Medical Image Analysis 16 (2012) 1490–1502

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Clavicle segmentation in chest radiographs

Laurens Hogeweg [a,*], Clara I. Sánchez [a], Pim A. de Jong [b], Pragnya Maduskar [a], Bram van Ginneken [a]

[a] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Geert Grooteplein Zuid 18, 6525 GA Nijmegen, The Netherlands
[b] Department of Radiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 12 January 2012
Received in revised form 11 May 2012
Accepted 25 June 2012
Available online 31 July 2012

Keywords:
Segmentation
Clavicles
Chest radiography
Pixel classification
System combination

ABSTRACT

Automated delineation of anatomical structures in chest radiographs is difficult due to superimposition of multiple structures. In this work an automated technique to segment the clavicles in posterior-anterior chest radiographs is presented in which three methods are combined. Pixel classification is applied in two stages and separately for the interior, the border and the head of the clavicle. This is used as input for active shape model segmentation. Finally dynamic programming is employed with an optimized cost function that combines appearance information of the interior of the clavicle, the border, the head and shape information derived from the active shape model. The method is compared with a number of previously described methods and with independent human observers on a large database. This database contains both normal and abnormal images and will be made publicly available. The mean contour distance of the proposed method on 249 test images is 1.1 ± 1.6 mm and the intersection over union is 0.86 ± 0.10.

1. Introduction

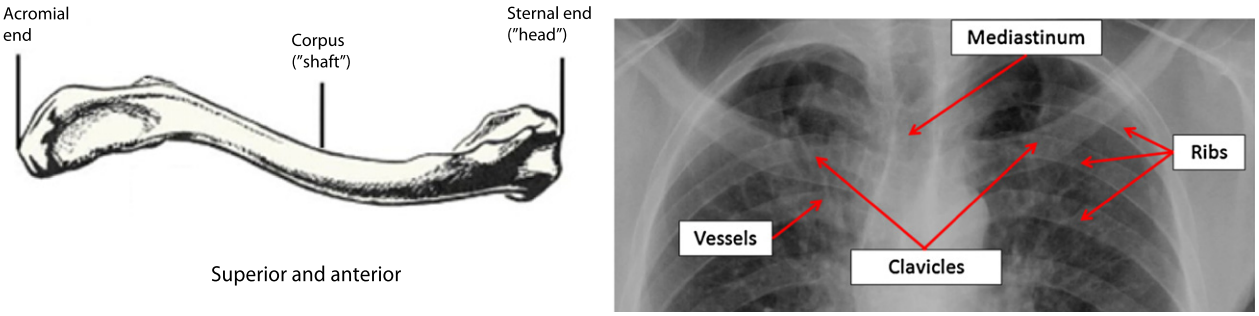
The automatic delineation of normal anatomical structures is a prerequisite for computerized analysis of medical images. The analysis of 2D radiographic images, such as chest radiographs, is a challenging task because superimposed normal and abnormal structures can make it difficult to discern the boundaries of particular objects. Detection, recognition and segmentation of these structures requires incorporating prior knowledge about their location and appearance. The large variation in both these properties in medical imaging as a result of variability in normal anatomy and presence of pathology can be handled through the use of supervised systems that learn from examples.

In this work we focus on the segmentation of the clavicles in chest radiographs. The clavicle is a cortical bone connecting the shoulder blade at the acromion to the breast bone at the sternoclavicular joint. Fig. 1a shows the anatomy of the clavicle in a schematic drawing. Fig. 1b shows the clavicle in a chest radiograph. The 2D projection in a radiograph causes several other structures to overlap the clavicle. Notably these are the ribs, the mediastinum and the larger vessels of the pulmonary vessel tree. In this paper the focus of the segmentation algorithm is the part of the clavicle contained inside the projection of the lung fields and the mediastinum. The lateral parts at the acromial end outside the lung fields are not considered.

Obtaining an accurate segmentation of the clavicles is useful for a number of applications. The segmentation can be used to digitally subtract the clavicle from the radiograph. Several algorithms have been published that address the suppression of bony structures (mainly ribs) in chest radiographs. Suzuki et al. (2006) suppressed ribs using an artificial neural network and showed that their technique increased the visibility of nodules. Loog and van Ginneken (2006a) presented a general filter framework based on regression, which has been applied to the suppression of bony structures on chest radiographs. The method gave promising results but was not further evaluated on a clinical problem. Recently it has been shown that commercially available software that suppresses bony structures in the chest radiograph can improve the radiologist's performance (Freedman et al., 2011; Li et al., 2011). Clavicle suppression in particular might aid radiologists to detect pathology in the lung apices, that are known to be difficult areas due to superimposing structures (Quekel et al., 1999). Performance improvement can also be expected if the automatic segmentation is used as input for computer-aided detection and diagnosis (CAD) systems. Certain lung diseases, such as tuberculosis, manifest themselves especially in the lung apex (van Ginneken et al., 2002). A good characterization of the structures in that area is needed to improve pathology detection and reduce false positives. Accurate localization of the medial parts of the clavicles can also serve to automatically determine possible rotation of the ribcage, an important quality aspect of chest radiographs. When chest radiographs are rotated, false abnormalities might appear in either or both of the lung fields due to apparent changes in parenchymal density.

* Corresponding author. Tel.: +31 24 3616651; fax: +31 243540866.
E-mail address: l.hogeweg@rad.umcn.nl (L. Hogeweg).

1361-8415/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2012.06.009



(a) Schematic drawing (b) Clavicles in the chest radiograph

Fig. 1. Anatomy of the clavicle. (a) Schematic drawing of clavicle with main parts labeled parts. (b) Appearance of clavicles in chest X-ray with surrounding structures indicated. The corpus and sternal end will be colloquially referred to as respectively the shaft and the head, respectively.

Only a few papers have addressed the segmentation of the clavicles. Yu et al. (2005) used dynamic programming and a nonlinear shape model to segment the clavicles but no quantitative error analysis was performed. van Ginneken et al. (2006) segmented the lung fields, heart and clavicles in chest radiographs from the JSRT database (Shiraishi et al., 2000) and compared several segmentation techniques. While results comparable to the interobserver variability were obtained for the heart and lung fields, the segmentation of the clavicles proved a more difficult task. Seghers et al. (2007) used a minimal shape and intensity cost path segmentation technique on the same database. The authors obtained a smaller error on the whole database than reported in van Ginneken et al. (2006), but individual results for the clavicles were not provided. Simkó et al. (2009) considered the task of detecting only the lateral part (diaphysis) of the clavicle. They attempted to find line shaped structures using the Radon transform on edge-enhanced images. The resulting segmentation was subsequently used to suppress the clavicles. Only a qualitative evaluation of the clavicle segmentation was provided.

The combination of different types of independently calculated information might help in the segmentation of anatomical structures, especially in difficult cases. We propose a method that combines different type of characteristics from the clavicles, such as border and head definition, local intensity and shape. These characteristics are extracted from the images by means of pixel classification (PC) and active shape models (ASM) (Cootes et al., 1995) and combined using dynamic programming. The advantage of ASM is that it can only produce plausible shapes, unlike PC which treats segmentation as a local classification problem and can produce segmentations of any shape. In the case of confusing border information, such as overlying ribs or the presence of abnormalities, PC can produce shapes which do not resemble clavicles. A disadvantage of ASM is that its mechanism to generate plausible shapes (through principle component modeling of the training shapes) also limits how accurate it can outline the borders of a previously unseen instance of the structure. Especially when the shape is complex or has a large variability, such as in the case of clavicles, this problem becomes pronounced. This motivated us to use ASM to provide a plausible but rough outline of the clavicle and use multiple dedicated pixel classifiers to refine the outline. Subsequently an optimal cost path is calculated that ensures a globally optimal solution and provides a convenient framework to combine shape information and multiple pixel classifiers.

