Available online at www.sciencedirect.com
ScienceDirect
Medical Image Analysis 11 (2007) 35–46
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Automatic rib segmentation and labeling in computed tomography scans using a general framework for detection, recognition and segmentation of objects in volumetric data

Joes Staal, Bram van Ginneken *, Max A. Viergever

Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Received 2 March 2006; received in revised form 29 August 2006; accepted 12 October 2006
Available online 27 November 2006

Abstract

A system for automatic segmentation and labeling of the complete rib cage in chest CT scans is presented. The method uses a general framework for automatic detection, recognition and segmentation of objects in three-dimensional medical images. The framework consists of five stages: (1) detection of relevant image structures, (2) construction of image primitives, (3) classification of the primitives, (4) grouping and recognition of classified primitives and (5) full segmentation based on the obtained groups.

For this application, first 1D ridges are extracted in 3D data. Then, primitives in the form of line elements are constructed from the ridge voxels. Next a classifier is trained to classify the primitives in foreground (ribs) and background. In the grouping stage centerlines are formed from the foreground primitives and rib numbers are assigned to the centerlines. In the final segmentation stage, the centerlines act as initialization for a seeded region growing algorithm.

The method is tested on 20 CT-scans. Of the primitives, 97.5% is classified correctly (sensitivity is 96.8%, specificity is 97.8%). After grouping, 98.4% of the ribs are recognized. The final segmentation is qualitatively evaluated and is very accurate for over 80% of all ribs, with slight errors otherwise.

Keywords: Supervised classification; Computed tomography; Detection; Elongated structures; Pattern recognition; Ribs; Computer-aided detection (CAD); Segmentation

1. Introduction

The automatic detection, recognition and segmentation of objects in three-dimensional medical images is an extremely challenging task, for a number of reasons: objects may have no clear boundaries; there may be similar structures in close vicinity of an object; the constellation of objects and their neighborhoods can vary widely from individual to individual; some objects or parts of objects may be missing, and scans obtained in clinical practice often contain pathology.

To overcome these challenges, prior knowledge should be included in the analysis. This may be done by learning characteristics about the objects from examples. Such supervised image analysis methods are becoming increasingly popular in the medical image analysis community. There is a rough dichotomy between local and global approaches. Local approaches, for example voxel classification, can be thought of as bottom-up processes. Due to their nature, they usually do not capture the overall structure and shape of the objects of interest. Global approaches, on the other hand, are usually top-down. They impose a model of a complete object that is fitted to the data. Examples are active appearance models (Cootes et al., 2001) and the m-reps framework (Pizer et al., 2003). However, the constraints set by global models may impede the incorporation of local knowledge to adapt to

* Corresponding author.
E-mail addresses: joes@isi.uu.nl (J. Staal), bram@isi.uu.nl (B. van Ginneken), max@isi.uu.nl (M.A. Viergever).

1361-8415/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2006.10.001

local structure. Furthermore, global object models often require initialization close to their actual position, which requires an initial detection and recognition step. They may also have difficulties in dealing with (partly) missing data; local abnormalities in the data (e.g. pathology or artifacts) may lead to a global failure of the segmentation method.

In many cases, it can be advantageous to detect a constellation of objects. In this way, anatomical knowledge about the expected location of structures relative to other structures can be exploited. Human observers use such strategies, sometimes perhaps without being aware of it. In this process, local and global analysis are intermingled.

In this work a general framework for detection, recognition and segmentation is presented that is a middle ground between local and global algorithms. It starts with a local description and works up to a more global description at every next stage. The frameworks consists of five stages: voxel detection, primitive construction, primitive labeling, primitive grouping and recognition, and finally, segmentation.

This framework is applied to the detection, labeling and segmentation of the complete rib cage in chest CT-scans. The latest generation of CT-scanners produces data with submillimeter resolution and isotropic voxel sizes. Radiologists are at risk of being overwhelmed by the amount of data they have to examine. To overcome this problem, computerized analysis methods are of great practical interest. In Summers (2003), numerous examples are given, of which automated rib segmentation is one. The ribs are always depicted in chest CT, so they should be reported on. In practice, rib anomalies and fractures are frequently missed (Prokop and Galanski, 2003). Hence our interest in a completely automatic extraction of the rib cage. This can be used for effective visualizations of the rib cage and for computerized detection of bone abnormalities. The segmented ribs can also act as reference objects to segment other structures.

Not much attention has been paid to rib cage segmentation. Some methods have been implemented that can be used to segment the rib cage, or, more generally, segment elongated structures in CT-data. In Aylward et al. (1996) and Aylward and Bullit (2002) centerlines are extracted from which the widths of the objects are estimated and rib segmentation is one of the applications considered. In Kim et al. (2002) a tracking algorithm is described, that proceeds from one 2D slice to the next. The method uses seeded region growing, for which the seeds must be supplied manually. The work in Kang et al. (2003) is based on 3D region growing using locally adaptive thresholds. In Kiraly et al. (2006) a method is proposed that finds seed points for ribs from a central coronal section of the scan and uses a centerline tracing algorithm previously developed for segmentation of pulmonary vasculature. From these centerlines a visualization of the rib cage is constructed but a complete rib segmentation is not performed. With the exception of the latter method, these previous works are not completely automatic. Moreover, no previous work used training data for learning the essentials of the detection, labeling and segmentation task.

The remainder of this paper is organized as follows. Section 2 outlines each step in the general framework for detection, recognition and segmentation. The material used for the application to rib cage detection and segmentation is described in Section 3 and the method is described in Section 4. Section 5 presents the experimental results. The paper ends with discussion and conclusions in Section 6.

2. Overview of the general framework

In this section the framework for detection, recognition and segmentation of objects in volumetric data is described in general terms. It consists of five stages:

1. detection of voxels with relevant image structure;
2. construction of primitives from these voxels;
3. classifying which primitives are part of the objects of interest;
4. grouping and recognition of the object primitives;
5. a full segmentation based on the extracted groups.

The implementation details of each stage will depend on the application. In principle each stage can be implemented as a supervised system, trained by examples. For the first two stages, however, it is typically more appropriate to use generic operations to detect blobs (spherical objects, e.g. nodules, tumors), ridge-like elongated structures (vessels, bones), edges (object boundaries), sheet-like patches (sternum, scapula, pelvis, skull) and heuristic grouping algorithms to construct primitives from these voxels. The primitives provide a more natural representation for the problem at hand than raw voxels. Moreover, the complexity of the classification problem, in the third stage, is vastly reduced because the number of primitives is orders of magnitudes smaller than the number of voxels in a scan. For most tasks, it will be difficult to design simple and effective rules for the classification and a supervised approach will be beneficial. A set of features should be computed, and each primitive is represented as a point in this feature space. The problem is now in the realm of statistical pattern theory and different classifiers and feature selection and/or extraction techniques can be explored see e.g. Duda et al. (2001), Hastie et al. (2001) and Fukunaga (1990).

Typically each object of interest will contain several primitives, otherwise the second and third stages of the framework would have solved the recognition problem already. The fourth stage groups the foreground primitives into objects and thus performs recognition. Here the coupling between local and global models takes place: interrelationships between primitives must be established. This stage can be very challenging, especially if the results of previous stages are far from perfect or when there are many primitives and objects to be grouped. If a satisfactory solution cannot be obtained in this stage, the system should reconsider the results of the earlier stages.

