, some algorithms have been proposed to correct the under-segmentation of juxtapleural nodules automatically. The rolling ball algorithm (RBA) is one of the early used methods to ameliorate the lung contour that is determined by thresholding.[4] The rolling ball is a circular

filter with a predefined radius. During scanning, the ball is successively placed tangential to each contour point. An indentation is identified when the ball filter contacts the contour at more than one locations. A new contour segment that linearly connects the endpoints of the indentation is constructed to bridge the indentation, therefore the image pixels that are newly encompassed by the contours are included within the lung segmentation regions. Gurcan *et al.*[10] designed an indentation detection method (IDM) for ameliorating the initial lung contour that is recognized by thresholding. This method computes the ratio between the minimum of the geodesic distances of two points on the border and the Euclidean distance of the linear segment formed by the same two points. If the ratio is larger than a predefined threshold, the lung border between the two points will be replaced by the straight line through the two points. Bellotti *et al.*[38] proposed a new active contour, called glued elastic band (GEB), to correct the contour of lung parenchyma. For each point on the contour, he introduces two types of forces, a constant adhesive force which is always directed to the inside of the contour and a compound force, which is the sum of two elastic forces that are determined by the closest points of the considered one. If the compound force and the adhesive force point to the opposite direction and their magnitude ratio (compound force/adhesive force) is larger than a predefined threshold, the lung border between the two closest points will be replaced by the straight line connecting the two points, therefore the considered point is encompassed in the lung parenchyma. Pu *et al.*[36] devised an adaptive border marching algorithm (ABM) to tackle this problem. In this algorithm, if a straight line between a start point and the considered one within the marching step is found to be on the right hand of a reference line, the reference line will be replaced by the straight line, therefore those points between the start point and the considered one will be included in the lung parenchyma. In the process, with a scaling factor the marching step length can be changed from an initial value to an optimized value corresponding to the specific appearance of a juxtapleural nodule under an adaptive condition in the expression that the ratio of the height to the chord length of the arc, which is composed of pixels between the start point and the considered one on the border, is smaller than a predefined threshold. Other solutions, which either focus on only partial features of juxtapleural nodules or are the variants of some of the algorithms,[39] such as Kim's texture-based analysis,[40] Wiemker's circular or semicircular structure searching along the lung wall[41] and Hu's morphological operations,[42] are also used to include missed juxtapleural nodules.

These scenarios demonstrate that some progresses have been made in lung segmentation. But due to the disadvantages of the employed algorithms, several challenges are still involved in this area, among which the most difficult is the threshold determination for contour correction. Algorithms such as RBA,[4] IDM,[10] GEB,[38] ABM,[36] etc. have been used to ameliorate lung region. But most of these algorithms have the difficulty of determining appropriate correction thresholds. Such

thresholds, for example, the radius in RBA, the ratio between the geometric and direct distances in IDM, the ratio between adhesive and internal forces in GEB, etc. correspond to the volumes of those juxtapleural nodules that will be included in the lung parenchyma, therefore, they have great impacts on these algorithms and should be determined with ground evidence. But unfortunately, most of the thresholds are decided empirically.

In this paper, a curvature-based correction algorithm (CBC) for automatic lung segmentation on chest CT images is proposed, in which the automatic selection of a threshold for the correction of the lung segmentation defects caused by juxtapleural nodules is more heavily concerned. This paper is organized as follows: In Sec. 2, the CBC is developed for the determination of the correction threshold after the establishment of Newton–Cotes-based smoothing algorithm (NCBS). The demonstration of how the proposed algorithms are applied in the process of the lung segmentation is given in Sec. 3. The experimental results of CBC on 25 datasets are illustrated in the first part of Sec. 4. In the second and third parts of Sec. 4, we investigate the effects of the key parameters in CBC and compare it with several other methods, respectively. The summary is given in the final section.

## 2. Methods

In lung segmentation, the initial contours drawn by thresholding-based methods are not smooth [see Fig. 1(a)] in general. From these rough contours we find it very difficult to choose an appropriate threshold automatically. So before the determination of a threshold, the rough contours should be smoothed in order to remove noises as many as possible. The CBC comprises of two stages: Threshold determination and correction implementation. The threshold determination is based on a smoothed contour produced by NCBS. After the initial contour is disposed with



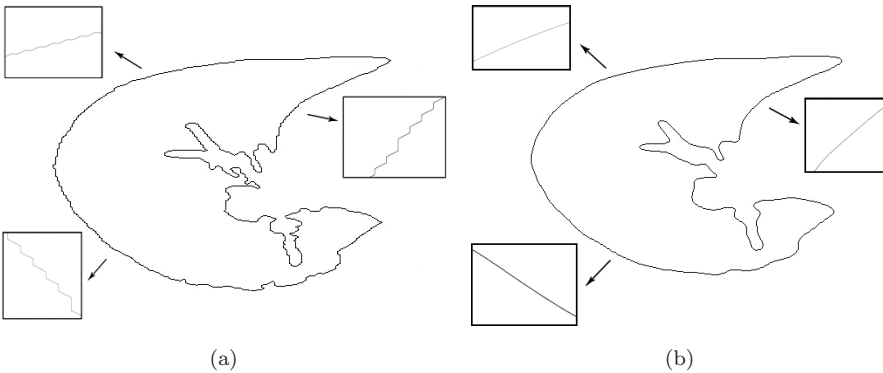(a)                                    (b)

Fig. 1.   The comparison between a rough contour and that after smoothed. (a) The rough contour drawn by thresholding method. The three magnified segments demonstrate the roughness of the contour. (b) The smooth contour produced by NCBS. The three magnified segments demonstrate the smoothness of the contour.

the smoothing algorithm, the CBC will compute the curvature for each pixel on the contour and decide a threshold to carry out the correction. The development of the algorithms is detailed in the following.

## 2.1. *Newton–Cotes-based smoothing algorithm NCBS*

For curve smoothing, several popular methods have been proposed, e.g., Gaussian smoothing (GS) or Fourier transform-based smoothing (FTS). As to GS, the key parameter, i.e., a proper standard deviation is not easy to be discovered for the smoothing of lung contour. If FTS is used, the amount of calculation for transformation is very great.[43] As the number of points on a lung contour is very large, it will not meet the real-time processing requirement in applications. The NCBS can overcome these disadvantages. With respect to the crescent outline of the lung border, its implementation is more direct and simpler and its key parameters are easy to be tuned to their optimal values.

The NCBS is built on Cotes-coefficient-based fitting formula. Before the fitting formula is brought forward, we give the Newton–Cotes integral formula which is derived from the mean value theorem of integrals. Given the function $y = f(x)$, let the integral range $[a, b]$ be divided into $n$ equal parts, the step length $h$ is represented as $h = \frac{b-a}{n}$, the $n + 1$ equal distance nodes are denoted as $x_k = a + kh (k = 0, 1, \ldots, n; x_0 = a, x_n = b)$. Then from Ref. 44 the Newton–Cotes integral formula can be represented as

$$I_n = (b - a) \sum_{k=0}^{n} C_k^n f(x_k), \tag{2.1}$$

where $I_n$ is the integral value, $C_k^n (k = 0, 1, \ldots, n)$ is called the Cotes coefficient and can be formulated as

$$C_k^n = \frac{h}{b-a} \int_0^n \prod_{j=0, j \neq k}^{n} \frac{t-j}{k-j} dt = \frac{(-1)^{n-k}}{nk!(n-k)!} \int_0^n \prod_{j=0, j \neq k}^{n} (t-j) dt. \tag{2.2}$$

In order to derive the Cotes-coefficient-based fitting formula, we give the mean value theorem of integrals as follows:

$$\int_a^b f(x) dx = (b - a) f(\xi), \tag{2.3}$$

where $\xi$ is a value between $a$ and $b$. Let the probability density function of variable $x$, $g(x)$, be denoted as

$$g(x) = \begin{cases} \dfrac{1}{b - a}, & a \leq x \leq b \\ 0, & \text{else} \end{cases}. \tag{2.4}$$

Since

$$E(f(x)) = \int_{-\infty}^{+\infty} f(x) g(x) dx = \int_a^b \frac{1}{b - a} dx, \tag{2.5}$$

we get

$$\int_a^b f(x)dx = (b-a)E(f(x)). \tag{2.6}$$

Considering formula (2.3) and (2.6), we conclude

$$E(f(x)) = f(\xi). \tag{2.7}$$

With respect to Newton–Cotes integral formula (2.1) together with formula (2.6), $E(f(x))$ can be approximated as below:

$$E(f(x)) \approx \sum_{k=0}^n \mathrm{C}_k^n f(x_k). \tag{2.8}$$

The above approximation is called Cotes-coefficient-based fitting formula which can be expressed as: If the variable $x \in [a, b]$ obeys uniform distribution, then the expectation of $f(x)$, $E(f(x))$, can be approximated to formula (2.8).

It is noted that when the condition that $C_k^n > 0 (n \le 7)$ is satisfied, the Cotes-coefficient-based fitting formula is stable, which means that the initial input error cannot be enlarged in the process of computation.[44] This important property makes the approximation formula to be employed in applications. Figure 2 illustrates that a segment on a contour is smoothed by replacing a noisy pixel $i$ with its fitted value $E(x)$ computed by Cotes-coefficient-based fitting formula in case of $n = 7$.

The NCBS is described as Algorithm 2.1.

---

**Algorithm 2.1** Newton–Cotes-Based Smoothing algorithm NCBS.

1: Resample for the initial contour with equal step length $l_1$;
2: For each point $i$, apply the Cotes-coefficient-based fitting formula for approximation. The new position $(x_i, y_i)$ of point $i$ is computed as $(n = 7)$:

$x_i = \frac{751}{17280}x_{i-4} + \frac{3577}{17280}x_{i-3} + \frac{1323}{17280}x_{i-2} + \frac{2989}{17280}x_{i-1} + \frac{2989}{17280}x_{i+1} + \frac{1323}{17280}x_{i+2} + \frac{3577}{17280}x_{i+3} + \frac{751}{17280}x_{i+4}$,

$y_i = \frac{751}{17280}y_{i-4} + \frac{3577}{17280}y_{i-3} + \frac{1323}{17280}y_{i-2} + \frac{2989}{17280}y_{i-1} + \frac{2989}{17280}y_{i+1} + \frac{1323}{17280}y_{i+2} + \frac{3577}{17280}y_{i+3} + \frac{751}{17280}y_{i+4}$;
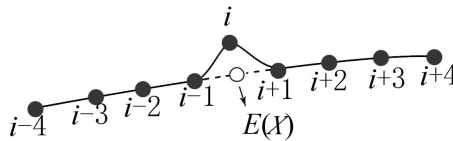
3: Repeat steps 1–2 for $m_1$ times.