This paper is organized as follows. In Section 2 the data that was used for both training and evaluation of the system is described. The proposed new method for clavicle segmentation is detailed in Section 3, and various other methods to which the new method is compared are also described there. Section 5 presents a number of experiments and results, and these results are discussed in Section 6. Section 7 concludes.

2. Data

A set of 548 consecutively obtained posterior-anterior chest radiograph were selected from a database containing images acquired at two sites in sub Saharan Africa with a high tuberculosis incidence. All subjects were 15 years or older. Images from digital chest radiography units were used (Delft Imaging Systems, The Netherlands) of varying resolutions, with a typical resolution of 1800×2000 pixels, the pixel size was 250 μm isotropic.

From the total set of images, 225 were considered to be normal by an expert radiologist, while 333 of the images contained abnormalities. Of the abnormal images, 220 contained abnormalities in the upper area of the lung where the clavicle is located. The data was divided into a training and a test set. The training set consisted of 299 images, the test set of 249 images. The development and optimization of the method was completely performed on the training set alone, the test set was only used to calculate the final results.

The database used in the paper will be made publicly available so that it can be used by other research groups to evaluate their segmentation methods. Details can be found in the appendix of this paper.

2.1. Manual segmentation of the clavicles

Manual tracings of the clavicle were used as the reference standard in this study. The clavicle is a high density object projected over a low density background (the lung parenchyma). The border of the clavicle typically appears as a ridge with a higher density than the inside. Even for human experts it can be difficult to delineate this border precisely. The complex cross-sectional shape of the clavicle causes multiple shadows on a chest radiograph. Often the border of the clavicle is partly aligned with the ribs. Especially the medial part of the clavicle is difficult to trace, because of multiple overlapping shadows from the vena cava, ribs, and mediastinal structures. The sternoclavicular joint is not always projected within the lung fields and can be hidden or very difficult to see.

Fixed points were used to determine the shape model for ASM and to evaluate different parts of the clavicle. The fixed points define three parts of the border of the clavicle: (1) the lower border, from fixed point 0–1, (2) the head from fixed points 1–2 and (3) the upper rib border, from fixed point 2–3. The following instructions to outline the clavicles were provided to human observers who provided manual tracings:

1. Start at the lateral inferior border of the clavicle at the projected crossing of the superior border of the scapula and the clavicle (fixed point 0).
2. Follow the inferior border until the start of the head (fixed point 1). The start of the head is defined as the location where the curvature of the border suddenly changes.

3. Follow the border until the end of the head (fixed point 2). The end of the head is defined as the location where the curvature of the border suddenly changes on the superior border.
4. Finish the segmentation by following the superior border until the projected crossing of the superior border of the scapula and the clavicle (fixed point 3).

The clavicles were outlined by three trained readers who were instructed by an expert radiologist. One of the readers was used to set the reference standard. The outlines of the other two readers are compared with the reference standard in the same manner as the automatic methods. These readers are referred to as the 2nd and 3rd observer. To ensure optimal outlines a regular review of the outlines was performed by the expert radiologist.

3. Methods

The proposed Hybrid Dynamic Programming/Active Shape Model/Pixel Classification algorithm (HDAP) combines a selection of existing methods in a structured way to improve on the results of the individual algorithms. A set of dedicated pixel classifier systems form the basis of HDAP, active shape modeling (ASM) is used to generate plausible shapes, and dynamic programming is used to find the exact boundary. Each of the individual algorithms use the output of the previous step(s) in the algorithm as their input(s). A schematic overview of the method is given in Fig. 2. A visual overview of its components is given in Fig. 4.

3.1. Pixel classification (PC)

Pixel classification forms the basis of the other methods. In this methodology the segmentation problem is recast into a pattern classification task (Jain et al., 2000; Duda et al., 2001). A number of continuous characteristics (features) are calculated for a number of samples (positions, pixels) in an image. A classifier is trained using labeled samples from a database of training images. Examples of labels are inside/outside clavicle and on/off the clavicle border. The classifier provides the mapping from features to class labels. In this work we use classifiers that provide a posterior probability that indicates how likely a sample should receive a label. Test images can now be segmented by computing the features for each position and applying the classifier.

3.1.1. Features and classification

Three types of features were calculated for each sample: texture features based on Gaussian derivatives, features derived from the Hessian matrix and position features. First each image is resized to a width of 1024 pixels. To capture local image structure (Florack et al., 1996) the output of Gaussian derivative filtered images of order 0 through 2 ($L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}$), at scales 1, 2, 4, 8, 16, 32 and 64 pixels were calculated. Hessian matrix derived features derived were used to detect the presence of line like structures (Frangi et al., 1998). Considering the two eigenvalues of the Hessian matrix $\lambda_1, \lambda_2, |\lambda_1| > |\lambda_2|$ two measures were derived: (1) $\sqrt{(\lambda_1^2 - \lambda_2^2)}$ to extract the lineness of the local image structure and the largest absolute eigenvalue $|\lambda_1|$ to indicate the strength of the response. The typical location of the clavicles in the image was captured through a number of spatial features: the (x, y) coordinates in the image resized to a width of 1024 pixels, the normalized (x, y) coordinates inside the bounding box of the unobscured lung fields, the distance of the pixel to the boundary of the lungs, and the distance of the pixel to the center of gravity of both lungs. An automatic lung segmentation algorithm was used to find the unobscured lung fields (van Ginneken et al., 2006).

In total 59 features were computed. Normalization of each feature to unit standard deviation and zero mean was performed beforehand in all cases. To reduce computation time, features were sampled every 2nd pixel on a regular grid in both the training and test images so that the resulting segmented images have a width of 512 pixels.

To construct the training set positive (clavicle) samples per image were randomly sampled (sample rates for each classification task are given below). For the negative samples the distance to the clavicle was used to control the sampling rate (see Section 3.1.2). One sixth of the positive and negative samples was used for the final training set, while the remaining 5/6th was used to evaluate the effect of feature selection and classifier selection.

3.1.2. Object body classification

A typical approach for segmentation using pixel classification is to detect all pixels inside the object (van Ginneken et al., 2006). Positive examples are selected from the inside of the object of interest and negative examples from outside. The clavicles are only a small part of the chest radiograph and to prevent the classifier focusing too much on the background structures the number of negative examples were sampled depending on the distance to the clavicle. 80% of the negative examples were sampled randomly within a distance of 10 mm of the clavicle, the remaining 20% was sampled from the rest of the image. A similar sample strategy was successfully used to segment fissures in thoracic computed tomography scans (van Rikxoort et al., 2008). The sampling rate for nearby negative and positive examples was 1.5%.

The computational burden of the classification was reduced by performing it in two steps. An initial rough classification of the object body was performed using a *kNN*-classifier with $k = 5$ and

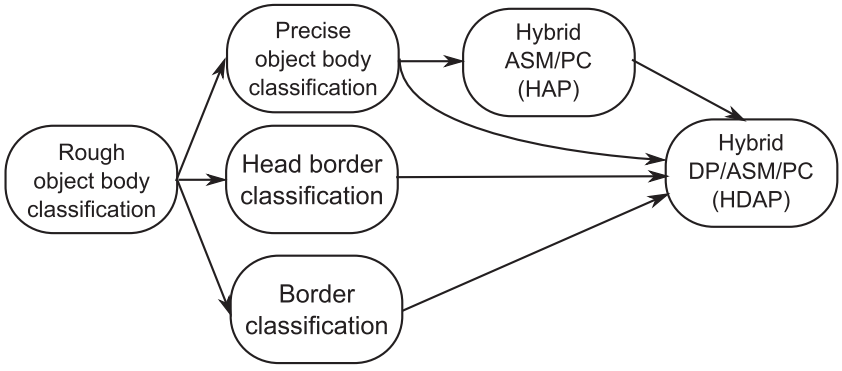


Fig. 2. Flowchart of HDAP. A number of independent dedicated pixel classifiers are combined using active shape modeling (ASM) and dynamic programming.

the first five features selected by feature selection. This quick classification gives a robust initial detection and localization of the clavicles. A rectangular area around this approximate detection was determined to restrict the search space for the second classification stage. Posterior probabilities p were thresholded with $p = 0.5$, the bounding box was determined and dilated by 20 mm. Construction of the training set and classification of samples in test images was performed only in this search area for the subsequent precise pixel classifiers.