Finally, in the fifth stage, the grouped primitives are used to obtain the full segmentation. These groups provide a "skeleton" or an approximate outline of the objects, and can be used as a precise initialization. A wide range of segmentation methods can be used, both heuristic such as region growing (Adams and Bischof, 1994), deformable models (Kass et al., 1987) level sets (Sethian, 1999; Vasilevskiy and Siddiqi, 2002), or supervised, e.g. active shape or appearance models (Cootes et al., 2001, 1995), m-reps (Pizer et al., 2003), deformable organisms (McInerney et al., 2002), etc.

3. Materials

Forty CT-scans of the thorax of different patients were used in this study. The scans were randomly selected from clinical practice and were acquired at the University Medical Center Utrecht on a 16-detector CT scanner (Mx8000IDT, Philips, Best, the Netherlands). All scans were realized in about 12 s in spiral mode with 16×0.75 mm collimation and 15 mm table feed per rotation (pitch = 1.3) in inspiration. Different protocols have been used, but in all scans the slice thickness is 1.0 mm and the slice spacing 0.7 mm. All scans were reconstructed to 512×512 matrices with a moderately soft kernel (Philips B). The in-plane resolution varied between 0.57 mm and 0.91 mm, using the smallest field of view (FOV) to include the outer rib margins at the widest dimension of the thorax. The number of slices varies from 407 to 706. In 36 of the examinations intravenous contrast material was administered.

The high resolution of the scans is not needed for the first four stages of the method. To reduce computation time, the scans were subsampled by a factor 2, resulting in sizes of $256 \times 256 \times 203$ to $256 \times 256 \times 353$ voxels. Listed parameter values for the first four stages pertain to this resolution. The final segmentation step (stage 5), is done on the full resolution scans.

In order to test the classifier with an independent test set, the scans are divided randomly in 20 scans for training and 20 for testing. Parameter values that are listed in the next section have been determined in pilot experiments that used only the training data; the test data was not used in any way during development of the algorithm.

4. Method

This section outlines the method for detecting, labeling and segmenting the complete rib cage from chest CT data, following the general five stage framework.

4.1. Stage 1: 1D ridge detection in 3D images

Because ribs are 1D elongated structures, we detect in the first stage 1D ridge voxels. To detect the ridge voxels of the bone structure in CT-data, the images are first thresholded at a Hounsfield $t_H = 100$ resulting in a binary image, see Fig. 1.

Ridge voxels are detected with the algorithm that is given below. This algorithm is a simplified version of the method presented in Kalitzin et al. (2001). Other schemes for ridge detection are described in Eberly (1996) and Lindeberg (1998).

An elongated structure in a 3D image is an 1D curve. The tangential vectors to this curve are determined by the eigenvector of the Hessian matrix $\mathbf{H}$ with smallest eigenvalue in absolute value. The matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \begin{pmatrix} L_{xx} & L_{xy} & L_{xz} \\ L_{yx} & L_{yy} & L_{yz} \\ L_{zx} & L_{zy} & L_{zz} \end{pmatrix}, \quad (1)$$

where L_{ij}, $i,j \in \{x,y,z\}$ represent second order derivatives of the image data with respect to the coordinates. Because taking derivatives of discrete images is an ill-posed operation, they are taken at a scale σ using the Gaussian scale-space technique (Florack, 1997; Lindeberg, 1994). The main idea is that the image derivatives can be taken by convolving the image with the derivatives of a Gaussian

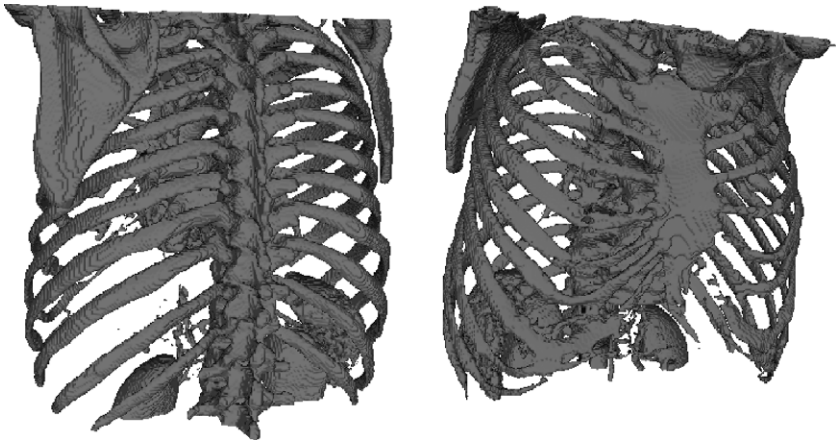

Fig. 1. A thresholded CT-scan surface rendered from two viewing directions. Note that because of the presence of contrast agent not only bony structures but also parts of the heart and great vessels are detected.

$$L_{ij} = L_{ij}(\mathbf{x};\sigma) = \int_{\mathbf{R}^3} L(\mathbf{x}) G_{ij}(\mathbf{x}-\mathbf{x}';\sigma)\,\mathrm{d}\mathbf{x}' \tag{2}$$

with $\mathbf{x} = (x,y,z)^{\mathrm{T}}$ and

$$G(\mathbf{x};\sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \exp\left(-\frac{\mathbf{x}\cdot\mathbf{x}}{2\sigma^2}\right) \tag{3}$$

the Gaussian kernel in 3D. We use $\sigma = 2.0$ voxel.

By ordering the eigenvectors $\mathbf{v}_i$, $i \in \{1,2,3\}$, in decreasing magnitude of their corresponding eigenvalues λ_i, i.e. $|\lambda_1| \geqslant |\lambda_2| \geqslant |\lambda_3|$, the vector tangential to the elongated structure is given by $\mathbf{v}_3$. The other two eigenvectors span the plane perpendicular to $\mathbf{v}_3$. A bright elongated structure has a maximum in this perpendicular plane, see Fig. 2. All eigenvectors are of unit length. As a result, detection of the 1D ridges is reduced to determining at every voxel whether there is a local maximum of the intensity in the plane normal to $\mathbf{v}_3$.

The detection of the maximum in the normal plane is done as follows. Around the voxel $\mathbf{x}$, the circle

$$\mathbf{c}(\theta) = \mathbf{x} + \rho(\mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta), \quad \theta \in [0, 2\pi) \tag{4}$$

is defined, with ρ the radius of the circle. The image has a maximum on a 1D ridge if

$$L(\mathbf{x}) - L(\mathbf{c}(\theta)) > 0 \quad \text{for all } \theta. \tag{5}$$

In the mathematical continuous world, the limit of ρ towards zero would be taken. On the discrete image lattice, a choice of $\rho = 1.0$ voxel seems natural. Furthermore, the polar angle θ must be discretized. We use 8 different angles, corresponding to an eight-connected neighborhood in 2D images. Note that in general $\mathbf{c}(\theta)$ will fall in between grid-points; to evaluate L at those points, trilinear interpolation is used.