---



Fig. 2.   A segment smoothed by replacing a noisy pixel $i$ with its fitted value $E(X)$. The points $i - 4$, $i - 3$, $i - 2$, $i - 1$, $i + 1$, $i + 2$, $i + 3$, $i + 4$ are used to compute the fitted value $E(X)$ of point $i$.

Figure 1(b) is the smoothed result of Fig. 1(a) by this algorithm. The selection for proper values of the key parameters such as $n$, $l_1$ and $m_1$ will be discussed in Sec. 4.

## 2.2. *Curvature-based correction algorithm*

The fact that the bending radius of the juxtapleural nodules are much smaller compared with that of the whole lung parenchyma is the foundation of the CBC. The determination of those pixels with certain curvature values on the contour to be included is the core of this algorithm, which is associated with three problems such as how to represent the curvature of a pixel, how to judge the concavity of a point and how to decide the curvature threshold for correction.

For an arc composed of three successive points $i-1$, $i$ and $i+1$ on the contour, let $R_i$ be the radius of the arc, $K_i$, the curvature of the arc, $L_i$, the Euclidean distance between $i$ and $i-1$ or $i+1$, and $W_i$ the height of the arc, i.e., the Euclidean distance from $i$ to the line connecting $i-1$ and $i+1$ (see Fig. 3). The relationship between $R_i$, $L_i$ and $W_i$ is formulated as below:

$$W_i = \frac{L_i^2}{2R_i} = \frac{L_i^2}{2}K_i. \tag{2.9}$$

In correction algorithm, as $L_i$ corresponds to the sampling step length and has a constant value, the expression (2.9) implies that the height $W_i$ of an arc is inversely proportional to the radius $R_i$ and proportional to the curvature $K_i$. Therefore, $W_i$ can be used as the measurement of $R_i$ and $K_i$. Since an arc is determined by the point $i$ and its two neighbors $i-1$ and $i+1$, the height $W_i$ and radius $R_i$ of an arc are also referred to the height $W_i$ and radius $R_i$ of point $i$.

In order to judge the concavity of a point $i$ on an enclosed contour, we define a reference vector $\mathbf{F}_i$, which originates from point $i$, plumbs to the line connecting $i-1$ and $i+1$, and always directs to the inside of the contour. If the height $W_i$ is extended to a vector $\mathbf{W}_i$ that originates from point $i$, therefore, the point at which the height vector $\mathbf{W}_i$ has opposite direction with reference vector $\mathbf{F}_i$ is concave,
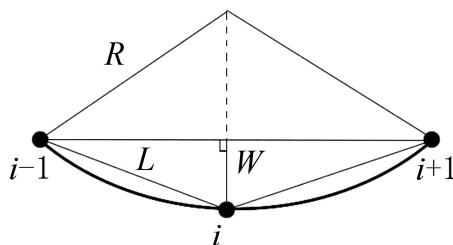


Fig. 3.   An arc composed of three successive points on the contour. $i-1$, $i$ and $i+1$ are three successive points on the arc, $R$ is the radius of the arc, $L$ is the Euclidean distance between $i$ and $i-1$, $W$ is the height of the arc.
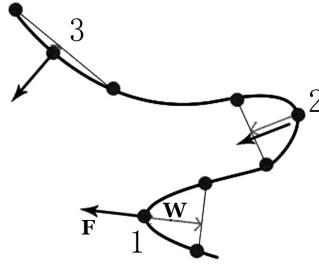
Fig. 4.    Judgment of concavity for points on a segment of an enclosed contour. 1, 2 and 3 are three points on the segment. **F** is the reference vector always pointing to the inside of the contour, **W** is the height vector of the current point. 1 and 3 are concave points as their reference and height vectors have opposite directions, 2 is convex point as its reference and height vectors have same directions.

otherwise is convex. In Fig. 4, the points 1 and 3 are concave and point 2 is convex. The concavity of a point $i$ can be represented by adding a sign to the value of $W_i$. The height $W_i$ which is set positive for a convex point and negative for a concave point is denoted as $W_i^+$. Now three variables can be used to describe the height of a point $i$: $\mathbf{W}_i$, $W_i$ and $W_i^+$. $\mathbf{W}_i$ is the height vector, $W_i$ is the magnitude of $\mathbf{W}_i$ and represents the measurement of the curvature, and $W_i^+$ has the same magnitude with $W_i$ and indicates the concavity by its sign. For the purpose of distinction in our algorithm, $W_i^+$ is referred to height value and $W_i$ is referred to height magnitude or curvature value.

The curvature threshold is the criteria for which points are selected to be corrected. It is noted that juxtapleural nodules locate at the concave parts of the contour and have very small bending radius, so those concave points with very large curvature values should be included in the lung parenchyma. In our correction algorithm, the height is employed as the correction criteria. As mentioned before, the magnitude of the height $W_i^+$ is inversely proportional to the radius value and its sign indicates the concavity of a point, thus those points with very large negative height values will be amended. For a smoothed contour, we can plot the height value $W_i^+$ of each pixel $i$ on a map so as to estimate which parts are the locations of juxtapleural nodules. Figure 5(b) is a plot for the height values of all pixels on a contour in Fig. 5(a) sequentially. In Fig. 5(b), corresponding to part a, b, c and d are convex points with large bending radius (small positive height value), concave points with large bending radius (small negative height value), convex points with small bending radius (large positive height value) and concave points with small bending radius (large negative height value), respectively. It is concluded that the part d on the contour is related to juxtapleural nodules.

For the determination of the curvature threshold value for lung contour amelioration, we should consider the shape features of a lung contour. An entire lung contour [see in Fig. 6(a), denoted as $IC_Q$] is not only smooth but also having an approximated crescent outline [see Fig. 6(b)]. So it can be largely partitioned into
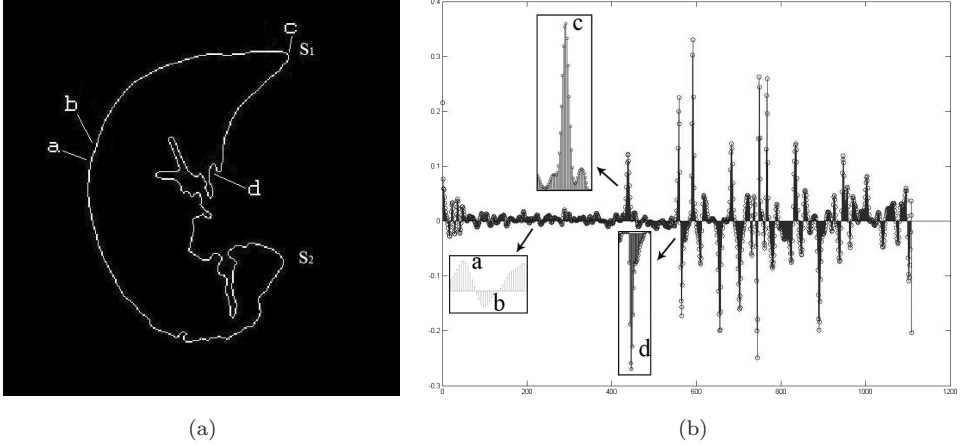
Fig. 5. The plot for height value $W_i^+$ of each pixel on a smoothed contour. (a) The original contour with parts a, b, c and d labeled. (b) The magnified parts a, b, c and d that respectively correspond to the parts a, b, c and d on original contour represent convex points with large bending radius, concave points with large bending radius, convex points with small bending radius and concave points with small bending radius, respectively. Part d is related to juxtapleural nodules.

two parts by two peaks S1 and S2 on the contour, the inner arc and the outer arc. Because of the crescent outline, the curvature magnitudes of most pixels on inner arc are smaller than most pixels on outer arc. Let $Q$ denote the sequence of all pixels on a contour, $Q_1$, the subsequence of all concave pixels, and $Q_2$, the subsequence of all convex pixels. Let $\overline{W}$ be the average of height magnitudes $W_i$ for pixels in a sequence. As most pixels in $Q_1$ locate on inner arc and most pixels in $Q_2$ locate on outer arc, we have [see Fig. 6(c)]

$$\overline{W}(Q_1) < \overline{W}(Q) < \overline{W}(Q_2). \tag{2.10}$$

But for a lung contour to be corrected [see Fig. 7(a), denoted as $IC_P$], it is always distorted far from a crescent outline. Let $P$, $P_1$ and $P_2$ denote the sequences of all pixels, concave pixels and convex pixels on this contour, respectively. Suppose that the contour in Fig. 6(a) is the ameliorated result of the contour in Fig. 7(a). According to the features of juxtapleural nodules analyzed (see Fig. 6), for the contour $IC_P$ only those pixels in $P_1$ with large height magnitudes are corrected. On the other hand, seen from the produced result $IC_Q$, those pixels in $Q_1$ locating near the line $\overline{W}(Q_1)$ [see in Fig. 6(c)] are protected. The selection of a curvature threshold value should maintain these two principles.