The output of pixel classifiers typically have a grainy noisy appearance. To reduce this effect and to obtain a single connected segmentation for each object (right and left clavicle) the output was post-processed. The output of the PC was blurred with $\sigma = 0.7$ mm and then thresholded with $p = 0.5$ to obtain a binary segmentation (van Ginneken, 2006). The two largest connected components were determined and holes were filled using a morphological closing with a circular kernel (radius = 10 pixels).

3.1.3. Object border classification

The border of the clavicles on chest radiographs has a distinct appearance from the body, appearing as a sharp dense ridge. In this work the border was separately classified from the object body to account for this difference in appearance. The rationale is that samples, features and classifiers can be independently optimized for body and border separately. Positive samples, representing the border, were selected within a distance of $n = 2$ pixels from the outline of the clavicles. The negative samples were selected with a minimum distance of $1.5n$ pixels to the outline to prevent sampling false negatives as a consequence of the inaccuracy in manual outlining the border. To focus on the task of distinguishing the object border from its direct surroundings, 50% of the background pixels were selected within 10 pixels from the outline, the other 50% was randomly sampled within the previously determined approximate clavicle bounding box. The sampling rate for nearby negative and positive examples was respectively 10% and 50%. The sampling rules are summarized in Table 1.

3.1.4. Head border classification

The head of the clavicle is a very difficult area to segment automatically and has an appearance that is distinct from the shaft of the clavicle. A separate classifier was constructed for the head border. The sample strategy was the same as for the object border classification, but positive examples were taken only from the outline of the head (between fixed points 1 and 2). The sampling rate for nearby negative and positive examples was 30%.

3.2. Hybrid ASM/PC (HAP)

Active shape modeling (ASM) (Cootes et al., 1995) is a popular method to segment structures in medical images. We use the implementation described in Cootes and Taylor (2001) which uses a global shape model, a multi-resolution appearance model and a multi-resolution search algorithm. The steps of the algorithm are shortly repeated here.

The shape of one or more objects is described by n points combined in a vector $x = (x_1, y_1, \ldots, x_n, y_n)^T$. A shape model is trained from a set of training shapes by determining the mean shape and the principal modes of variations using Principal Component Analysis (PCA). Gray value profiles are sampled perpendicular to the object border at each point of the shape. The first derivative of the profile is calculated and the profile is normalized. The best position of a point during search is determined by minimizing the Mahalanobis distance of the derivative profile to the appearance model created from profiles of corresponding points in the training set. The fitting of the shape is performed in an iterative scheme where points are alternately moved to their optimal position according to the appearance model and then projected on the shape model. This projection is performed using least square fitting with a bound on the maximal variation. A multi-resolution appearance model is used to prevent the search algorithm finding local optima. Shape alignment is not performed before PCA to increase the ability of the model to fit a large variability of shapes and locations.

ASM is run on the output of the object body classifier instead of the original gray values, combining pixel classification and ASM into a hybrid ASM/PC (HAP) algorithm. The probabilistic output of the pixel classifier is used directly as input for the ASM method, without the post-processing described in Section 3.1.2.

The ASM algorithm requires a set of training shapes with corresponding landmarks. The correspondence is provided by employing the four fixed points indicated during manual annotation (see Section 2.1). Between these fixed points a constant number of landmarks were interpolated over the initially annotated contour. On the lower clavicle border 20 points were interpolated, on the head 21 points were taken and on the upper clavicle border again 20 points, leading to a total of 65 landmarks per clavicle. The two clavicles are combined into one shape model containing two objects, consisting of 130 landmarks. By combining both clavicles in one model unlikely configurations, such as gross asymmetry, will be prevented during fitting as these configurations do not typically occur in (normal) chest radiographs.

Table 1
Sample strategy for border classification.

Distance to border (pixels)	Class	Sampling rate
0–2	Positive	50% of available samples
2–3	–	Not sampled
2–10	Negative	10% of available samples
10+	Negative	Same number as nearby negative samples

3.3. Hybrid DP/ASM/PC (HDAP)

We note that the pixel classification method has high accuracy in areas where the border of the clavicles can be easily discerned. In areas where the output of the pixel classifier is uncertain another step is needed to integrate the local decisions. Contextual methods such as applying smoothing, mathematical morphology, iterated contextual classification (Loog and van Ginneken, 2006b) or Markov Random Fields (Vittitoe et al., 1999) will generally improve the results by including information of the local surroundings but they do not provide a global integration of the available information.

Optimal path based methods (Bellman, 1962; Montanari, 1971), on the other hand, can provide this global context by finding the combination of local decisions that form the best solution given the evidence provided by the local (pixel classifiers) and global (shape) information sources. In an appropriate formulation the optimal path can be easily found using dynamic programming.

To detect the border of the clavicles with dynamic programming the image must be warped in an appropriate coordinate system. In the coordinate system we use here the border is an approximately straight line. The coordinate system is created by sampling profiles of pixel values from points on the border to the closest point at the primary medial axis (PMA, see Section 3.3.1) of the (estimated) object. The spatial sampling frequency is adjusted to the length of the profile so that in the transformed coordinate system the distance from PMA to the border is approximately the same amount of pixels. As dynamic programming must be able to improve the detected border, the profiles are extended on the outside of the object. A large margin of twice the PMA-contour distance was chosen as the length of the profiles.

The result of this transformation is a rectangular optimal path space with one axis aligned to the clavicle border and the other axis aligned to the profiles.

3.3.1. Primary medial axis (PMA)

The skeleton of the output of the HAP method is found using the medial axis transform (Blum, 1967). At the end of elongated objects the skeleton will typically have short branches emanating from the corners and running to the main centerline unless the ends are perfectly spherical. These (shorter) branches are removed until only one segment remains. The remaining single centerline of the elongated object is called the primary medial axis (PMA).

3.3.2. Cost function

Any type of image can be projected in the coordinate system of the optimal path space to form a cost function for the dynamic programming. The outputs of the different pixel classifiers described above and the output of HAP are combined in one cost function

$$C = C_{border} + \alpha C_{body} + \beta C_{shape}(\sigma) \quad (1)$$

where C is the final cost function, C_{border} the sum of the outputs of the object border and head border classification, C_{body} is the derivative of the output of the body classification in the profile direction (in optimal path space) and C_{shape} represents the influence of the shape found by HAP (see Fig. 3a–e). The conversion of C_{body} by taking the derivative is needed to convert the original step edge formed by the border to a ridge in the cost function. The profile of C_{shape} is formed by Gaussian blurring (the scale controlled by the parameter σ) the clavicle outline that was obtained using HAP. The parameters α and β control the relative influence of each of the components. The optimal path is found by maximizing the cost function (see Fig. 3f).

4. Experimental

The classifier settings, and the weights in the cost function combination were determined in a number of pilot experiments.

4.1. Feature and classifier selection

The optimal classifier for each of the pixel classifiers was determined using classifier and feature selection. The following classifiers were evaluated. A *kNN*-classifier with $k = 5$, 15, 30, 50, 100, and 200 and a linear discriminant classifier (LDC) with and without Principal Component Analysis to reduce the number of features. For *kNN*-classification the fast tree-based implementation by Arya et al. (1998) was used. Previous work indicated that the approximate solution given by this implementation does not influence the classification results (van Ginneken et al., 2006). Feature selection was performed using Sequential Forward Selection (SFFS) (Pudil et al., 1994).