4.2. Stage 2: from ridge voxels to line elements

The next step is the construction of primitives in the form of line elements from the set of 1D ridge voxels. A growing process is used that takes into account the local orientation of the ridge. The neighborhood around a ridge voxel is investigated and those ridge voxels in the neighborhood that meet certain requirements are added to the primitive. The first condition is that the candidate ridge voxel at $\mathbf{x}_c$ is in the neighborhood of the seed voxel $\mathbf{x}_s$

$$\|\mathbf{x}_c - \mathbf{x}_s\| \leqslant \epsilon_c, \tag{6}$$

where $\epsilon_c \in [0,\infty)$ is the connectivity radius. The size of the radius determines the size of small gaps that can be closed. A second test investigates the similarity of the local orientation with an inner product

$$|\mathbf{v}_{3,c} \cdot \mathbf{v}_{3,s}| \geqslant \epsilon_0, \tag{7}$$

where $\epsilon_0 \in [0,1]$ is a threshold that determines the orientation sensitivity. A higher threshold means stricter selectivity. Finally, it could happen that a candidate ridge voxel obeys the connectivity and orientation demands, but that it is located on a parallel ridge, see Fig. 3. To prevent this, a third test is applied. The orientation of the unit-length vector $\mathbf{r}$ that points from $\mathbf{x}_c$ to $\mathbf{x}_s$ is compared to the local orientation vector at the seed

$$|\mathbf{r} \cdot \mathbf{v}_{3,s}| \geqslant \epsilon_p. \tag{8}$$

The threshold $\epsilon_p \in [0,1]$ sets the parallell sensitivity. Again, a higher threshold means stricter selectivity. The primitives that are found with these constraints are coined *affine convex sets* or convex sets for short. Convex, because they approximate straight lines (the only 1D convex sets in 3D) and affine because Euclidean distance is replaced with geodesic distance (i.e. distance along the set) (Staal et al., 2004).

The seeds for this region growing process are chosen randomly from the available ridge voxels. After extraction of a convex set a new seed is chosen randomly from the remaining ridge voxels. To prevent growing from one

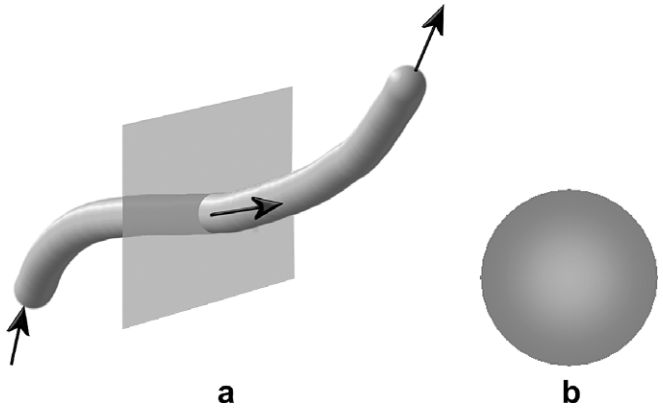

Fig. 2. (a) A bright elongated structure in 3D. Along the planes perpendicular to its centerline (sub-dimensional) maxima are found. (b) Image intensities in a plane perpendicular to the centerline.

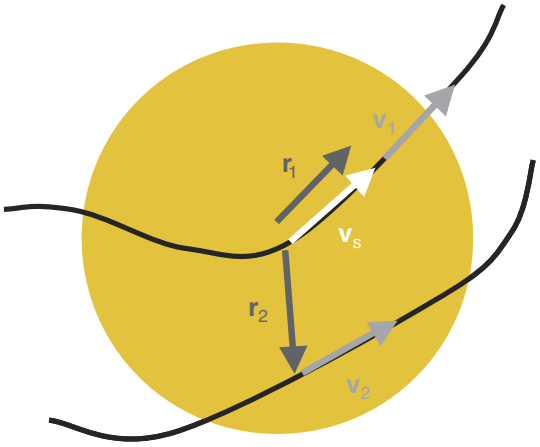

Fig. 3. The dark curved lines are two ridges. The diameter of the disk is ϵ_c. $\mathbf{v}_s$ is the eigenvector belonging to the seed voxel, $\mathbf{v}_1$ and $\mathbf{v}_2$ are the eigenvectors of candidate voxels. The vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are unit vectors pointing from the seed voxel to the candidate voxels. The voxel that belongs to the same ridge will be added, because it satisfies the conditions in Eqs. (6)–(8). The voxel on the parallel ridge does not satisfy condition Eq. (8) and will not be added.

elongated structure in to another, a threshold ϵ_m can be set on the maximum size a convex set may attain. To extract the rib primitives, the following settings were used: $\epsilon_c = 5.0$, $\epsilon_0 = 0.9$, $\epsilon_p = 0.8$ and $\epsilon_m = 20$.

4.3. Stage 3: classification of the primitives

After the formation of primitives, some of them will belong to the objects we want to segment (ribs) and some will not (background). We use a classifier to distinguish ribs from other primitives. A classifier maps a feature vector extracted from the objects under investigation to class numbers or to a vector of probabilities, where each element in the vector is the probability of belonging to that class. The mapping must be learned from example data, where the examples have been labeled to the appropriate class numbers.

Two types of features have been used in this work: local features that encode information from a single primitive and features that encode a relationship between two primitives. Were only the first type of features available, any classifier from pattern recognition theory could have been applied, e.g. linear and quadratic discriminant classifiers, neural networks, kNN-classifiers (Duda et al., 2001; Bishop, 1995). But as we want to include features of the second type as well, there is no straightforward method to use these classical classifiers. Therefore, we employed a spin-glass classifier that was presented in Staal et al. (2005). In this classifier every primitive is considered to possess a spin that can be up or down, signifying foreground (rib) or background. The system is subject to an energy functional consisting of a local and a bilocal part, which encode the features from a single primitive and the interaction features, respectively. A Metropolis algorithm is applied to find the expected values of the state variables of the system, that is, the probabilities for the labels of the primitives. For a detailed description of this classifier we refer to Staal et al. (2005).

For the local features geometrical information of the primitives and intensity based information is extracted. The geometrical information used is the Euclidean length of the primitive, its curvature and the projection of the main axis of the primitive to the x, y and z axis (5 features). For the intensity based features combinations of Gaussian derivatives are computed on 4 scales ($\sigma = 1.0$, 2.0, 4.0 and 8.0 voxels). The values of the derivatives are averaged over the primitive, i.e. a feature $\phi_i = \frac{1}{N}\sum_j D_j$, where j runs over the locations of the primitive, N is the number of voxels in the primitive and D_j is the derivative (combination) of interest. All zeroth, first and second order partial derivatives are computed ($1 + 3 + 6$ features times 4 scales), resulting in 40 features. The gradient magnitude, the determinant of $\mathbf{H}$ and the trace of $\mathbf{H}$ are also included, adding another 12 features (3 features times 4 scales). So, in total 57 local features are extracted.

For the features that encode interactions between two primitives ξ_i and ξ_j we compute a measure for distance (distance between the closest end-points), a measure for mutual orientation (inner product between the unit vectors aligned with the primitives) and an alignment measure, see Fig. 4. Furthermore we use the mean and the absolute difference of the curvatures, the mean and absolute difference of the gray values, the mean and absolute difference of the lengths, the absolute difference of the projection of the primitives' main axis to the x, y and z coordinate and the absolute difference of the average value of the x, y and z coordinate of the primitives. As a result, 15 interaction features are extracted. We consider only interactions between neighboring primitives if either of their endpoints is within a distance of 10 voxels of each other.