Analogous to an entire lung contour $IC_Q$, a distorted lung contour $IC_P$ can be approximated to a crescent with some small holes on its edge, each of which corresponds to a missed juxtapleural nodule. For simplicity these small holes are aggregated to a larger one [see in Fig. 7(b)]. Wherever $IC_P$ has small or large holes on its edge, it always has more pixels with large height magnitudes (these pixels
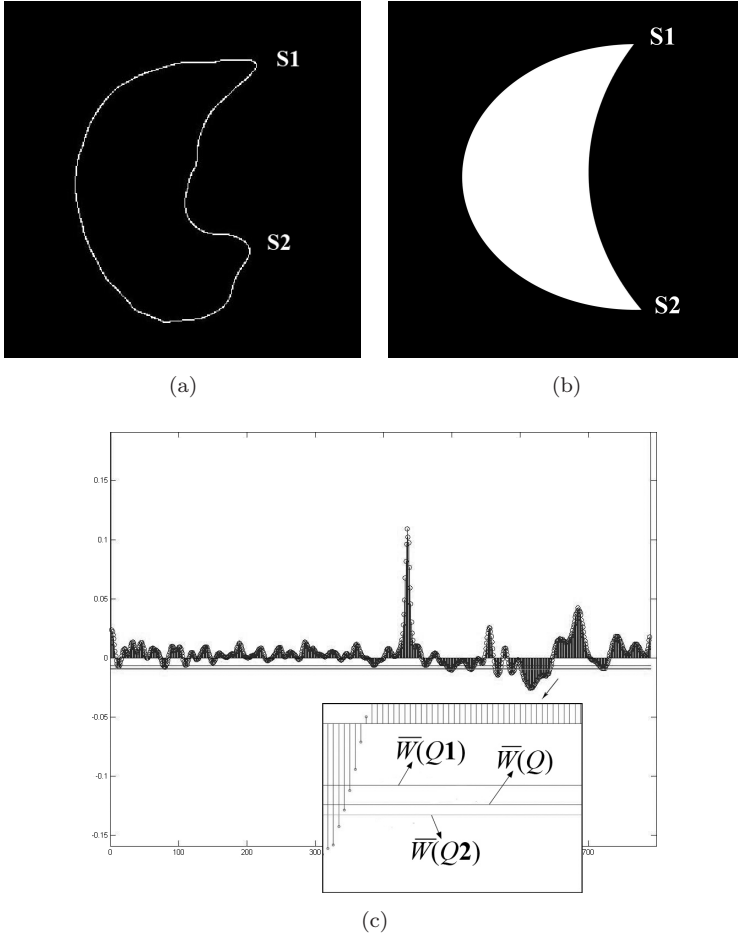
Fig. 6.   An entire lung contour. (a) The entire lung contour is partitioned into two main parts by two peaks S1 and S2. (b) The entire lung contour is approximated to a crescent. (c) The plot for height value $W_i^+$ of each pixel on the entire lung contour and three lines for average height magnitudes $\overline{W}(Q)$, $\overline{W}(Q_1)$ and $\overline{W}(Q_2)$. $\overline{W}(Q)$, $\overline{W}(Q_1)$ and $\overline{W}(Q_2)$ are the average height magnitudes for all pixels, concave pixels and convex pixels on the entire lung contour, respectively. For the convenience of comparison, the three lines are drawn on the lower half part. The magnified segments for the three lines show $\overline{W}(Q_1) < \overline{W}(Q) < \overline{W}(Q_2)$.

are always related to juxtapleural nodules) than $IC_Q$, we get

$$\overline{W}(Q) < \overline{W}(P). \tag{2.11}$$

Combining formula (2.10) and (2.11), we have [see Fig. 7(c)]

$$\overline{W}(Q_1) < \overline{W}(Q) < \overline{W}(P). \tag{2.12}$$

So for the contour $IC_P$ we select $\overline{W}(P)$ as the curvature threshold value for correction, ensuring that only those concave pixels with higher deviations from $\overline{W}(Q_1)$
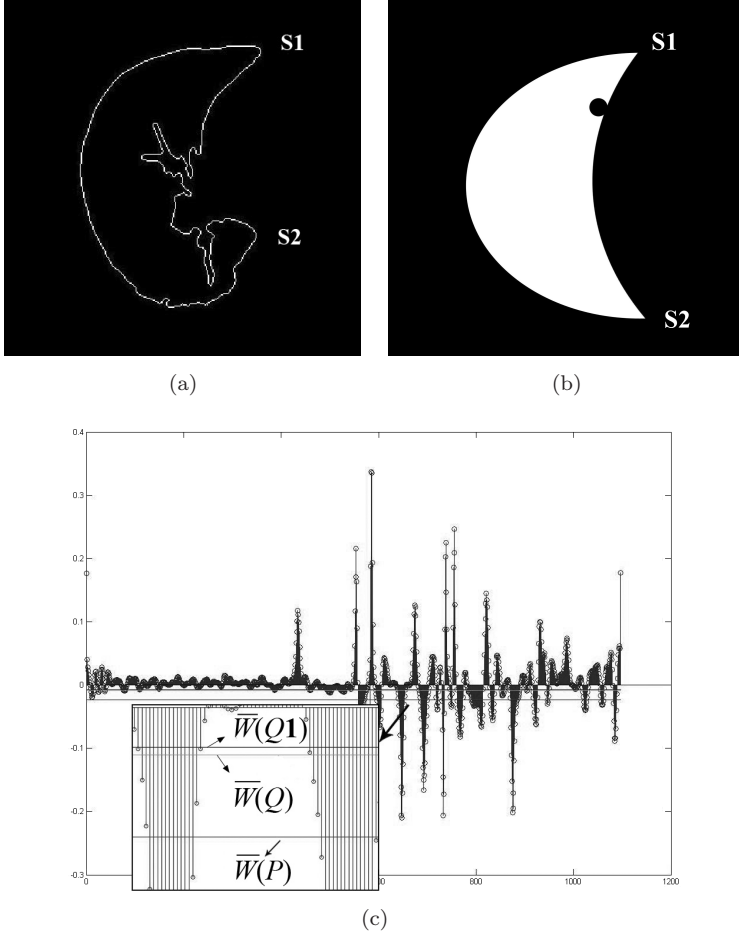
Fig. 7.   A distorted lung contour. (a) The distorted lung contour is partitioned into two main parts by two peaks S1 and S2. (b) The distorted lung contour is approximated to a crescent with a hole on its edge. (c) The plot for height value $W_i^+$ of each pixel on the distorted lung contour and three lines for average height magnitudes $\overline{W}(Q)$, $\overline{W}(Q_1)$ and $\overline{W}(P)$. $\overline{W}(Q)$ and $\overline{W}(Q_1)$ are the average height magnitudes for all pixels and concave pixels on the entire lung contour (the ameliorated result of the distorted contour), respectively. $\overline{W}(P)$ is the average height magnitude for all pixels on the distorted lung contour. For the convenience of comparison, the three lines are drawn together on the lower half part. The magnified segments for the three lines show $\overline{W}(Q_1) < \overline{W}(Q) < \overline{W}(P)$.

will be amended and those near the line $\overline{W}(Q_1)$ will be left unchanged. Note that the curvature threshold value $\overline{W}(P)$ is computed on a smoothed contour, it will not have a much higher deviation value from $\overline{W}(Q)$. This makes sure that most juxtapleural nodules with certain bending radius will be included in the restored lung parenchyma.

After the three associated problems have been resolved, we summarize the CBC as Algorithm 2.2.

---

**Algorithm 2.2** Curvature-based Correction algorithm CBC.

---

1: For a contour $IC_P$, apply NCBS to get a smoothed contour;

2: Resample for the smoothed contour with equal step length $l_2$;

3: For each pixel $i$, compute height magnitude $W_i$. Then Compute the average of the height magnitudes for all the pixels $\overline{W}(P)$. Set curvature threshold value $\theta_{\text{correct}} = \overline{W}(P)$;

4: Resample for the contour $IC_P$ with equal step length $l_2$;

5: For each pixel $i$ in $P$, compute height magnitude $W_i$, height vector $\mathbf{W}_i$, reference vector $\mathbf{F}_i$ and height value $W_i^+$. if $W_i^+ < 0$ and $W_i > \theta_{\text{correct}}$, then it will be amended: Join its two neighbors $i - 1$ and $i + 1$ with a line and remove it from the sequence $P$ (Therefore the contour $IC_P$ is also updated);

6: Repeat steps 4–5 until the length of the contour does not change.

---

In this algorithm, steps 1–3 are used to decide the curvature threshold value for a contour to be corrected, the following steps 4–6 are used for judgment and correction. The variables and their computations in this algorithm can be found in the preceding discussions.

## 3. Lung Segmentation

The CBC algorithm is applied in our designed computer aided lung nodule detection system. In this CAD system, the lung segmentation process is fully automatic, it consists of four modules: Thorax extraction, lung contour detection, lung contour amelioration and lung parenchyma extraction (see Fig. 8).

The process is started with thorax extraction. This stage comprises the removal of all artifacts external to the patient's body, among which are bed sheets, the air that involves him and the surface on which he lies. On gray-level images $G_{CT}$ (in this paper original CT images are transformed into gray-level images with pixel value ranging from 0 to 255) these artifacts are differentiated from the thoracic wall and mediastinum by a threshold $\theta_{HR}$ [see Fig. 9(a)]. This threshold $\theta_{HR}$ can be determined by analyzing the cumulative gray-level histogram of the pixels that
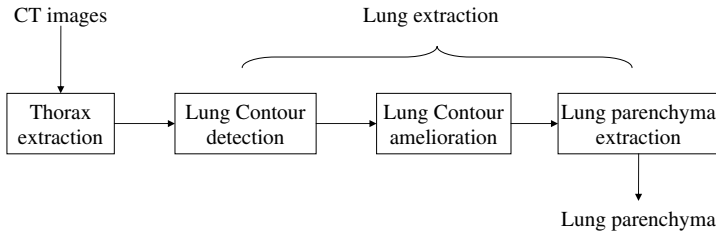


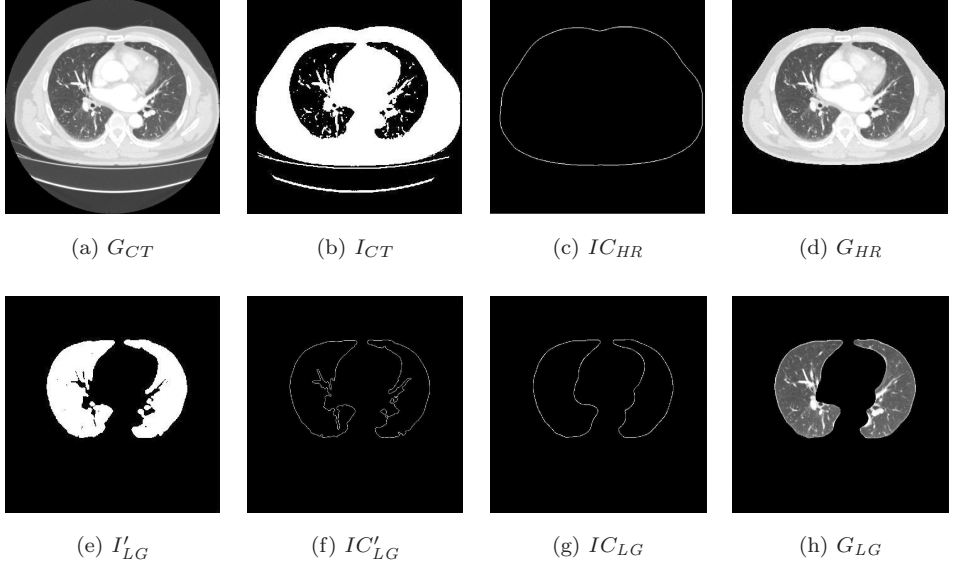Fig. 8.    The automatic lung segmentation process.