The effect of classifier and feature selection was evaluated using the area under the ROC curve (A_z) on the pixel datasets. LDA classifiers performed in general much worse than *kNN*-classifiers. Minor differences in A_z were found between different values for k. Improvements in classifier performance (A_z) were not directly reflected in the segmentation performance measures (Ω and MCD). A similar effect was found for the feature selection where small increases in classifier performance were observed with feature selection but no increase or even a decrease in segmentation performance. For this reason all subsequent results are shown for a *kNN*-classifier ($k = 15$) using the full set of 59 features and feature normalization.

4.2. Cost function weights

C_{border} and C_{body} in Eq. 1 both indicate the border likelihood with the same scaling, the factor α controlling their combination was therefore set to 1. The influence of the shape model, determined by β and σ, was optimized by doing a grid search on these parameters in the training set. For β the values $(0.0, 0.05, 0.10, 0.20, 0.5, 1.0)$ and for σ the values $(1, 2, 4, 6, 8, 12)$ were tested. On the training set the combination $(\beta, \sigma) = (0.1, 4.0)$ yielded the best results and these values were used in subsequent experiments.

5. Results

5.1. Segmentation performance metric

To measure the performance of a segmentation algorithm, a 'goodness' index is required. For a two class segmentation problem, one can distinguish true positive (TP) area (correctly classified as object), false positive (FP) area (classified as object, but in fact background), false negative (FN) area (classified as background, but in fact object), and true negative (TN) area (correctly classified as background). From these values, measures such as accuracy, sensitivity, specificity, kappa and overlap can be computed. In this work we use the intersection divided by union as an overlap measure, given by

$$\Omega = \frac{TP}{TP + FP + FN}. \quad (2)$$

This is a well accepted measure, but one should be aware that objects that are small or thin or have a complex shape usually achieve a lower Ω than larger and more spherical objects (Gerig et al., 2001). For many purposes the part of the clavicle inside the lung fields is most relevant, and the overlap was calculated only inside the convex hull of the automatically segmented lung fields obtained using van Ginneken et al. (2006).

In addition, the mean absolute contour distance (MCD) is computed. For each point on contour X, the closest point on contour Y is computed; these values are averaged over all points; this is repeated with contours X and Y interchanged to make the measure symmetric (Gerig et al., 2001). The distances are given in millimeter. One pixel corresponds to 0.78–1.22 mm on the images of 512 pixels wide depending on the width of the original image.

For comparisons between methods, paired Student's t-tests to test the difference between means were used. Differences are considered significant if $p < 0.05$.

5.2. Segmentation results

The proposed HDAP method was compared with a number of other systems. Performance measures were calculated by considering the annotations of the first observer as the reference standard. The annotations of the 2nd and 3rd observer, the tuned version of the ASM from van Ginneken et al. (2006), as well as the object body classification with post-processing (PC-postproc) and the hybrid ASM/PC (HAP) method were evaluated. For ASM the same configuration as for HAP was used, except original gray values instead of posterior probabilities were used as input. In addition four different configurations of HDAP were evaluated to study the effects of including and excluding various components of the algorithm. 'HDAP' is the complete algorithm as described in Section 3.3. 'HDAP: no border' only uses shape and the object body classification to create the cost function in Eq. 1. 'HDAP: no head' does not include the dedicated head border classification in the cost space but does use the object border classification. The

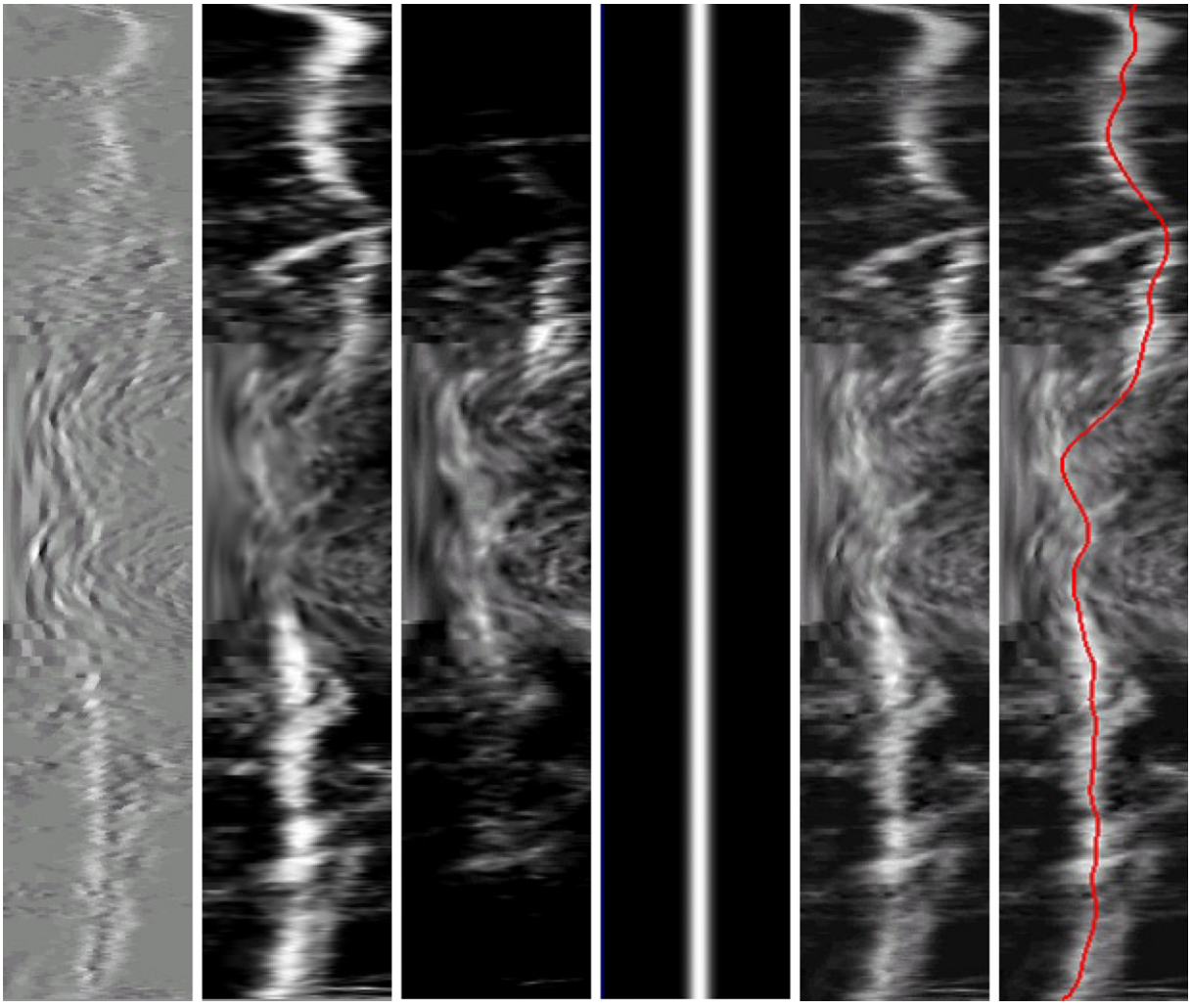


Fig. 3. Composition of cost function in optimal path space. High values indicate a high clavicle border likelihood. (a)–(d) form the components of the cost function (see Eq. 1). (a)–(c) are derived from pixel classifiers by sampling profiles: (a) is based on Fig. 4c, (b) is based on Fig. 4d, (c) is based on Fig. 4e. (d) Is the ASM fitted shape, based on 4f. The total cost is shown in (e), where (f) shows the optimal path.

'HDAP: no shape' variant does not include the term in Eq. 1 that controls the influence of the contour found by HAP.

Results of the different evaluated methods are shown in Tables 2 and 3 for respectively MCD and Ω. In Figs. 5 and 6 the same results are shown graphically as boxplots (McGill et al., 1978). In both figures and tables the results are ordered according to the median of the results (best performing first). Methods that show significant improvement compared to the previously listed method are indicated with an asterisk.