Beforehand, it is unknown which features produce good results and which features deteriorate the classification process. A means to determine the appropriate features is to conduct feature selection (Duda et al., 2001). The feature selection method that has been applied is sequential forward selection (Whitney, 1971). This algorithm starts with a null feature set and, for each step, the best feature according to some criterion function is included with the current feature set. In this work, A_z, the area under the ROC-curve, is taken as the criterion function (Metz, 1978). If a feature is included, the performance of the current set of features is stored. After all features have been included, the subset that gives the best overall performance is chosen. For the feature selection process the training set was randomly split in 10 scans for training and 10 scans for evaluation (the test set was not involved in the feature selection process in any way).

After training and feature selection, new data can be classified. However, since the output of the classifier is a probability, we must set a threshold t_P to get a classification in foreground and background. This threshold can be estimated from the training set, by taking that value

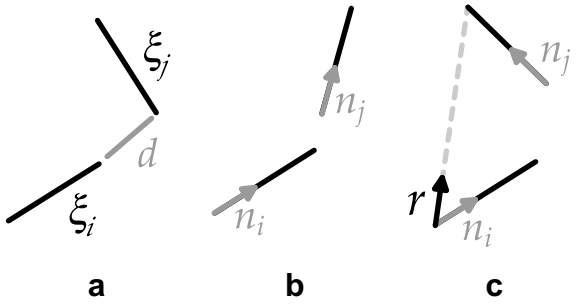

Fig. 4. (a) The shortest distance between the end-points of two line elements is taken as the distance d between the elements. Note that there are four distances between the end-points of two elements. (b) The angle between two line elements is characterized by the absolute value of the inner product of the unit vectors $\mathbf{n}_i$ and $\mathbf{n}_j$, that are aligned with the line elements. (c) A (symmetric) measure for alignment is found by looking for the end-points which are closest to each other and forming a vector $\mathbf{r}$ of unit length along the line between the two other end-points. Note that those end-points are not necessarily the end-points with the longest distance between the two line elements. The alignment measure is now defined as the mean of the absolute values of the inner product of $\mathbf{r}$ with $\mathbf{n}_i$ and $\mathbf{n}_j$: $\frac{1}{2}(|\,\mathbf{r}\cdot\mathbf{n}_i\,| + |\,\mathbf{r}\cdot\mathbf{n}_j\,|)$.

$t_{P,\ max}$ which maximizes the accuracy of the system. The accuracy measures the rate of correct predictions made by the classifier over the complete set.

4.4. Stage 4: grouping of primitives

The goal of this stage is twofold: to group the rib primitives into centerlines of the specific ribs and to label the centerlines to rib side and number (e.g. 10th left rib).

For the centerline construction, the mechanism presented in Section 4.2 is used again, but this time on the ridge voxels that constitute the primitives classified as foreground only. Due to the absence of the background primitives, the constraints on orientation and parallelism can be relaxed to $\epsilon_0 = 0.5$ and $\epsilon_p = 0.7$ and the maximum size requirement can be removed. To encourage closing small gaps, $\epsilon_c = 11$ voxels is taken.

For the labeling of the centerlines two simple heuristic algorithms prove to be sufficient. First, the centerlines are divided in left and right parts by computing the center of gravity of all centerlines. Comparing the center of gravity of a specific centerline with the total center of gravity decides to which side the centerline belongs.

Then, to obtain the rib numbers for one side, the longest centerline from that side is picked. Next, the centerlines that are closest above and below are detected. The centerline found above is used to find the next centerline above. In a similar fashion, the next centerline below is detected. This scheme is repeated until no more centerlines can be detected. To determine which centerline is closest, we define a distance between two centerlines as follows. Every ridge voxel in the first centerline is mapped to the closest (in Euclidean distance) ridge voxel in the second centerline and vice versa. Note that multiple-to-one mapping is possible. Now, the distance between two centerlines is taken as the average of the distances between the mapped ridge voxels. At the start of the number assigning, it has to be decided which of the two closest centerlines is above and which is below the longest centerline. The average z value of the constituting ridge voxels is used for that purpose. The algorithm stops when 11 ribs are found. Then, the final rib, which is not present in all scans, is searched for. A final rib is only included if it is not too small (at least 80% of the closest centerline) and if its distance to the closest rib is between 0.9 and 1.1 of the mean distance of the already recognized centerlines.

4.5. Stage 5: full segmentation

After the classification stage the centerlines of the objects that must be segmented have been estimated. For the full segmentation, we have chosen to use a seeded region growing algorithm as described in Adams and Bischof (1994). At the start of the algorithm, the mean gray values of the image per centerline are computed. A list is initialized with neighboring voxels of all centerlines. The voxel in this list that has the smallest difference in gray value as compared to the mean of its neighboring centerline is removed from the list and added to the centerline. Next, the mean of the centerline is updated and the neighbors of the added voxel, if not already present, are added to the list. The algorithm continues until the list is empty.

In this seeded region growing algorithm, only voxels in a neighborhood of 40 voxels around every centerline are taken into account and voxels with a Hounsfield unit below 130 are considered to belong to background. To circumvent leaking in non-bone structures, only voxels above a gray level threshold t_b and closer than a distance d_{max} to the centerline are added to the boundary list. To prevent leaking into structures that are not foreground, the region growing is performed on both the centerlines and the non-grouped background primitives.

5. Experiments and results

This section presents the results of the experiments for each processing stage. The output of stage 1, the ridge detection, is shown in Figs. 5(a) and (b) for two of the data

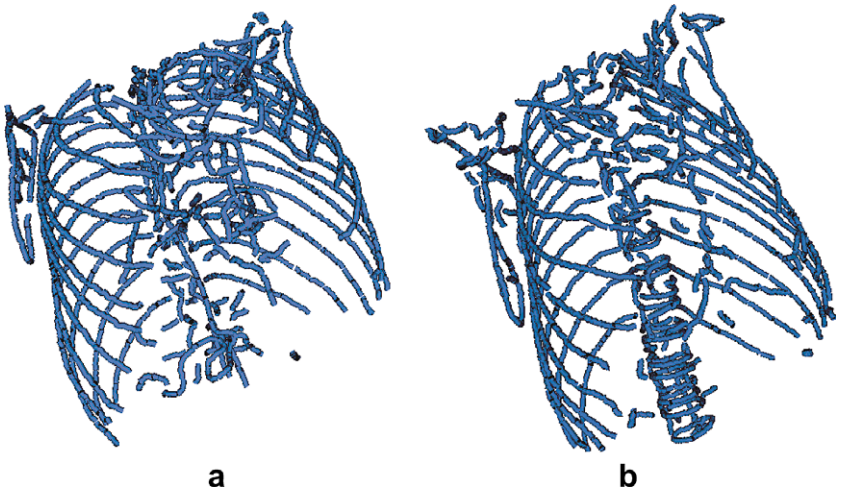

Fig. 5. Ridges of two chest CT-scans. (a) Input image size is $256 \times 256 \times 235$. (b) Input image size is $256 \times 256 \times 288$.

sets. In Fig. 6, primitives obtained in the second stage for the datasets in Fig. 5 are shown.