(a) $G_{CT}$      (b) $I_{CT}$      (c) $IC_{HR}$      (d) $G_{HR}$

(e) $I'_{LG}$      (f) $IC'_{LG}$      (g) $IC_{LG}$      (h) $G_{LG}$

Fig. 9.   Automatic lung segmentation sequence. (a) Original CT image $G_{CT}$. (b) Transform $G_{CT}$ into a binary image $I_{CT}$ by thresholding. (c) Get the inner border $IC_{HR}$ with boundary tracking algorithm. (d) Compute the thoracic image $G_{HR}$ by region growing method. (e) Acquire the binary lung image $I'_{LG}$ by thresholding to $G_{HR}$. (f) Produce the inner border $IC'_{LG}$ of $I'_{LG}$ with boundary tracking algorithm. (g) Change $IC'_{LG}$ into the final inner border $IC_{LG}$ with CBC. (h) Obtain the final lung parenchyma $G_{LG}$ with region growing method.

constitute the main diagonals of the exam CT slice.[4] By threshold $\theta_{HR}$, the original gray-level image $G_{CT}$ is transformed to a binary image $I_{CT}$, in which the pixel values with higher and lower gray levels in $G_{CT}$ are set 1 and 0, respectively [see Fig. 9(b)]. Next, applying boundary tracking algorithm[45] to binary image $I_{CT}$ we get the inner border $IC_{HR}$ of the thoracic region [see in Fig. 9(c)]. The thoracic image $G_{HR}$ [see Fig. 9(d)] is computed by region growing method.[10]

The purpose of the three successive stages, commonly referred to as lung extraction, is to identify the thoracic wall and mediastinum voxels, making possible the work on the next stages just with the region which forms the pulmonary parenchyma in which juxtapleural nodules with right volumes are included as many as possible.

In the stages of lung extraction, the thresholding method is also used to get the lung region. In order to decide the threshold $\theta_{LG}$, we build a gray-level histogram[4] for the thoracic image $G_{HR}$ obtained previously. The gray-level that maximizes the separation between the two major peaks in this histogram is selected as the threshold. From the thoracic image $G_{HR}$, we acquire the initial binary lung image $I'_{LG}$ [see in Fig. 9(e)] by eliminating those surrounding tissues composed of high-intensity pixels with values greater than $\theta_{LG}$, such as muscles, fats, bones, etc. As in thorax extraction, the boundary tracking algorithm[45] is applied to produce the inner border $IC'_{LG}$ [see Fig. 9(f)] of binary lung image $I'_{LG}$. Since some juxtapleural

nodules are excluded from thoracic region enclosed by $IC'_{LG}$, it will be ameliorated in the next stage.

The amelioration of the inner border is achieved only by the implementation of CBC. With this algorithm, the initial inner border $IC'_{LG}$ is changed into the final inner border $IC_{LG}$ [see Fig. 9(g)].

The last stage is to obtain the correct lung parenchyma $G_{LG}$ [see Fig. 9(h)], which can be done by the region growing method[10] that is used in thorax extraction.

## 4. Results and Discussions

The performances for lung segmentation will be evaluated by three sets of indices adopted in Ref. 36. First is the reinclusion ratio of juxtapleural nodules in the segmented lung parenchyma. Let $n'_{ND}$ and $n_{ND}$ be the numbers of juxtapleural nodules in the segmented lung parenchyma and reference standards, respectively. $n^0_{ND}$ is the number of juxtapleural nodules included in the initial lung contour obtained by thresholding method, then the ratio $ri$ is represented by

$$ri = (n'_{ND} - n^0_{ND})/(n_{ND} - n^0_{ND}). \tag{4.1}$$

Second is the percentages of over- and under-segmentation, indicating the over- and under-segmented volumes for lung parenchyma, respectively. The volumes of lung parenchyma are measured by pixels contained inside. For slice image $k$ of a CT scan, let $v'_k$ and $v_k$ be the volumes of lung parenchyma on the segmented thoracic region and reference standard, respectively, then the percentages of over- and under-segmentation for a slice lung volume, $pct'_k$ and $pct_k$, and for the total lung volume, $pct'$ and $pct$, are respectively represented by

$$pct'_k = (v'_k - v_k)/v_k \times 100\%, \tag{4.2}$$

$$pct_k = (v_k - v'_k)/v_k \times 100\%, \tag{4.3}$$

$$pct' = \left(\sum_k v'_k - \sum_k v_k\right) \Big/ \sum_k v_k \times 100\%, \tag{4.4}$$

$$pct = \left(\sum_k v_k - \sum_k v'_k\right) \Big/ \sum_k v_k \times 100\%. \tag{4.5}$$

The third is the over- and under-segmentation fitness measured by the lung surface distance between the segmented lung parenchyma and reference standard. For slice image $k$ of a CT scan, the distance between the over-segments on the segmented lung parenchyma and the reference standard, $d'_k$, is defined as the maximum of the shortest distances from pixels on the over-segments to the contour of the reference standard. Analogously, the distance between the under-segments on the reference standard and the segmented lung parenchyma, $d_k$, is defined as the maximum of the shortest distances from pixels on the under-segments to the contour of the

*ACB*: over-segments     *ADB*: under-segments



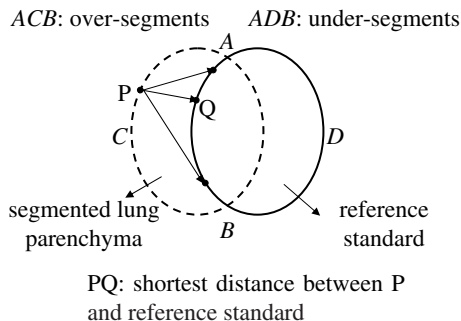PQ: shortest distance between P
and reference standard

Fig. 10. Illustration of the shortest distance from a pixel to a segment. A and B are cross points between over-segments and under-segments. P, C and Q, D are points on over-segments and under-segments, respectively. For a slice image, over-segment is defined as the border of the segmented lung parenchyma not encompassed in the reference standard and under-segment is defined vice versa.

segmented lung parenchyma (see Fig. 10). The total lung surface distances for estimating the fitness of over- and under-segmentation, $d'$ and $d$, are respectively computed as

$$d' = \max_k(d'_k), \tag{4.6}$$

$$d = \max_k(d_k). \tag{4.7}$$

### 4.1. *Experimental results*

The CT scans for 25 outpatients (age 18–79, mean 53; 17 male/8 female) are used in our experiments. Each scan has about 400–800 slices. The interval between slices is 0.4–0.6 mm. Each slice has a matrix size of $512 \times 512$ pixels, the size of each ranging 0.5–0.8 mm. By our developed methods, the lung parenchyma for each slice of CT scans is segmented. In the interest of verification, each scan is examined by three experienced radiologists and the contours of lung parenchyma are delineated on original CT images manually. The consensus about the lung parenchyma identified by three radiologists is used as reference standards. In order to discover whether juxtapleural nodules are included in the segmented thoracic region, the positions and outlines of juxtapleural nodules are also marked on both segmented lung parenchyma and reference standards by radiologists.

By testing against the 25 datasets, our experiments show that all 129 juxtapleural nodules that are missed by initial thresholding method are correctly reincluded and added to the segmentation. The CBC gains a 100% reinclusion ratio for the 25 datasets. An example of the comparison between the original thoracic region, initial lung region obtained by thresholding method and new lung region obtained by CBC is shown in Fig. 11. It is seen that all the 13 excluded juxtapleural nodules from initial lung region are all correctly reincluded in the new lung region.
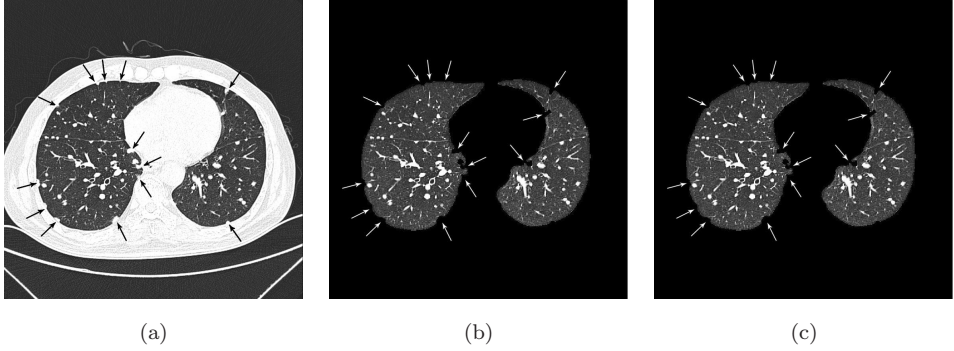
Fig. 11.   Demonstration of excluded juxtapleural nodules reincluded. (a) Original thoracic region with juxtapleural nodules marked by radiologists. (b) The initial lung region obtained by thresholding, from which 13 juxtapleural nodules were excluded. (c) The new lung region obtained by CBC, in which 13 excluded juxtapleural nodules were reincluded.
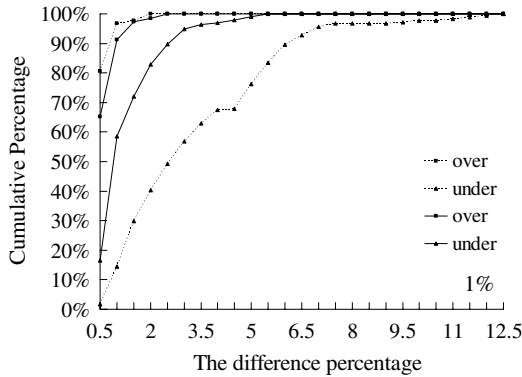


Fig. 12.   The cumulative difference percentage distribution for the slices of a CT scan. The scales (unit is 1%) on $x$-axis represent the over- or under-segmentation percentages ($pct'_k$ or $pct_k$). The scales on $y$-axis represent the cumulative percentages for the slices of one CT scan. A point $(a, b)$ on the two curves of this plane shows that among all the slices, the percentage of the number of the slices which have over- or under-segmentation percentages smaller than $a$ is $b$.