All the HDAP methods improve significantly compared to the other automatic methods measured by both MCD and overlap. From the HDAP methods the no shape variant is the best performing method, with a slightly higher MCD and overlap than HDAP. HDAP: no border is the least performing of the variants, indicating that object border classification is an important part of the algorithm. When the head border classification is included in the cost function results improve significantly, especially the MCD.

Fig. 7 shows the outlines provided by the reference standard and five compared segmentations (2nd observer, HDAP: no shape, HAP, ASM and PC-postproc) for four selected cases. The four cases are chosen by ranking all results according to the average MCD of the four shown automated methods and then choosing for display the 0%, 33%, 66% and 100% percentile, respectively position #1, #85, #168 and #249. The performance increase as a result of the border refinement from HDAP to HAP can be clearly seen for all the cases. The last, worst segmented case, exemplifies one of the advantages of using HAP instead of ASM. In this radiograph the lung tops have been cut off due to poor collimation. ASM is not robust against these kinds of outliers resulting in 0 overlap with the reference standard, while the object body classification provides enough information for HAP to provide a reasonable segmentation.

5.3. Evaluation of different parts of the clavicle

The lateral parts of the clavicles are often clearly visible and relatively easy to segment, while the medial part of the clavicle is often obscured by numerous other structures such as the mediastinum or large vessels. The first row of Table 4 shows for the best performing method, HDAP: no shape, how the algorithm performs for different parts of the clavicle. Shown is the mean MCD of all the test cases for three different parts of the clavicle (see Section 2.1).

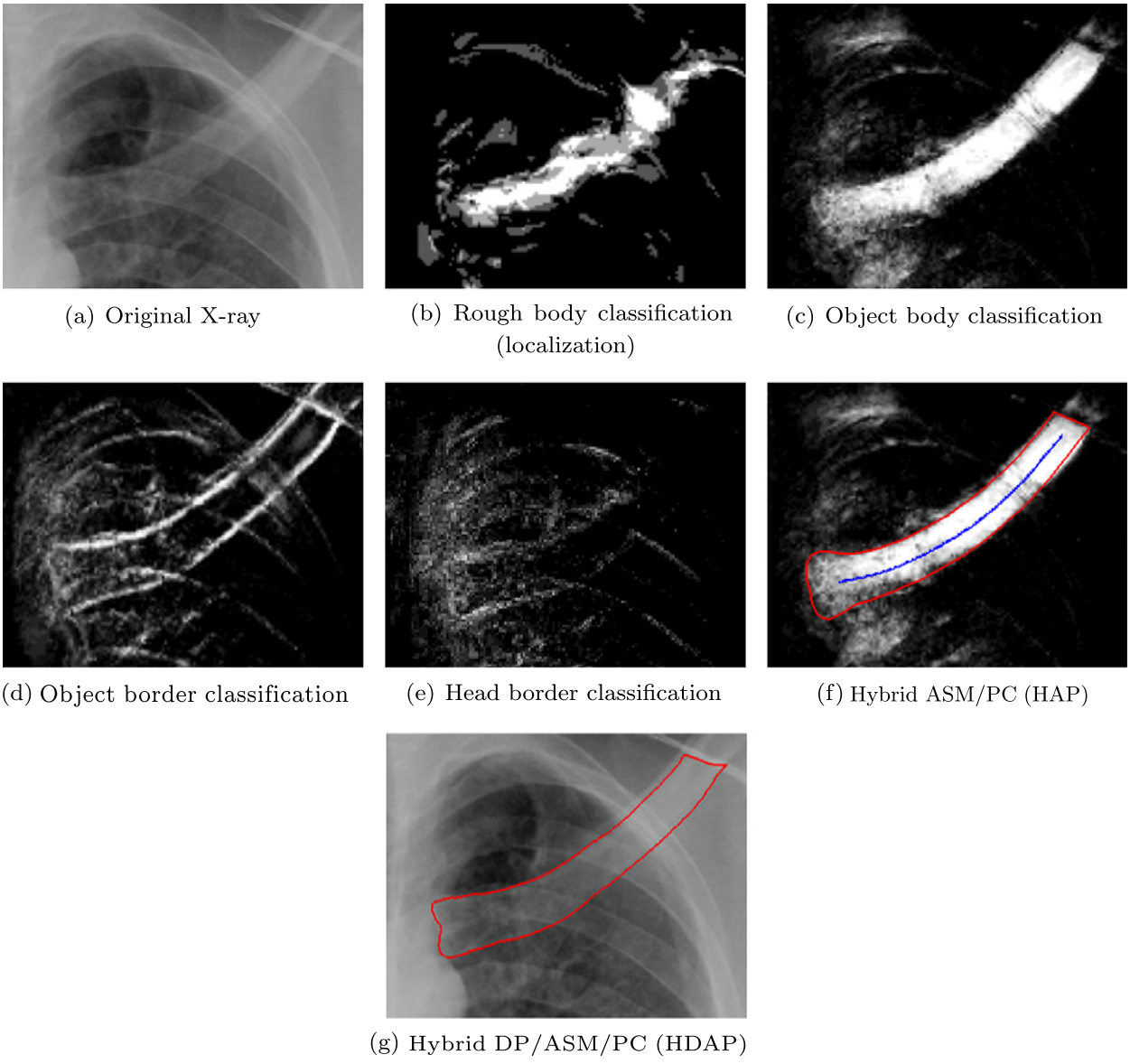


Fig. 4. Outputs of the algorithm shown for one case. Corresponding cost space images are shown in Fig. 3.

Table 2
Segmentation results for the different methods. The mean contour distance (MCD; in mm) is given. Methods are ranked according to their median MCD. Methods which significantly ($p < 0.05$) improve over the method below it are indicated with an asterisk.

MCD	Mean	Stdev	Min	$q1$	Median	$q3$	Max
3rd observer	0.48	0.28	0.19	0.32	0.41	0.54	2.48
2nd observer *	0.49	0.25	0.19	0.34	0.43	0.57	1.94
HDAP: no shape *	1.09	1.57	0.28	0.57	0.81	1.15	22.74
HDAP *	1.10	1.57	0.27	0.56	0.82	1.16	22.65
HDAP: no head *	1.15	1.63	0.25	0.60	0.86	1.25	23.53
HDAP: no border *	1.49	1.58	0.39	0.92	1.20	1.59	21.44
HAP *	1.83	1.62	0.56	1.20	1.56	2.01	22.04
ASM	3.62	7.78	0.55	1.35	1.79	2.60	87.57
PC-postproc	2.74	4.46	0.94	1.50	1.86	2.47	44.28

Table 3
Segmentation results for the different methods. The overlap Ω is given. Methods are ranked according to their median Ω. Methods which significantly ($p < 0.05$) improve over the method below it are indicated with an asterisk.

Ω	Mean	Stdev	Min	$q1$	Median	$q3$	Max
3rd observer	0.93	0.04	0.70	0.92	0.94	0.95	0.97
2nd observer *	0.93	0.04	0.73	0.92	0.94	0.95	0.97
HDAP: no shape *	0.86	0.10	0.03	0.83	0.88	0.91	0.96
HDAP *	0.85	0.10	0.03	0.83	0.87	0.91	0.95
HDAP: no head *	0.85	0.10	0.02	0.82	0.87	0.91	0.96
HDAP: no border *	0.80	0.10	0.05	0.77	0.82	0.86	0.93
HAP *	0.77	0.10	0.05	0.74	0.78	0.83	0.92
PC-postproc *	0.73	0.11	0.18	0.70	0.75	0.80	0.89
ASM	0.69	0.19	0.00	0.65	0.75	0.80	0.90

The lower and upper border are most accurately segmented with an average error of approximately 0.5 mm. The head section of the clavicle is much harder to segment and has an error of about 2.5 pixels. To see if the adding of the head border classifier improved performance the second row shows the results for the HDAP: no head method. The addition of the head border classification improves the MCD on the head by about 0.16 mm (0.2 pixels)