To enable manual labeling a 3D interface was developed that enabled an observer (the first author) to select primitives by clicking. The observer had the opportunity to view slices of the original CT-scan together with the primitives and it was possible to perform rotations, zooming and window leveling. Labeling of one scan took about 5 min. The observer added information to which rib a primitive belonged (e.g., left side, second rib). Because the first left and right ribs were not always visible in the scans, these were not marked in the training and test data. We did not encounter any cases where a rib existed in the data, but no primitive was found by the ridge detection.

In the training set 3290 primitives out of 9455 primitives were labeled as rib (35%), in the test set this number was 3208 out of 10,016 (32%). After the feature selection, 17 local features and 7 interaction features were retained. Table 3 shows which features were selected in what order and which value of A_z was achieved by adding each feature. The optimal threshold on the posterior probabilities $t_{P,\,max}$ was found to be 0.85. Running the classifier on the test set resulted in the ROC-curve of Fig. 7. The value for A_z was 0.992. The corresponding confusion matrix after thresholding the posterior probabilities with $t_{P,\,max}$ is shown in Table 1. From Table 1 we find that the accuracy is 0.975, the sensitivity 0.968 and the specificity 0.978.

An example of the classification results on the primitives of Fig. 6 is shown in Fig. 8. Note that in Fig. 8(b) some primitives of the fifth right rib are misclassified (the upper rib is the second rib of the subject).

Next, the classified primitives were grouped together. Fig. 9 depicts the results. In 3 scans only 20 of the 22 selected ribs were found. In two of these images the left and right 12th ribs are missing, in 1 scan a second and a 12th rib were missing. In one scan 21 ribs were found (twelfth rib missing). There were two scans in which 23 ribs were detected. In both cases a first rib was detected.

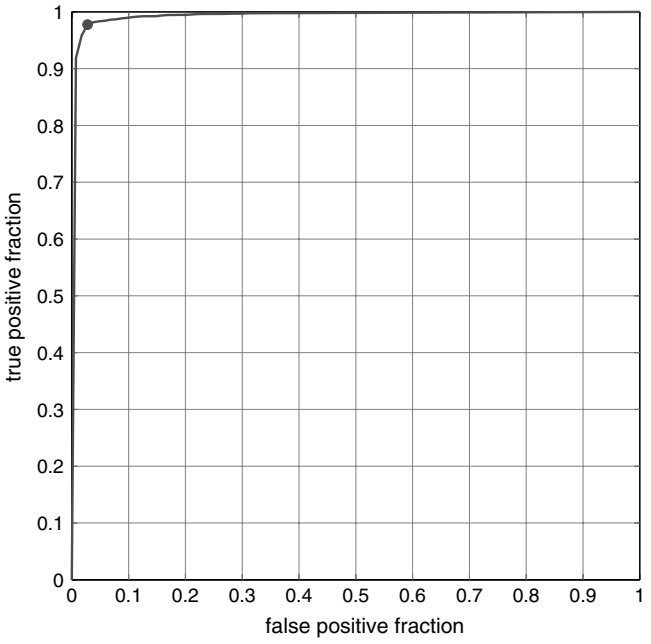

Fig. 7. The ROC-curve of the system, computed over the test set. The area under the curve is 0.992. The circle denotes the point of optimal accuracy (determined from experiments on the training set) with $t_{P,\,max} = 0.85$.

Table 1
Confusion matrix for the primitives in the test set consisting of 20 scans for $t_{P,\,max} = 0.85$

		Reference		Total
		Rib	Non-rib	
Computer	Rib	3105 (31.0%)	147 (1.5%)	3252 (32.5%)
	Non-rib	103 (1.0%)	6661 (66.5%)	6764 (67.5%)
	Total	3208 (32.0%)	6808 (68.0%)	10016

As the first ribs were not always visible, we had decided not to mark them as rib when setting the truth in both training and test scans. In these two test scans they were

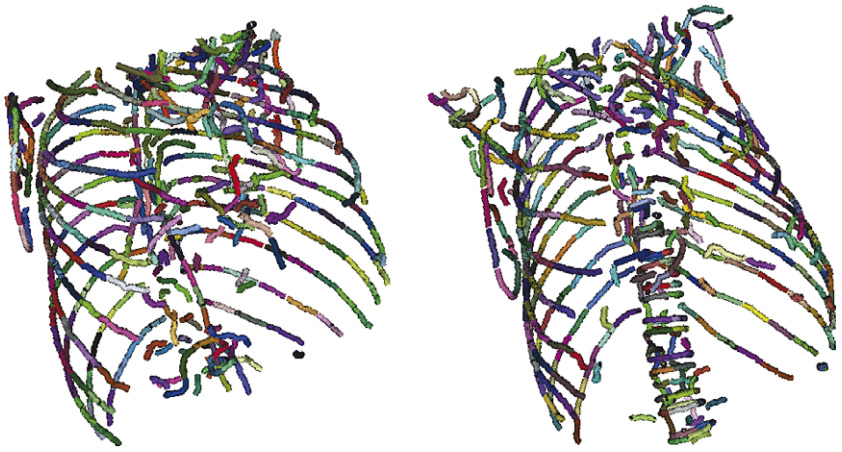

Fig. 6. The primitives of Fig. 5(a) and (b). Every set has a distinct random color. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

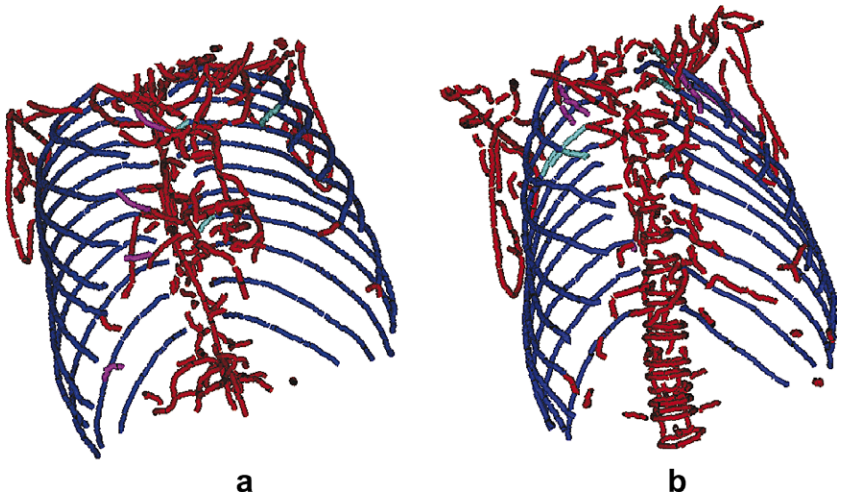

Fig. 8. Results of the classification of the primitives of Fig. 6. True positives are shown in dark blue, true negatives in red, false positives in magenta and false negatives in cyan. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