The average over- and under-segmentation percentages for the 25 scans are 0.47% and 1.22%, respectively. The cumulative difference percentage distribution for the slices of one CT scan from the 25 datasets is shown in Fig. 12 (see the two solid lines). It is seen that about 65.3% of the total slices have over-segmentation within 0.5% ($pct'_k \leq 0.5\%$), while only about 16.7% have under-segmentation within the same percentage ($pct_k \leq 0.5\%$). As for all the slices, the over- and under-segmentation are within 2.41% and 5.15%, respectively, i.e., the largest differences for over- and under-segmentation are 2.41% and 5.15%, respectively. In order to see how these two indices change after correction, the cumulative difference percentage distribution for the same slices after initial segmentation is also given in Fig. 12

(see the two dotted lines). It is seen that for the initial segmentation the over- and under-segmentation within 0.5% are 80.66% and 1.66%, respectively. As for all the slices, the over- and under-segmentation are within 1.80% and 12.15%, respectively. In Fig. 12, the two indices changing curves of correction (the two solid lines) locate between the two ones of initial segmentation (the two dotted lines), which means that the correction process is a compromise between over- and under-segmentation. Figure 12 shows that after correction the under-segmentation becomes much better but the over-segmentation becomes a little worse. This approves that the correction algorithm can include the missed juxtapleural nodules but at the same time encompass some unnecessary parts into the lung parenchyma.

The average over- and under-segmentation fitness for the 25 scans is 2.65 pixels and 3.51 pixels, respectively. The cumulative error fitness distribution for the slices of the same CT scan is shown in Fig. 13 (see the two solid line). The largest errors for over- and under-segmentation are 8.82 and 18.66 pixels, respectively. Analogous to Fig. 12, the cumulative error fitness distribution for the same slices after initial segmentation is also given in Fig. 13 (see the two dotted lines). It is also seen that after correction the under-segmentation becomes much better and the over-segmentation becomes a little worse. This further approves the former conclusion that the correction algorithm is always a compromise between over- and under-segmentation.

## 4.2. *Parameter analysis*

In order to gain insight into the proposed algorithms, we conduct some experiments to discover the relationships between different parameters so as to decide their
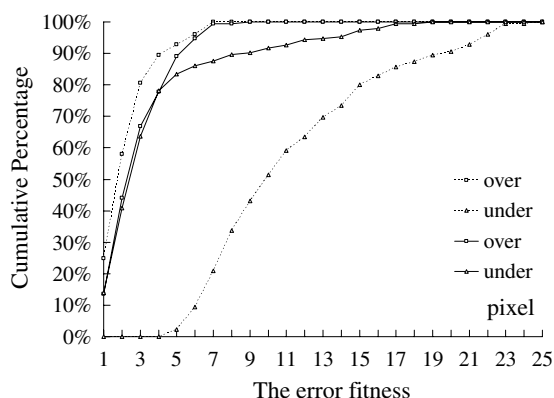


Fig. 13. The cumulative error fitness distribution for the slices of a CT scan. The scales (unit is pixel) on $x$-axis represent the over- or under-segmentation fitness ($d'_k$ or $d_k$). The scales on $y$-axis represent the cumulative percentages for the slices of one CT scan. A point $(a, b)$ on the two curves of this plane shows that among all the slices, the percentage of the number of the slices which have over- or under-segmentation fitness less than $a$ is $b$.

optimal values in real applications. We only focus on the tuning of the involved parameters used in NCBS such as: $n$, the base for the Cotes coefficients; $l_1$, the sampling step length; $m_1$, the epochs for the smoothing algorithm to be executed. Since the parameter $l_2$, the sampling step length for CBC, are set to be 0.9, a value that is often cited by other algorithms.

It is noted that the three parameters $n$, $l_1$ and $m_1$ may be intertwined. But as parameter $n$ represents the size of the neighbors which are used to fit the center point, it has the main effect to the smoothing result[45,46] and should be first determined. For the CBC, the smoothing result can be measured by a predefined value of the threshold $\theta_{\text{correct}}$. It is known that a too small value of $n$, e.g., 1 or 2, will have no evident smoothing result. On the other hand, a too big value of $n$ may have a bad smoothing result.[46] For NCBS, we find that if $n$ is greater than 7 no satisfied result can be obtained despite how the other two parameters $l_1$ and $m_1$ are tuned, so we will study the fitted effects with $n = 3, 4, 5, 6$ and 7, respectively. Figure 14 shows that for the given values of parameters $n$ and $l_1$, the epochs $m_1$ are required at least to get a predefined threshold ($\overline{W}(P) = 0.025$). From Fig. 14, we can see that to get a predefined threshold $\overline{W}(P)$ for a given value of step length $l_1$, the epoch $m_1$ required for an odd value of $n$ (for example $n = 3$) is smaller than an even value (for example $n = 4$), and the larger the base $n$, the smaller the epoch $m_1$. Based on these facts we chose the optimal values for base $n$ to be 7.

Next, we will decide the optimal value for parameter $l_1$. For NCBS the sampling step length $l_1$ is another inherent parameter and has more impact than the exterior parameter $m_1$. Although the smoothed effect due to incorrectly selected value of $l_1$ can be compensated by adjusting the value of $m_1$, it will be seen hereafter that the compensation is very limited. For the given $n$, we investigate the smoothness of contours produced by NCBS for different sampling step length $l_1$. As to curve


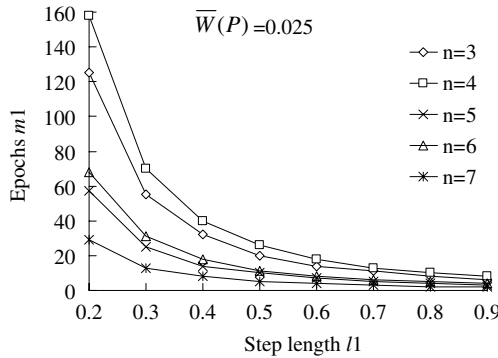
Fig. 14.    Epochs $m_1$ required to get a predefined threshold for different $n$ and $l_1$. $\overline{W}(P) = 0.025$ is the predefined threshold for the NCBS to achieve. The scales on $x$-axis represent the step length $l_1$ adopted by the NCBS. The scales on $y$-axis represent the epochs $m_1$ for the NCBS to be iterated. A point $(a, b)$ on the five curves of this plane shows that to achieve the predefined threshold, the NCBS need to be iterated for $b$ epochs ($m_1 = b$) if $a$ is adopted as the step length ($l_1 = a$).

smoothing, visual inspection is always used to select the optimal values of some parameters, especially under the case that the radian of a curve changes with large magnitudes.[46,47] Figures 15(a)–15(h) are the plots for the height values of consecutive pixels on the contours smoothed by the algorithm with the given $n = 7$ for