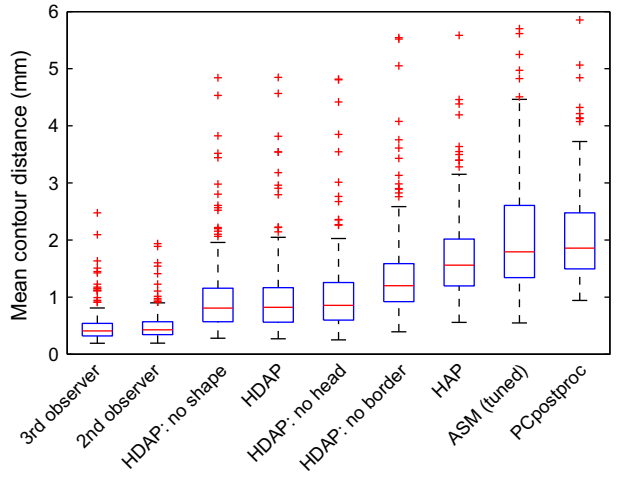


Fig. 5. Boxplots of the mean contour distance on the test set of 249 cases for the different methods. The central line indicates the median, the box edges the 25th and 75th percentiles and the whiskers the extremes of the data excluding outliers. Points are considered outliers if they lie more than 2.7σ from the mean, corresponding to 99.3% percent of the data. The corresponding numbers are listed in Table 2.

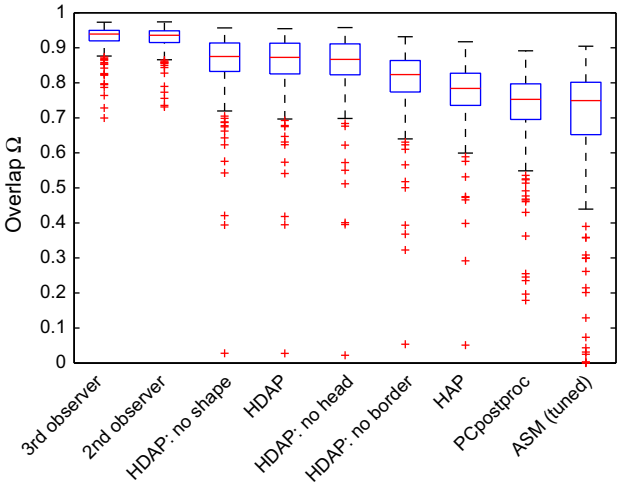


Fig. 6. Boxplots of the overlap Ω on the test set of 249 cases for the different methods. The central line indicates the median, the box edges the 25th and 75th percentiles and the whiskers the extremes of the data excluding outliers. Points are considered outliers if they lie more than 2.7σ from the mean, corresponding to 99.3% percent of the data. The corresponding numbers are listed in Table 2.

on average. Results for HAP are shown in the last row and indicate that the addition of dynamic programming improves result by approximately 0.5 pixels for the lower and upper border and by 0.25 pixels for the head.

The right part of the table shows the same results for the subset of images in the test set containing abnormalities in the upper lung area ($n = 103$). The mean MCD for the lower and upper border is only slightly higher compared to the results for all images, while errors increase more in the head. MCD for images containing abnormalities in other parts of the lung ($n = 49$) were similar to those of normal images.

5.4. Effects of ASM and DP

Fig. 8 shows the effects of the different components of HDAP on the MCD. Each point in the plot is a radiograph, if a point lies on the identity line no improvement was observed between the two compared methods, if the point is above the line the error has increased, if it is below the line the error has decreased. Fig. 8a shows the change of error when ASM is applied to the pixel classification output (HAP). In most cases the error decreases, indicating the ability of HAP to correct for some of the errors made by PC-postproc. A number of cases show a large reduction in the error. Even with perfect PC output HAP will still make errors as a result of the limited capability of the model to fit complex shapes. To demonstrate this, reference outlines were directly fitted to the shape model providing an upper bound for ASM/HAP. A mean MCD and overlap of respectively 0.93 ± 2.11 mm and 0.86 ± 0.050 were found, indicating the need to refine the contour provided by HAP.

Fig. 8b shows the same type of scatter plot for the change in error when adding dynamic programming (from HAP to HDAP). For the large majority of cases the error decreases again. Cases with a large error when using HAP also typically have a large error after applying HDAP, indicating that HDAP especially improves on cases where the true border has been found already approximately.

6. Discussion

A hybrid segmentation algorithm, HDAP, has been presented to segment the clavicles in chest radiographs and has been evaluated on a large database of normal and abnormal radiographs. Two main conclusions can be drawn from the results: (1) the addition of dynamic programming to a combination of other algorithms significantly improves the segmentation performance compared to the previous state-of-the-art algorithm (2) the automated segmentation of clavicles in chest radiographs is a difficult problem and does not yet achieve the same performance as human observers. In this discussion the merits of the presented HDAP algorithm will be pointed out first, including a comparison with previously published work. Then possible reasons for the lower performance compared to human readers are discussed. Finally a number of recommendations for future research are given.

6.1. Optimal path searching as contour refinement

Searching for an optimal path in a cost space in order to segment objects is an approach with some favorable properties. By modifying the cost space in an appropriate way it is easy to encode extra information into the algorithm (Timp and Karssemeijer, 2004). The use of dynamic programming to locate the optimal path ensures a global solution to the problem. HDAP provides a double integration of local and global information. First local cues provided by the object body classification are integrated with shape knowledge using HAP and subsequently additional local cues provided by different pixel classifiers are added. For the sake of simplicity and elegance it might be argued that one should focus on using and optimizing one individual algorithm instead of combining an array of methods. It is unlikely though that one single algorithm can robustly deal with all kinds of variation encountered in medical image segmentation.

The basic problem in all combinations of methods is to retain the good parts of each method and discard the bad parts. In the case of segmentation it can be expected that certain algorithms perform better at localizing certain parts of the object or perform better for certain cases. In general it is very difficult to determine for a single specific case whether a method has succeeded (at specific locations) or not. If a procedure would exist to detect errors in method for particular cases it is often also possible to correct these errors. A few of such attempts have been performed. In an application to segment breast masses on ultrasound (Cui et al., 2009)