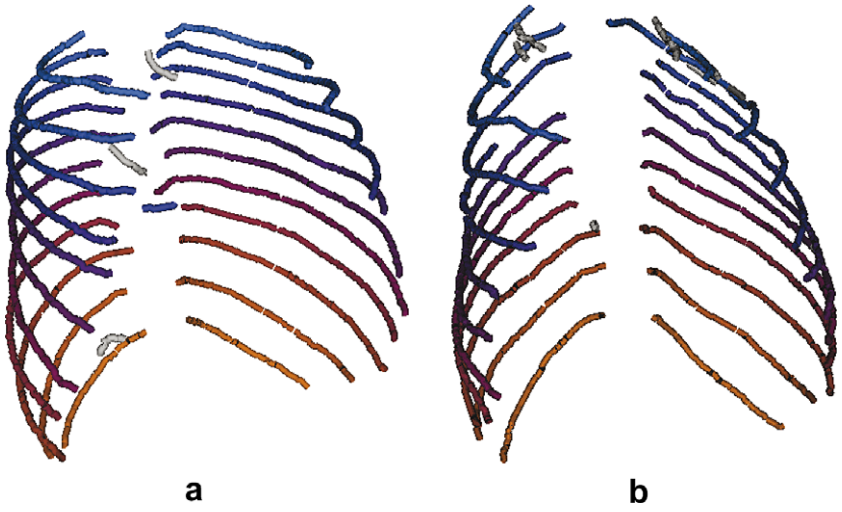

Fig. 9. Result of grouping the foreground primitives of Fig. 8. Centerlines not considered ribs are in light gray. The numbering of the ribs is color coded: from the upper centerline in blue via purple to the lower centerline in orange. Some of the primitives of the fifth right rib (on the left of the image, number four from the top) in Fig. 8(b) have been misclassified, in (b) these parts are missing, and as a result the structure representing that rib is shorter than the other ones. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

nevertheless detected by the system. All other ribs in all scans were correctly found and labeled. The detection results are further specified in Table 2. Taking only rib 2 up to and including rib 12 into account, $433/440 \times 100\% = 98.4\%$ of the ribs was correctly recognized. Notice from Fig. 9 that some false positives have been detected, because these primitives could not be assigned to a group.

In Fig. 10, the results of the full segmentation are shown for the data sets of Fig. 9, which served as initialization for the region growing algorithm. The 6th and 7th rib on the left side in Fig. 10(a) are dented, an observation that is not easily made when inspecting the scan slice by slice. A rendering from a different perspective highlighting this detail is shown in Fig. 11. Note that due to the missing primitives of the 5th right rib in Fig. 9(b), also the full segmentation in Fig. 10(b) is incomplete.

Since we have no manual reference for the full segmentation, a qualitative evaluation has been conducted. An observer (the first author) inspected surface renderings, similar to the ones in Fig. 9, to check whether the ribs had grown into the bony structures of the spine. Next, the observer inspected the scans slice by slice with the segmentation overlaid on it. The observer counted the ribs that are too long (growing into the cartilage) and too short (typically because a primitive has been misclassified, see e.g. Figs. 9(b) and 10(b)). The results of this evaluation are shown in Table 4. The majority of the ribs has the correct length. Two scans are responsible for 16 ribs which are too short. Careful inspection of these scans revealed that

Table 2
Results of rib detection divided in left and right ribs

	Rib 1	Rib 2	Ribs 3–11	Rib 12
Left	2	19	20	17
Right	0	20	20	17

Every entry gives the number of ribs found for the 20 test scans. For rib 1 no labels were set in the test set. For rib 2 up to and including rib 12, 20 detections is the maximum possible.

the threshold set in the first stage ($t_H = 100$) is too high. Also, in one scan eight ribs are found which are too long (Fig. 12). The cartilage of this subject is heavily calcified, so that it appears to be bone. Finally, because the ribs are connected to the spine, the seeded region growing algorithm tends to oversegment that part of the ribs (see Fig. 13).

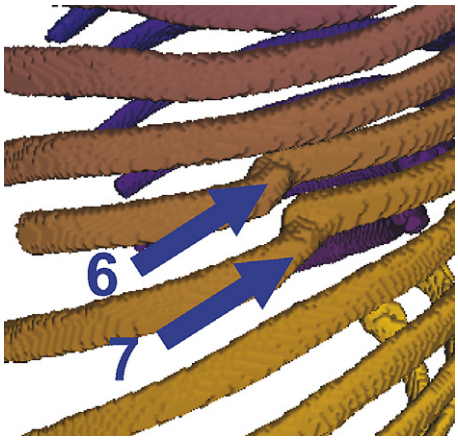

Fig. 11. Detail of the dented ribs 6 and 7 on the left side of Fig. 10(a). Surface rendering is taken from a different perspective.

6. Discussion

A five stage method has been presented for automatic detection, recognition and segmentation of elongated structures in 3D data. The approach can easily be adapted to other types of structures, such as ridge or edge surfaces, for which primitives in the form of patches can be constructed. The method can also be used in other data dimensionalities.

Local segmentation methods like voxel classification are not the most suitable approaches when large data sets have to be processed. There are a few reasons. First of all, obtaining (manually) labeled reference sets, which are needed for training, is an extremely cumbersome process. Second, when a large amount of features is needed or a complex classifier is employed, the time required for computing the features, training the classifier and segmenting the voxels can be prohibitively large. And finally, local methods yield no recognition of the segmented objects.

Global methods, like active shape models and active appearance models, return a segmentation in which recognition is granted, that is, they tell the user exactly which object is where. However, these methods rely on a careful initialization. Using these models e.g. for rib cage segmentation can easily lead to results where the ribs in the detected rib cage are shifted with respect to the ground truth. This is due to the fact that the model will locally be well fitted to the data, and therefore cannot escape from such a local minimum. Another important issue with these schemes is that example data is needed with reliably annotated corresponding landmarks, which is difficult in 3D data sets. This hampers the applicability of these models, especially because they need a great many example sets to capture the anatomical variability that is inherently present in 3D medical data.

In our proposed approach, we try to surmount some of the limitations of the purely local and purely global approaches discussed above. The construction of primitives

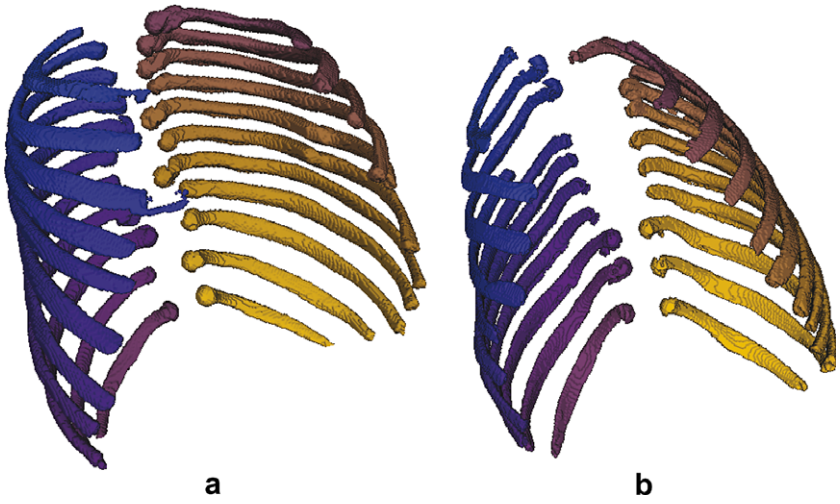

Fig. 10. Full segmentation of the data sets of Fig. 9. (a) The 6th and 7th rib on the left side are dented, an observation that is not easily made when inspecting the scan slice by slice (see Fig. 11 for a more detailed view). (b) The fifth right rib on the right is not complete, because the primitive that should have initialized the region growing algorithm is not classified as rib.