(a) $l_1 = 0.2$

(b) $l_1 = 0.3$

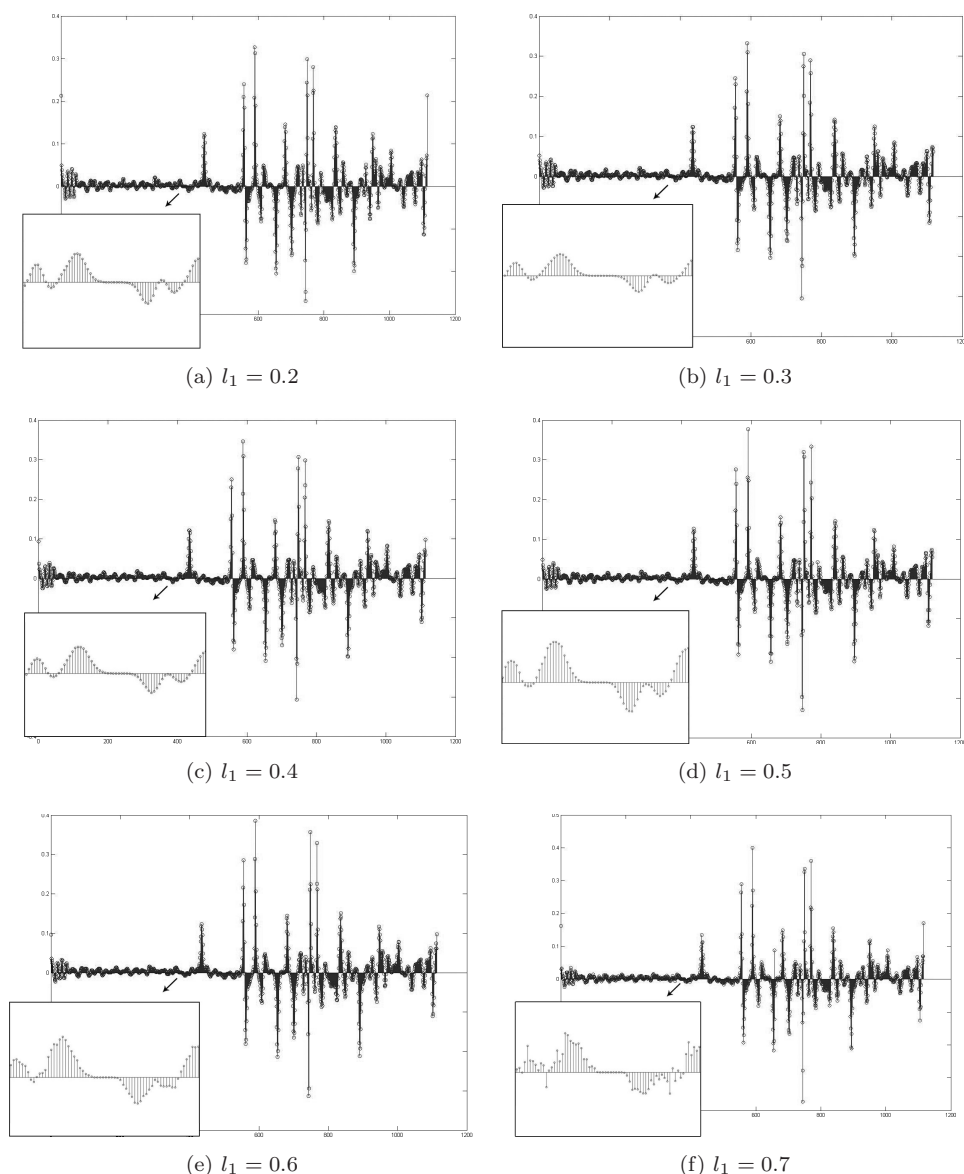(c) $l_1 = 0.4$

(d) $l_1 = 0.5$

(e) $l_1 = 0.6$

(f) $l_1 = 0.7$

Fig. 15. Plots for the height values of consecutive pixels on the smoothed contours by different sampling step length $l_1$. 15(a)–(h) are the plots for contours produced by the NCBS with the given $n = 7$ for $l_1 = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and $0.9$, respectively. The magnified parts on each plot demonstrate the smoothness of the contours.
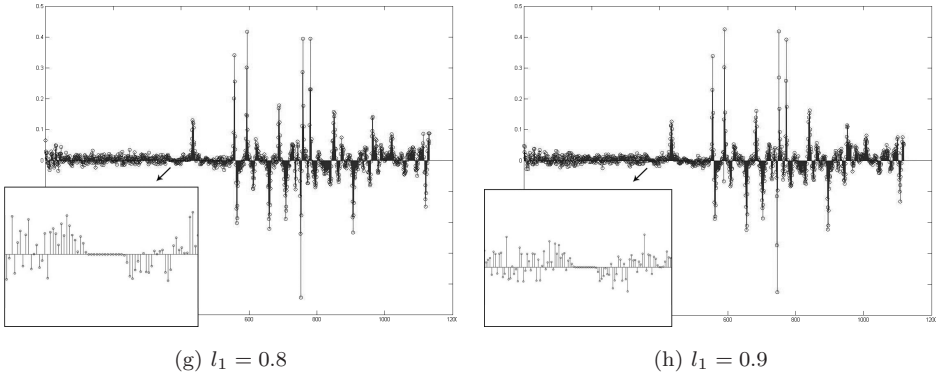
(g) $l_1 = 0.8$                                          (h) $l_1 = 0.9$

Fig. 15.    (*Continued*)

$l_1 = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and $0.9$, respectively. Note that a small segment on each plot is magnified so as to judge its smoothness. From these plots we see that the produced contour with $l_1 = 0.4$ can get best smoothness. Here for CBC, in order to verify the impact of smoothness to the final correction result, we also give the final quantitative error measurements for the same lung contour corrected by CBC under the eight different values of $l_1$ (see Fig. 16). Note that in Fig. 16 the unit for indices $pct'$ and $pct$ is 1% and the unit for indices $d'$ and $d$ is pixel. At the point 0.4 in Fig. 16, the CBC gains a better overall quantitative error measurements than at other points, this verifies that the more smoother the curve is, the better result the correction has. Therefore, for CBC the smoothness judged by visual inspection is consistent with the final correction result.

The last parameter to be tuned is $m_1$. For the given $n$ and $l_1$, some experiments are conducted to find out the impact of parameter $m_1$ on the threshold $\overline{W}(P)$. Figure 17 displays the different thresholds $\overline{W}(P)$ gained by iterating the smoothing
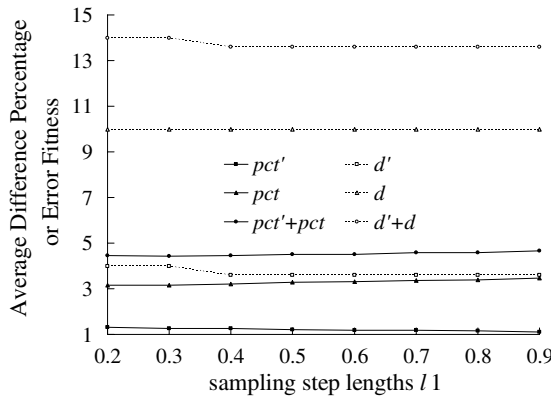


Fig. 16.    The final quantitative error measurements on the same lung contour corresponding to the eight different values of $l_1$.
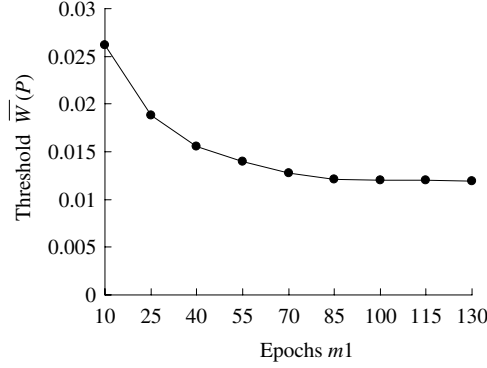
Fig. 17. Thresholds $\overline{W}(P)$ changed with epochs $m_1$.

algorithm with $n = 7$ and $l_1 = 0.4$ for different epochs $m_1$. It is seen from this figure that the thresholds $\overline{W}(P)$ change with slight magnitude.

In order to select the optimal value for epoch $m_1$, we investigate the effects of different epochs $m_1$ on the average performances of CBC to the 25 datasets. Figure 18 shows the set of the performance indices $pct'$, $pct$, $d'$, $d$, $pct' + pct$ and $d' + d$ achieved by the correction algorithm under different epochs $m_1 = 10, 25, 40, 55, 70, 85, 100, 115$ and $130$, respectively. This figure also shows that when the parameter $m_1$ has reached a certain value, most of the indices change with small
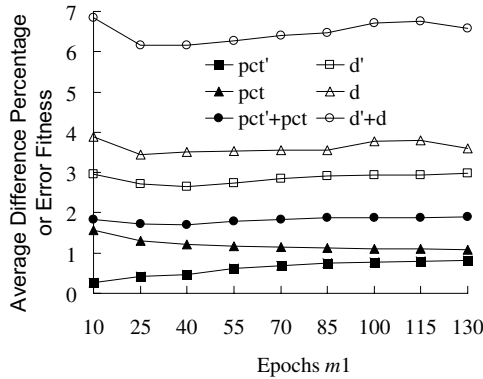


Fig. 18. Average performances obtained by different epochs $m_1$ for the 25 datasets. The scales on $x$-axis represent the epochs $m_1$ for the NCBS to be iterated. The scales on $y$-axis represent the average performance indices $pct'$, $pct$, $d'$, $d$, $pct' + pct$ and $d' + d$ for the 25 datasets. $pct'$ and $pct$ (unit is 1%) are the average over- and under-segmentation percentages. $d'$ and $d$ (unit is pixel) are the average over- and under-segmentation fitness. $pct' + pct$ is the sums of $pct'$ and $pct$. $d' + d$ is the sum of $d'$ and $d$. A point $(a, b)$ on the six curves of this plane shows that if the NCBS repeats a epochs ($m_1 = a$), the average performance index $b$ is achieved by applying the correction algorithm to 25 datasets.

magnitudes. The consistent experimental results shown in Figs. 17 and 18 have sufficiently approved the former conclusion that the compensation by augmenting the value of $m_1$ is very limited. It is seen from Fig. 18 that the correction algorithm obtains similar $pct' + pct$ but larger $d' + d$ with more epochs $m_1$. Considering the effects of less over-segmentation $pct'$ and $d'$ with similar under-segmentation $pct$ and $d$, we take the optimal value for epoch $m_1$ as 40.

### 4.3.  *Comparison with other methods*

At present, some methods including RBA,[4] IDM,[10] GEB[38] ABM,[36] etc. are developed to tackle the lung segmentation problem. But most of them are confronted with the difficulty of selecting correct thresholds for parameters related to the radius of the juxtapleural nodules. Thresholds such as the radius in RBA, the ratio between the geometric and direct distances in IDM, the ratio between adhesive and internal forces in GEB, etc., are not selected straightforward due to the size variance of juxtapleural nodules and have to be decided empirically prior to the implementation of the algorithms. The ABM is the one that the threshold for the marching step length can be tuned adaptively during the process of computation. In ABM method, there are three parameters affecting the result[36]: the initial marching step length *IMS*, adaptive threshold $\lambda_0$, and the scale factor $\delta$. The method will produce an unsatisfied result except that the three parameters are tuned perfectly. First, *IMS* should be assigned a value that is larger than the circumference of any juxtapleural nodule and a smaller value will result in a better fitness between the obtained border and the original border. Next, $\lambda_0$ should be set a value less than 0.5 and should be large enough to make sure that other concave regions will not be incorrectly regarded as juxtapleural region. And the selection of $\delta$ should guarantee that the marching step length will be optimized to a desired value and should help avoid undue computational burden. In our CBC algorithm, the parameter having significant impact on the result is the curvature threshold $\theta_{\mathrm{correct}}$ with an assigned value $\overline{W}(P)$. The only requirement for the computation of $\overline{W}(P)$ is a smoothed contour produced by NCBS. As already analyzed, a smoothed contour-based $\overline{W}(P)$ has little effect on the performances of our correction algorithm. By definition, $\overline{W}(P)$ is a measurement that reflects the outline feature of a contour as a whole. Therefore, the determination of the curvature threshold value in our correction algorithm is not only completely automatic, but also fully self-adaptive.