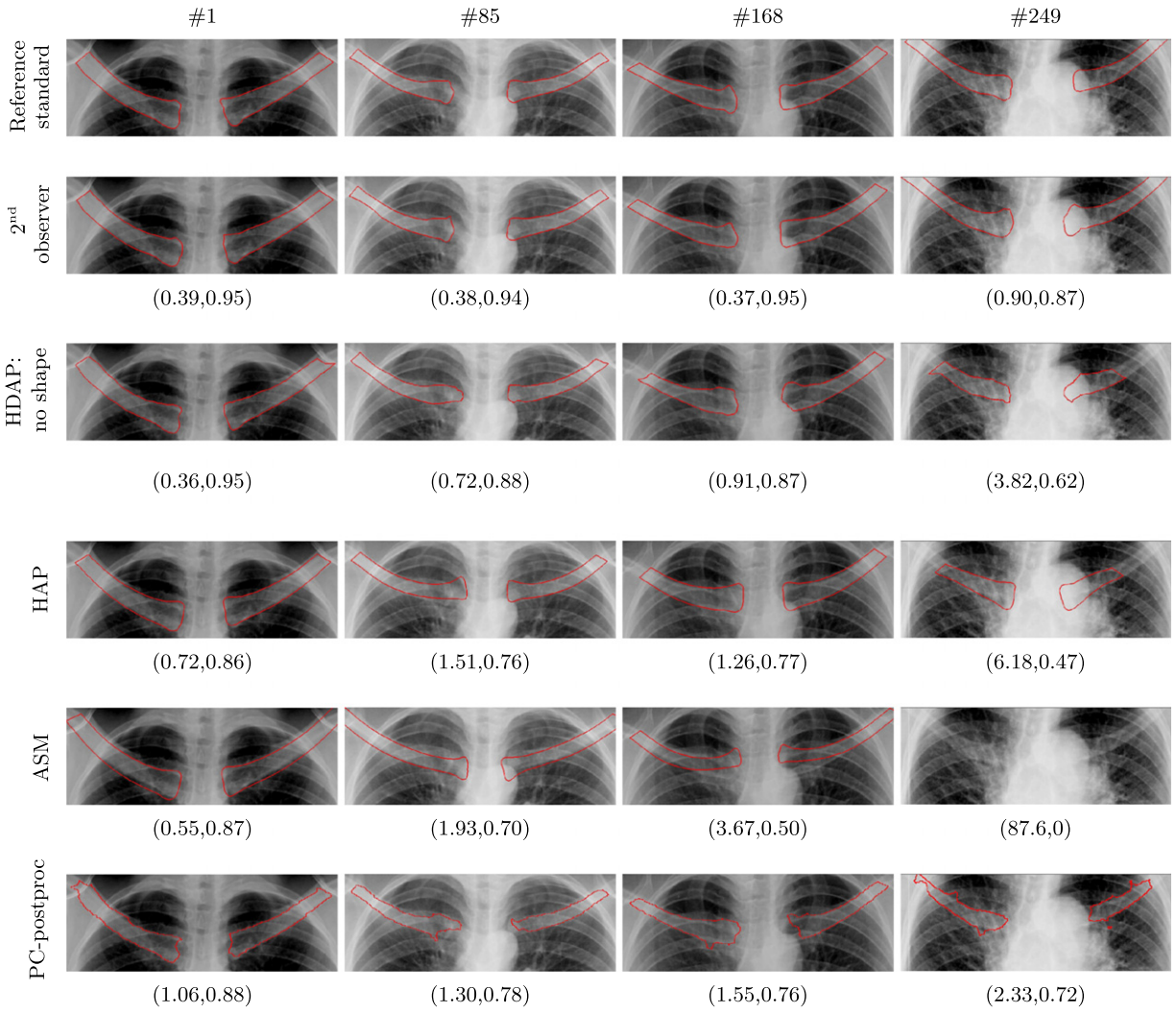


Fig. 7. Examples for selected methods. Shown from left to right are respectively the #1, #85, #168 and #249 cases ranked to the average MCD of the displayed methods (from top to bottom). Indicated below each example are the MCD (in mm) and the overlap Ω.

Table 4
Mean contour distance (MCD; in mm) on the different sections of the clavicle. Values given are mean ± standard deviation. The first row shows the best performing method. The second row indicates the performance difference when the head border cost function is not used. For comparison the results of HAP are also shown.

Method	All images			Images with abnormalities in upper lung area		
	Lower border	Head	Upper border	Lower border	Head	Upper border
HDAP: no shape	0.49 ± 1.27	2.84 ± 3.33	0.61 ± 1.64	0.60 ± 1.79	3.56 ± 4.36	0.63 ± 1.50
HDAP: no head	0.49 ± 1.24	3.00 ± 3.33	0.59 ± 1.61	0.60 ± 1.76	3.72 ± 4.31	0.61 ± 1.41
HAP	1.18 ± 1.15	3.39 ± 3.06	1.08 ± 1.72	1.27 ± 1.43	4.03 ± 3.97	1.10 ± 1.40

defined a goodness-of-fit criterion for points on a contour found using a snake model. If the points were not properly located, a second specialized fitting stage was used to improve the segmentation. van Rikxoort et al. (2009) used error detection based on shape statistics derived from a training set of failed segmentation results to switch to a more advanced (and computationally more expensive) segmentation algorithm in case of a failure. In the general case, when such a procedure does not exist, some sort of averaging or majority voting can be used (Warfield et al., 2004; van Ginneken et al., 2006; Artaechevarria et al., 2009).

We argue that HDAP uses a special kind of averaging where evidence is accumulated from different types of pixel classifiers by adding them in the cost space. In general increasing the number of components in a system is expected to increase robustness as long as the errors of the individual components are not too large and are complementary to each other, a result well known from the field of classifier combination (Kuncheva, 2002). For this particular application of segmenting the clavicles especially detecting the border separately of the object proved to yield a large performance improvement.

The idea of refining an initial rough object detection has been explored before in Brejl and Sonka (2000). A shape variant Hough transform was used to solve the problem of generating an initial detection and outline of the object of interest. A border appearance model was then created by sampling profiles perpendicular to the border and calculating a number of features from them. The fit values of border points (based on Mahalanobis distance) in a test image is then used in an active snake model to refine the initial

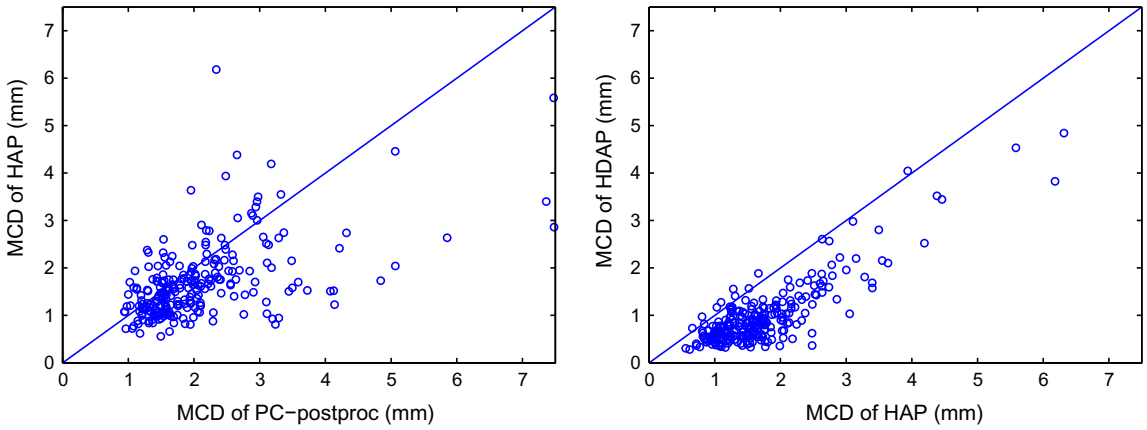


Fig. 8. Effect of steps in HDAP on MCD. The left scatterplot shows the change in MCD from the output of the post-processed body pixel classification (PC-postproc) to HAP, each point being one case. For most cases an improvement of the MCD is observed. The right image shows a similar plot for the change from HAP to HDAP, in almost all the cases an improvement in MCD occurs.

outline. While similar in overall approach of the problem there are some important differences with our work. The initial detection of HDAP is based on the appearance of the texture of the object and not only on the appearance of the border. This adds improved robustness when the border detection provides confusing results in individual images or when it is difficult to generate an accurate border appearance model. The generation of the cost space for border refinement also differs; HDAP uses cost functions derived from a set of dedicated pixel classifiers which can be easily expanded by choosing different training sets for different sections.

6.2. Comparison with previous approaches to clavicle segmentation

In previous work on the segmentation of clavicles in chest radiographs (van Ginneken et al., 2006) the set-up for the ASM was different as also the heart and the lungs were included in the model. Adding the lungs to the shape model might give a benefit in some cases, but due to the inherent limitation of flexibility of shape models it is not expected that performance will actually increase. The database used in this work is also different from the JSRT database used in van Ginneken et al. (2006). The JSRT database is a lung nodule database and in only a few cases the clavicular area is affected by the presence of a nodule. Instead, the database used in this work contained a considerable number of abnormal images.

To be able to compare the results, Table 5 shows results for the images containing no abnormalities in the lung top. The mean MCD for ASM is higher in this work than in van Ginneken et al. (2006). The use of (severely) abnormal training images can be an explanation for this. Also the "tuned" parameters for ASM were actually tuned on the database from van Ginneken et al. (2006). These issues were not further investigated because ASM only serves as a reference method. HAP shows higher performance than ASM here, while on the JSRT database the reverse is observed. This change can be a consequence of the higher robustness of HAP against pathology and the better performance of the object body classification. Most importantly, HDAP (all variants including no shape, the best one), shows errors smaller than the other methods on either of the two databases.