Table 3
Results of the feature selection

Number	Feature	Type	A_z
1	$\lvert\mathbf{n}_i \cdot \mathbf{n}_j\rvert$	i	0.9365
2	Absolute difference mean y coordinates	i	0.9466
3	Mean curvature	i	0.9492
4	Absolute difference mean x coordinates	i	0.9542
5	Absolute difference projection main axis on z-axis	i	0.9541
6	Absolute difference curvature	i	0.9522
7	Length	l	0.9482
8	Projection main axis on z-axis	l	0.9713
9	$\det \mathbf{H}$, $\sigma = 4.0$	l	0.9776
10	L_{zz}, $\sigma = 1.0$	l	0.9794
11	L_{yz}, $\sigma = 1.0$	l	0.9834
12	L_{xy}, $\sigma = 1.0$	l	0.9839
13	Absolute difference projection main axis on y-axis	i	0.9838
14	L_{xx}, $\sigma = 4.0$	l	0.9843
15	L, $\sigma = 2.0$	l	0.9855
16	Curvature	l	0.9866
17	L_{xx}, $\sigma = 2.0$	l	0.9874
18	L_{zz}, $\sigma = 4.0$	l	0.9876
19	L_{yz}, $\sigma = 4.0$	l	0.9881
20	$\lVert\nabla L\rVert$, $\sigma = 1.0$	l	0.9890
21	$\det \mathbf{H}$, $\sigma = 1.0$	l	0.9893
22	Projection main axis on x-axis	l	0.9898
23	L_{yy}, $\sigma = 1.0$	l	0.9898
24	L, $\sigma = 4.0$	l	0.9902

The first column gives the number of selected features. The second column gives the name of the feature. The third column denotes the type of the feature (l for local, i for interaction). The last column shows the achieved area under the ROC-curve on the scans used for evaluation.

Table 4
Results of a qualitative evaluation of the rib segmentations

	Into spine	Too short	Too long
Total	43	32	11
Average per scan	2.15	1.6	0.55
Percentage of detected ribs	9.9%	7.4%	2.5%

For each rib it has been determined if it has grown into the bony structures of the spine, if it is too long or if it is too short. The first row shows the total number of ribs, the second the average per scan (20 scans are evaluated) and the last row the results as percentage of the number of detected ribs (in total 435 ribs are detected).

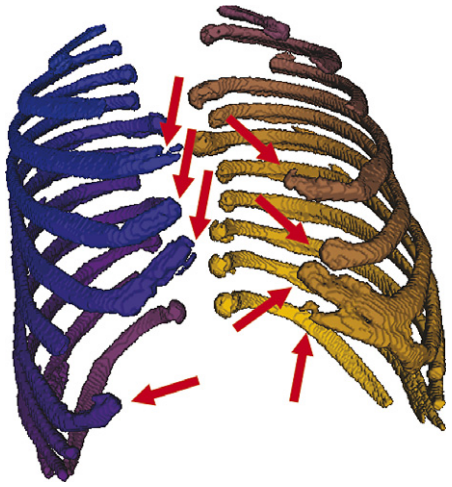

Fig. 12. The red arrows point to 8 ribs which are oversegmented due to calcified cartilage in the ribs. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

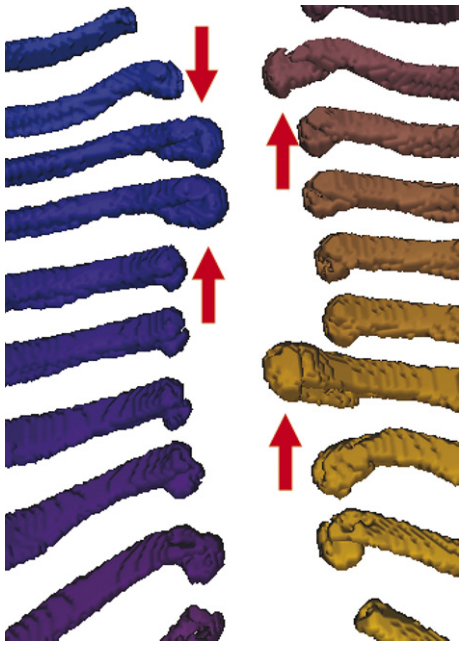

Fig. 13. The arrows point to pieces of the ribs which have grown into the spine.

allows for a quick classification scheme, because the number of primitives is order of magnitudes lower than the number of voxels. For example, the CT-scans in this study contain about 10^7 voxels, as opposed to only 500 primitives on average. Not only is the computation of a lot of features now possible, but manual labeling is feasible too. In the CT-scans on average 160 primitives had to be clicked per scan (the other 340 being background) and manually labeling the primitives of one scan took about 5 min.

The computation time for the rib application is in the order of a few minutes. Computing the features for a scan takes about 1 min, the detection and classification of the primitives about 2.5 min and the region growing about 3 min. All computations were done on a PC with an AMD Athlon XP 1800+ processor and 1 Gb memory. The code was written in C++ and has not been optimized.

For the classification and recognition of the primitives, primitives that are as large as possible are favorable. The size of the primitives is governed by the four parameters that are introduced in Section 4.2. The connectivity parameter ϵ_c is used to bridge gaps which might have been caused by interference of other image structure (crossing of ridges, noise). A too large value of this parameter will give erroneous results if no good candidates are found in the close neighborhood. A value that is too small will give many fragmented primitives. However, the setting is not critical:

larger values (between 5 and 10 voxels) gave similar results. The choice for the parameters for orientation and parallellity, ϵ_0 and ϵ_p, have been chosen fairly strict in the construction for the primitives. If the conditions are weakened (say we choose 0.7 for both parameters) primitives will grow into different objects. For the same reason the threshold ϵ_m on the maximum size was included. If it is decreased the number of primitives that has to be classified increases. If it is increased, primitives which belong to rib and other structures will merge. So there is clearly a tradeoff. After selection of the foreground primitives, however, relaxation of these conditions gives good grouping power, as is evident from the excellent rib detection results.

In the majority of the scans, the ridge voxels of certain vessels were detected because the patients had been administered intravenous contrast agent. However, the use of a classifier enabled the discrimination between vessel and rib primitives. Thus the presented method is relatively insensitive to the presence of contrast material in the scan.

Of course, there are disadvantages with our approach as well. Misclassification of a single primitive is costly: a substantive structure is missed or added. An example is shown in Figs. 9(b) and 10(b). A human being would probably not make such a mistake, because he or she would infer from global reasoning that a part is missing. We conjecture, that in our proposed model, similar behavior can be introduced by a feedback coupling between the fourth and the third stage. However, the details of such a feedback coupling scheme remains an open question at this moment and is subject of future research. In the construction of the rib cage centerlines, we are able to employ simple heuristics to group the primitives and feedback seems not to be necessary for most of the detected ribs (only 1.6 ribs per scan turn out to be too short). However, a better global model for the ribs and rib cage might improve the results, not only for undersegmented objects but also for oversegmented objects, such as those shown in Fig. 12.