The computational complexity of RBA is $O(r^2 \times m^2)$, where $r$ is the radius of the ball and $m$ is the width of the lung CT image. The computational complexities of IDM, GEB and ABM are $O(n^2)$, where $n$ is the number of points forming the contour. The computational complexity of our CBC algorithm is $O((m_1 + m_2)n)$, where $n$ is the number of points forming the contour and $m_1$ and $m_2$ are the epochs for the smoothing algorithm (i.e., steps 1 and 2 of NCBS) and correction algorithm (i.e., steps 4 and 5 of CBC) to be executed, respectively. Compared with the number of points on the contour or the width of the lung CT image, the epochs $m_1$ and $m_2$

are much smaller. The execution of our CBC algorithm is more efficient than the other methods.

To compare the correction effects between our CBC algorithm and the other methods, we again conduct the tests by implementing RBA, IDM, GEB and ABM on the same datasets with the same error metrics as presented in Sec. 4.1. For the reinclusion ratio on the 25 datasets, the RBA, IDM, GEB and ABM have achieved values of 100%, 88.32%, 94.47% and 100%, respectively. While our CBC has gained a value of 100%.

The cumulative over- and under-segmentation percentage and fitness distributions for the slices of the same CT scan are shown in Figs. 19(a)–(d), respectively. The over- and under-segmentation percentages and fitness for the same 30 selected slices are given in Figs. 20(a)–(d), respectively. The values for some key error metrics
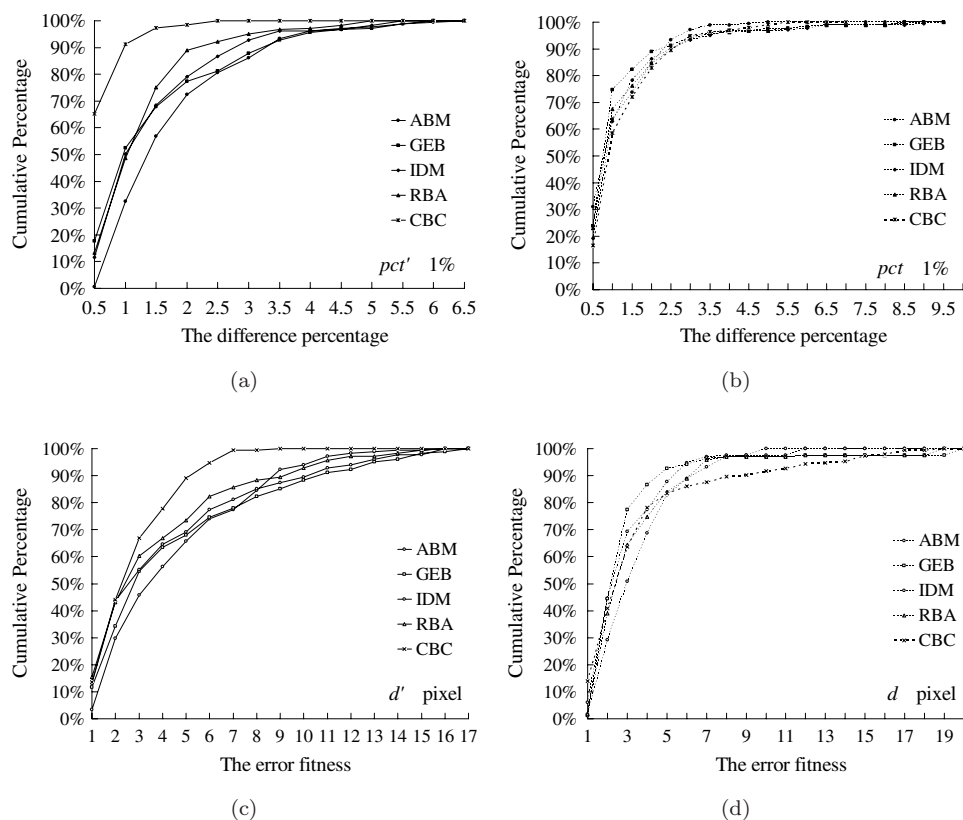


Fig. 19. The cumulative over- and under-segmentation percentage and fitness distributions for the slices of the same CT scan. (a) The cumulative over-segmentation percentage distributions. (b) The cumulative under-segmentation percentage distributions. (c) The cumulative over-segmentation fitness distributions. (d) The cumulative under-segmentation fitness distributions.
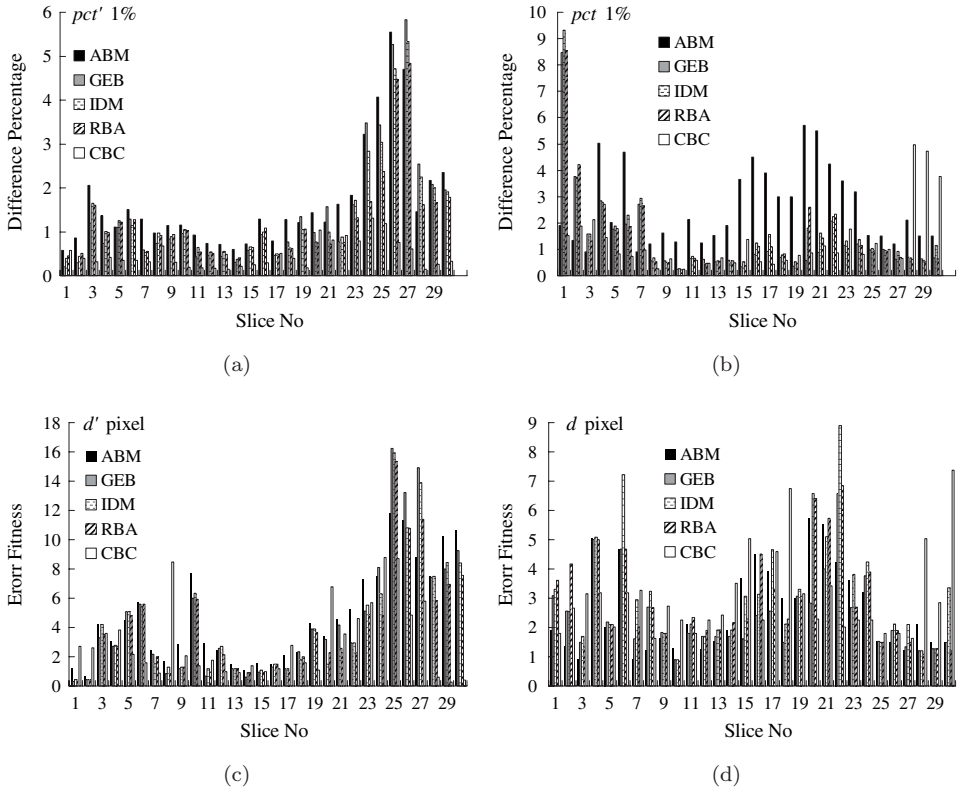
Fig. 20.   The over- and under-segmentation percentages and fitness for the same 30 selected slices. (a) The over-segmentation percentages. (b) The under-segmentation percentages. (c) The over-segmentation fitness. (d) The under-segmentation fitness.

Table 1.   The values for the different error metrics.

| Data sets | Error metrics | ABM | GEB | IDM | RBA | CBC |
|---|---|---|---|---|---|---|
| 25 scans | Re-inclusion ratio | 100% | 94.47% | 88.32% | 100% | 100%$^\star$ |
| | Avg over pct[a] | 1.67% | 1.41% | 1.35% | 1.21% | 0.47%$^\star$ |
| | Avg under pct[b] | 1.13% | 1.12%$^\star$ | 1.37% | 1.26% | 1.22% |
| | Avg over fit[c] | 4.34 | 4.35 | 4.02 | 3.63 | 2.65$^\star$ |
| | Avg under fit[d] | 2.64 | 2.53$^\star$ | 3.27 | 2.97 | 3.51* |
| Slices of | Cul over dist[e] | 0.55% | 17.68% | 11.61% | 13.26% | 65.29%$^\star$ |
| one scan | Cul under dist[f] | 30.99%$^\star$ | 23.70% | 19.23% | 23.09% | 16.58%$^\star$ |
| | Largest over pct[g] | 6.19% | 5.83% | 5.34% | 4.84% | 2.41%$^\star$ |
| | Largest under pct[h] | 4.69%$^\star$ | 8.48% | 9.32% | 8.56% | 5.15% |
| | Largest over fit[i] | 15.18 | 16.77 | 15.93 | 15.36 | 8.82$^\star$ |
| | Largest under fit[j] | 11.17 | 19.63 | 9.55$^\star$ | 18.41 | 18.66 |

*Note*: $^\star$Best value for the different error matrices.
*Worst value for the different error matrices.

are summarized in Table 1. These error metrics include average over- and under-segmentation percentages and fitness for 25 scans (labeled by superscript a, b, c and d, respectively), the cumulative over- and under-segmentation percentage distributions for slices of one scan within 0.5% (labeled by superscript e and f, respectively) and the largest over- and under-segmentation percentages and fitness (labeled by superscript g, h, i and j, respectively).

From Fig. 19, we see that our CBC has much less over-segmentation than the other methods, but it produces a little bigger under-segmentation. This has also been demonstrated by the percentages and fitness for the randomly selected 30 slices of one scan in Fig. 20. In Table 1, the best and worst values for the different error metrics are labeled with a superscript $\star$ and $\ast$, respectively. Table 1 shows that our CBC has achieved best error metrics such as reinclusion ratio (100%), average over-segmentation percentage (0.47%), average over-segmentation fitness (2.65 pixels), cumulative over-segmentation percentage distribution within 0.5% (65.29%), largest over-segmentation percentage (2.41%), largest over-segmentation fitness (8.82 pixels) but also gained two worst error metrics such as average under-segmentation fitness (3.51 pixels) and cumulative under-segmentation percentage distribution within 0.5% (16.58%) and some other unsatisfied error metrics. For over-segmentation, the CBC has the best values for all the error metrics, some are even highly dominant to other methods, e.g., the cumulative over-segmentation percentage distribution within 0.5% (65.29%). While for under-segmentation, the CBC has some worst or unsatisfied values for all the error metrics, but the gaps between these values and the best ones are not big, some of which are very small, e.g., the average under-segmentation percentage (1.22%) is only 0.1% bigger than the best one (1.12%). These experimental results show our CBC has gained significant better over-segmentation by the cost of a little worse under-segmentation. For CT lung contour amelioration, our CBC is a much more rewarded compromise between the over- and under-segmentation than the other methods.

## 5. Summary

As for lung segmentation in computer analysis, we have developed a CBC algorithm to solve the contour correction problem. Unlike other methods, it provides an effective approach to overcome the difficulty for the selection of a threshold. The proposed algorithm focuses particularly on the crescent outline features of the lung and can amend the lung contour with a fully automatic selected curvature threshold. The accuracy has been proved through experiments with real CT scans, the effectiveness and efficiency have been compared with related algorithms and the feasibility has been verified according to the discussion on the determination of optimal parameter values. The algorithm introduced in this paper has been put into use in the early stages of our currently developed lung nodule CAD system and has achieved satisfied results.

## Acknowledgments

## References

1. Cancer facts & figures, American Cancer Society — Cancer statistics. URL Available at http://www.cancer.org, 2012.
2. Tan M, Deklerck R, Jansen B, Bister M, Cornelis J, A novel computer-aided lung nodule detection system for CT images, *Med Phys* **38**:5630–5645, 2011.
3. Armato III S, McLennan G, Bidaut L, McNitt-Gray M, Meyer C, Reeves A, Zhao B, Aberle D, Henschke C, Hoffman E *et al.*, The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans, *Med Phys* **38**(2):915–931, 2011.
4. Armato S, Giger M, Moran C, Blackburn J, Doi K, MacMahon H, Computerized detection of pulmonary nodules on CT scans, *Radiographics* **19**(5):1303–1311, 1999.
5. Ko J, Betke M, Chest CT, Automated nodule detection and assessment of change over time — Preliminary experience, *Radiology* **218**(1):267–273, 2001.
6. Zhao B, Ginsberg M, Lefkowitz R, Jiang L, Cooper C, Schwartz L, Application of the LDM algorithm to identify small lung nodules on low-dose MSCT scans, *Proc SPIE* **5370**:818–823, 2004.
7. Camarlinghi N, Gori I, Retico A, Bellotti R, Bosco P, Cerello P, Gargano G, Lopez Torres E, Megna R, Peccarisi M *et al.*, Combination of computer-aided detection algorithms for automatic lung nodule identification, *Int J Comput Assist Radiol Surg* **7**(3):455–464, 2012.
8. Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T, Automated detection of pulmonary nodules in helical CT images based on an improved template matching technique, *IEEE Trans Med Imaging* **20**(7):595–604, 2001.
9. Way T, Sahiner B, Chan H, Hadjiiski L, Cascade P, Chughtai A, Bogot N, Kazerooni E, Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features, *Med Phys* **36**(7):3086–3098, 2009.
10. Gurcan M, Sahiner B, Petrick N, Chan H, Kazerooni E, Cascade P, Hadjiiski L, Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system, *Med Phys* **29**:2552–2558, 2002.
11. Kubo M, Yamada N, Kawata Y, Niki N, Eguchi K, Ohmatsu H, Kakinuma R, Kaneko M, Kusumoto M *et al.*, CAD system for lung cancer based on low-dose single-slice CT image, *Proc SPIE* **5032**:819–826, 2003.
12. Yamada N, Kubo M, Kawata Y, Niki N, Eguchi K, Omatsu H, Kakinuma R, Kaneko M, Kusumoto M, Nishiyama H *et al.*, ROI extraction of chest CT images using adaptive opening filter, *Proc SPIE* **5032**:869–876, 2003.
13. Mekadaa Y, Kusanagi T, Hayase Y, Mori K, Hasegawa J, Toriwaki J, Mori M, Natori H, Detection of small nodules from 3D chest X-ray CT images based on shape features, *Int Cong Ser* **1256**:971–976, 2003.

14. Enquobahrie AP, Reeves AP, Yankelevitz DF, Henschke C, Automated detection of pulmonary nodules from whole lung helical CT scans: Performance comparison for isolated and attached nodules, *Proc SPIE* **5370**:791–800, 2004.

15. McNitt-Gray M, Hart E, Wyckoff N, Sayre J, Goldin J, Aberle D, A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results, *Med Phys* **26**:880–888, 1999.

16. Li Q, Doi K, Analysis and minimization of overtraining effect in rule based classifiers for computer-aided diagnosis, *Med Phys* **33**:320–328, 2006.

17. Ozekes S, Osman O, Computerized lung nodule detection using 3D feature extraction and learning based algorithms, *J Med Syst* **34**(2):185–194, 2010.

18. Kawata Y, Niki N, Ohamatsu H, Kusumoto M, Kakinuma R, Mori K, Yamada K, Nishiyama H, Eguchi K, Kaneko MN, Pulmonary nodule classification based on CT density distribution using 3D thoracic CT images, *Proc SPIE* **5369**:413–422, 2004.

19. Agam G, Armato III S, Wu C, Vessel tree reconstruction in thoracic CT scans with application to nodule detection, *IEEE Trans Med Imaging* **24**(4):486–499, 2005.

20. Lo SCB, Li YH, Matthew TFMD, Lure YMF, Zhao H, Classification of lung nodules in diagnostic CT: An approach based on 3D vascular features, nodule density distributions and shape features, *Proc SPIE* **5032**:183–189, 2003.

21. Li Q, Li F, Shiraishi J, Katsuragawa S, Sone S, Doi K, Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules, *Med Phys* **30**:2584–2593, 2003.

22. Aoyama M, Li Q, Katsuragawa S, Li F, Sone S, Doi K, Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images, *Med Phys* **30**:387–394, 2003.

23. Armato S, MacMahon H, Automated lung segmentation and computer-aided diagnosis for thoracic CT scans, *Int Cong Ser* **1256**:977–982, 2003.

24. N. Memarian, J. Alirezaie, P. Babyn, Computerized detection of lung nodules with an enhanced false positive reduction scheme, in *Image Processing, 2006 IEEE Int Conf*, pp. 1921–1924, 2006.

25. Suzuki K, A supervised "lesion-enhancement" filter by use of a massive training artificial neural network (MTANN) in computer-aided diagnosis (CAD), *Phys Med Biol* **54**(18):S31–S45, 2009.

26. Matsuki Y, Nakamura K, Watanabe H, Aoki T, Nakata H, Katsuragawa S, Doi K, Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: Evaluation with receiver operating characteristic analysis, *Am J Roentgenol* **178**(3):657–663, 2002.

27. Sluimer I, Schilham A, Prokop M, van Ginneken B, Computer analysis of computed tomography scans of the lung: A survey, *IEEE Trans Med Imaging* **25**(4):856–864, 2006.

28. Wang P, DeNunzio A, Okunieff P, O'Dell W, Lung metastases detection in CT images using 3D template matching, *Med Phys* **34**:258–266, 2007.

29. Pu J, Zheng B, Leader J, Wang X, Gur D, An automated CT based lung nodule detection scheme using geometric analysis of signed distance field, *Med Phys* **35**(8):3453–3461, 2008.

30. Tanino M, Takizawa H, Yamamoto S, Matsumoto T, Tateno Y, Iinuma T, A detection method of ground glass opacities in chest xray CT images using automatic clustering techniques, *Proc SPIE* **5032**:1728–1737, 2003.

31. Takizawa H, Yamamoto S, Recognition of lung nodules from X-ray CT images using 3D Markov random field models, in *Pattern Recognition, 2002. Proc. 16th Int Conf*, Vol. 1, pp. 99–102, 2002.

32. Kakar M, Olsen D, Automatic segmentation and recognition of lungs and lesion from CT scans of thorax, *Comput Med Imaging Graph* **33**(1):72–82, 2009.

33. Fotin S, Reeves A, Biancardi A, Yankelevitz D, Henschke C, A multiscale Laplacian of Gaussian filtering approach to automated pulmonary nodule detection from whole-lung low-dose CT scans, *Proc SPIE* **7260**:72601Q1–72601Q8, 2009.

34. Armato III S, Sensakovic W, Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis1, *Acad Radiol* **11**(9):1011–1021, 2004.

35. Goo J, Tongdee T, Tongdee R, Yeo K, Hildebolt C, Bae K, Volumetric measurement of synthetic lung nodules with multi-detector row CT: Effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy, *Radiology* **235**(3):850–856, 2005.

36. Pu J, Roos J, Yi C, Napel S, Rubin G, Paik D, Adaptive border marching algorithm: Automatic lung segmentation on chest CT images, *Comput Med Imaging Graph* **32**(6):452–462, 2008.

37. Kalender W, Fichte H, Bautz W, Skalej M, Semiautomatic evaluation procedures for quantitative CT of the lung, *J Comput Assist Tomogr* **15**(2):248–255, 1991.

38. Bellotti R, De Carlo F, Gargano G, Tangaro S, Cascio D, Catanzariti E, Cerello P, Cheran S, Delogu P, De Mitri I *et al.*, A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model, *Med Phys* **34**:4901–4910, 2007.

39. Pastor L, Pousse A, Manzoni P, Kastler B, A pulmonary nodule modeling tool as a diagnostic aid for lung HRCT images, *Comput Med Imaging Graph* **29**(8):631–637, 2005.

40. Kim D, Kim J, Noh S, Park J, Pulmonary nodule detection using chest CT images, *Acta Radiol* **44**(3):252–257, 2003.

41. Wiemker R, Rogalla P, Zwartkruis A, Blaffert T, Computer-aided lung nodule detection on high-resolution CT data, *Proc SPIE* **4684**:677–688, 2002.

42. Hu S, Hoffman E, Reinhardt J *et al.*, Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images, *IEEE Trans Med Imaging* **20**(6):490–498, 2001.

43. Chen TJ, Wang YL, Pu XY, The application of fast Fourier transform in the noise smoothing of weak Raman spectrum, *J Yunnan Univ* **27**(6):509–513, 2005.

44. Li QY, Wang NC, *Numerical Analysis*, 4th edn., Tsinghua University Publication, 2006.

45. Sonka M, Hlavac V, Boyle R, *Image Processing, Analysis, and Machine Vision*, 3rd edn., Thomson, 2007.

46. Zhou Y, Wang ZX, Chen ZY, Li DH, An adaptive algorithm of curve smoothing, *Comput Eng* **26**(4):13–16, 2000.

47. Zheng XD, Zhou NN, Huang XH, New method of 2D shape representation and edge smoothing, *Appl Res Comput* **8**:171–175, 2006.