In the grouping stage, prior knowledge can be introduced in the form of global information: it is known that human beings have 24 ribs. We exploit this knowledge when the 11 most plausible centerlines on each side are taken as ribs and a possible 12th rib on both sides is searched for. As a result of using global information, the number of false positives is reduced. Again, feedback coupling between the classification and grouping stage might also be advantageous to reduce the number of false negatives. In general, we argue that modelling multiple parts to detect one object increases the robustness of the detection and recognition stage.

The classification task in this paper has been examined with other classifiers as well, such as kNN-classifiers and quadratic discriminant classifiers (Duda et al., 2001). There is no straightforward way to incorporate information between primitives with these classifiers, so we tested them with local features only. Their performance is not bad but significantly worse than that of the spin-glass based classifier.

In the final stage of our algorithm, we have used a seeded region growing algorithm (Adams and Bischof, 1994), because of its simplicity and the nature of the data. This algorithm sometimes leaks a rib into the spine. An anatomical shape model of the part of the rib close to the spine might obtain better results. In order to obtain better segmentation results, the found centerlines can serve as initialization for other, more powerful segmentation schemes, such as deformable models (Kass et al., 1987), active shape models (Cootes et al., 1995), active appearance models (Cootes et al., 2001), level sets (Sethian, 1999; Vasilevskiy and Siddiqi, 2002) and m-reps (Pizer et al., 2003). The obtained results are already good enough to use in further processing, for example to find rib anomalies such as the ones displayed in Fig. 11.

We believe the framework is applicable to a range of medical image analysis tasks. We distinguish three categories: (1) segmentation of a single object (no recognition needed), (2) segmentation of multiple similar objects and (3) segmentation of multiple different objects. As examples we mention segmentation of the aorta (cat. (1)), segmentation of the rib cage (cat. (2)), segmentation of the aorta and iliac and renal arteries (cat. (2)) or the full skeleton (cat. (3)). For category (1) and (2) segmentation tasks only a single type of primitives is needed. Category (3) tasks typically require different types of primitives.

In some category (2) tasks it is possible to construct all primitives directly from the output of the first stage. However, in cases where the objects to be segmented are of different sizes, like aorta and iliac and renal arteries, a multi-scale construction is probably preferable. For category (3) tasks different primitives must be constructed at multiple scales, and the classification stage should be performed on any of these different types of primitives in turn. Categories (1)–(3) pose increasingly harder challenges. For category (1) tasks, recognition is performed by classifying each primitive to foreground or background and establishing the topology between the foreground primitives. For category (2) tasks, putting primitives in the right order is not sufficient for recognition, and interrelationships between grouped objects need to be established. The same is true for category (3) tasks, although recognition of certain parts might be easier due to the different nature of the primitives.

In this work we have considered a task of category (2). Although the framework is generic in nature, it will not be straightforward to simply copy the procedure to solve other multiple object segmentation tasks, especially complex ones. At the level of the individual steps the required methods may be entirely different. It may be necessary to introduce feedback between stages and modeling the relations between primitives may be much more challenging.

In summary: the key contributions that we identify are (1) the presentation of a general applicable framework for detection, recognition and object segmentation; (2) a completely automatic scheme that groups and recognizes image primitives; (3) successful application of the framework

to the task of finding, labeling and segmenting the ribs in chest CT data.

References

Adams, R., Bischof, L., 1994. Seeded region growing. IEEE Trans. Pattern Anal. Mach. Intell. 6 (16), 641–647.

Aylward, S.R., Bullit, E., 2002. Initialization noise singularities and scale in height ridge traversal for tubular object centerline extraction. IEEE Trans. Med. Imag. 21 (2), 61–75.

Aylward, S.R., Bullitt, E., Pizer, S., Eberly, D., 1996. Intensity ridge and widths for tubular object segmentation and description. In: Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis, pp. 131–138.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Cootes, T.F., Cooper, D., Taylor, C.J., Graham, J., 1995. Active shape models – their training and application. Comp. Vis. Image Understanding 61 (1), 38–59.

Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. 23 (6), 681–685.

Duda, R.O., Hart, P.E., Stork, H.G., 2001. Pattern Classification, second ed. Wiley-Interscience, New York.

Eberly, D., 1996. Ridges in Image and Data Analysis. Kluwer Academic Publishers, Dordrecht.

Florack, L.M.J., 1997. Image Structure. Kluwer Academic Press, Dordrecht.

Fukunaga, K., 1990. Statistical Pattern Recognition, second ed. Academic Press, New York.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning. Springer-Verlag.

Kalitzin, S.N., Staal, J.J., ter Haar Romeny, B.M., Viergever, M.A., 2001. A computational method for segmenting topological point sets and application to image analysis. IEEE Trans. Pattern Anal. Mach. Intell. 23 (5), 447–459.

Kang, Y., Engelke, K., Kalender, W.A., 2003. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. IEEE Trans. Med. Imag. 22 (5), 586–598.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: active contour models. Int. J. Comput. Vis. 1 (4), 321–331.

Kim, D., Kim, H., Kang, H. S., 2002. An object-tracking segmentation method: vertebra and rib segmentation in CT images. In: SPIE Proceedings: Medical Imaging 2002, vol. 4684, pp. 1662–1671.

Kiraly, A. P., Qing, S., Shen, H., 2006. A novel visualization method for the ribs within chest volume data. In: SPIE Proceedings: Medical Imaging 2006, vol. 6141, pp. 51–58.

Lindeberg, T., 1994. Scale-space Theory in Computer Vision. Kluwer Academic Publishers, Dordrecht.

Lindeberg, T., 1998. Edge detection and ridge detection with automatic scale selection. Int. J. Comput. Vis. 30 (2), 117–154.

McInerney, T., Hamarneh, G., Shenton, M., Terzopoulos, D., 2002. Deformable organisms for automatic medical image analysis. Med. Image Anal. 6 (3), 251–266.

Metz, C.E., 1978. Basic principles of ROC analysis. Sem. Nucl. Med. 8 (4), 283–298.

Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Graham Gash, A., Glotzer, J.M., Jiroutek, M.R., Lu, C., Muller, K.E., Tracton, G., Yushkevich, P., Chaney, E.L., 2003. Deformable m-reps for 3D medical image segmentation. Int. J. Comput. Vis. 55 (2–3), 85–106.

Prokop, M., Galanski, M. (Eds.), 2003. Spiral and Multislice Computed Tomography of the Body. Thieme, Stuttgart.

Sethian, J.A., 1999. Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press.

Staal, J.J., Abrámoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B., 2004. Ridge based vessel segmentation in color images of the retina. IEEE Trans. Med. Imag. 23 (4), 501–509.

Staal, J.J., Kalitzin, S.N., Viergever, M.A., 2005. A trained spin-glass model for grouping of image primitives. IEEE Trans. Pattern Anal. Mach. Intell. 27 (7), 172–1182.

Summers, R.M., 2003. Road maps for advancement of radiologic computer-aided detection in the 21st century. Radiology 229 (1), 11–13.

Vasilevskiy, A., Siddiqi, K., 2002. Flux-maximizing geometric flows. IEEE Trans. Pattern Anal. Mach. Intell. 24 (12), 565–1578.

Whitney, A.W., 1971. A direct method of non parametric measurement selection. IEEE Trans. Comput. 20 (9), 100–1103.