HDAP shows some similarities to the method proposed by Yu et al. (2005). These authors used a method that alternates the application of nonlinear shape model fitting and dynamic programming contour refinement to the segmentation of both lungs and clavicles in chest radiographs. While they stress the importance of nonlinear shape models compared to standard PCA, their method fails to improve over standard ASM for the lung fields. A reason for this could be that the cost function in their search space is based only on intensity differences (to find the edges) instead of on multiple specialized border detectors as in HDAP. No numerical results are given in Yu et al. (2005) for the accuracy of the segmentation of the clavicles.

Table 5
Comparison of results in this paper with van Ginneken et al. (2006). For a fair comparison only images containing no abnormalities in the lung top (148/249) are used to calculate the mean MCD.

Method	van Ginneken et al. (2006)	This work
ASM	2.04	3.52
Hybrid ASM/PC	2.78	1.60
PC-postproc	2.90	1.92
HDAP: no shape	–	0.89
2nd observer	0.68	0.44

Simkó et al. (2009) developed a method to detect the diaphysis (shaft) of the clavicle using an initial detection with a Radon transform and a subsequent contour refinement using active contour fitting. Only qualitative results on the segmentation accuracy were reported. These authors also showed the ability of suppressing the previously automatically detected clavicle to reduce the number of false positives of a nodule detection system applied to the publicly available JSRT database (Shiraishi et al., 2000).

Seghers et al. (2007) used a combination of local (point) appearance models and local shape information based on the orientation vector between two consecutive points to generate a cost space. By using dynamic programming to find the optimal path in this space, the problem of generating plausible shapes with control points at plausible locations is solved at once. This method, called minimum intensity and shape cost path (MISCP), has the important advantage that no iterated methods, such as ASM, which suffer from local minima, are used. The immediate drawback of this is that no global shape information is encoded and that the method might produce unrealistic results when individual points cannot be detected reliably. Using the PC output as input for ASM in HDAP largely solves the problem of local minima during fitting. Also the border refinement in HDAP ensures a precise segmentation over the whole border of the object instead of only at a number of fixed points as in MISCP.

6.3. Improving HDAP

The mostly linear relation between the MCD of HAP and HDAP (Fig. 8) indicates that if the initial border is located too far away from the real border, HDAP cannot improve on HAP. The maximal

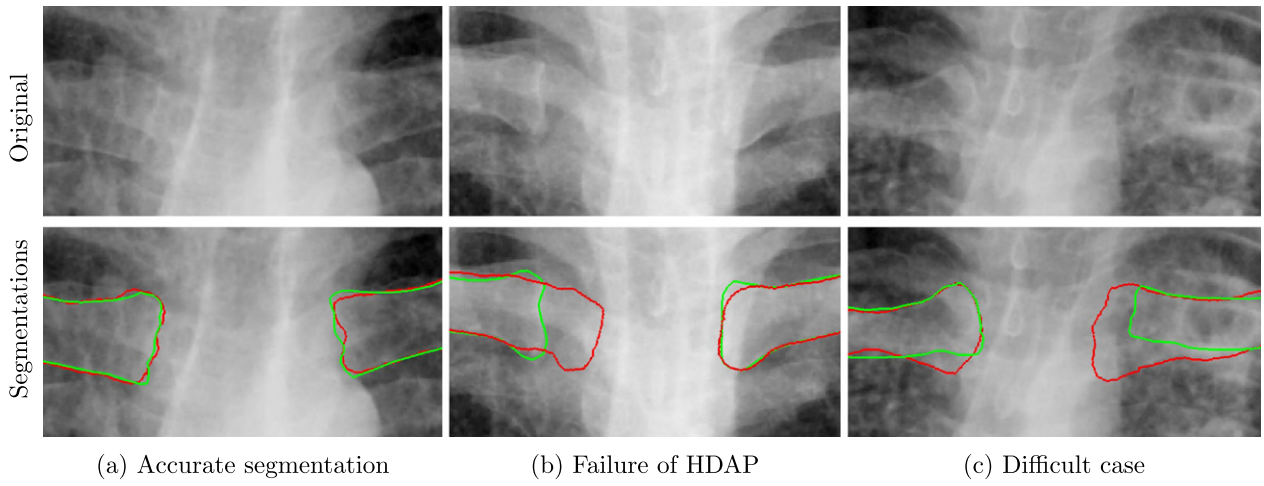


Fig. 9. Examples of head border segmentation. The green line is the outline provided by the reference standard, the red line from the best performing method HDAP: no shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

correction that HDAP can achieve is determined by the length of the profiles in optimal path space. This length was set to double the original distance between border and centerline (PMA) of the object, but was not optimized. If the initial pixel classification contained very large errors, HAP cannot improve over PC-postproc.

The slightly higher MCD on the upper clavicle border (see Table 4) can be explained by the presence of a double ridge which is often seen on the superior side of the clavicle but not on the inferior. For human observers it is relatively easy to choose the correct line if the clavicle border is traced from medial to lateral. HDAP sometimes chooses the wrong line when there is a more optimal path for the dynamic programming. This problem could be partly solved by adding a term to the cost function which prevents sudden changes in the smoothness of the optimal path. In practice, such corrections are typically not easily made without introducing errors in the segmentation of other images.

The error at the head of the clavicle is considerably larger than the errors at the diaphysis of the clavicle. In some cases HDAP can locate the border of the clavicle at the head very precisely (Fig. 9a); in other cases the algorithm fails even when a clearly visible edge can be seen (Fig. 9b). Finally in a considerable part of the cases the medial edge of the clavicle is hidden behind the mediastinal structures and both HDAP and the human observers fail to agree on where the border should be (Fig. 9c). How to improve the performance of automated segmentation methods in such difficult areas as the head of the clavicle is an open question. Human observers likely include much more context information in their reasoning. This context knowledge is provided by general knowledge, specific domain knowledge and their ability to reason about that knowledge and weigh alternative options. Encoding a comprehensive part of this knowledge in a computer system has so far proven very difficult.

6.4. Computation time

HDAP was implemented on a 3 GHz Intel Core 2 Duo with single threaded C++ code. The total computation time for one case is about 18 min. Most of the time is spent in the calculation of the features for the images (about 13 min), followed by the pixel classifiers (about 5 min). The combination of the information through HAP and HDAP costs only a few seconds of the total time. Running time is high but our implementation was not optimized. A more in-depth feature selection analysis could reduce the number of features that need to be calculated, e.g. Gaussian derivatives calculated at large scales are expensive to compute but mostly contribute to the localization of the clavicle and not to the exact determination of the border. A large speedup can also be obtained by using a recursive implementation of Gaussian derivatives (Deriche, 1993). Alternatively, efficiently computed filter families such as Haar wavelets could be used (Viola and Jones, 2001). Speedup for pixel classification can be obtained by parallelizing the classification or by adopting faster classifiers with similar performance. Also a number of specific strategies can be adopted to reduce the number of samples that need to be classified, such as multi-resolution schemes (van Ginneken et al., 2006) or sparse pixel classification (Dam and Loog, 2008).

6.5. Application to other tasks

While the proposed method is applied to the specific task of clavicle segmentation it can be generalized so that it can be applied to other segmentation tasks in (medical) images as well. Using a common framework (an optimal path formulation in our case) to combine global and local information sources is the core concept. The specific configuration and types of subcomponents to be used depend on the nature of the task. We expect that especially tasks where local cues can be ambiguous will benefit from an integration of global and local information.

In chest radiographs other tasks of interest are lung field segmentation and rib segmentation. In normal images an automatic lung field segmentation algorithm has been shown to perform as well as a human observer (van Ginneken et al., 2006). For chest radiographs containing (severe) abnormalities it is not known whether similarly good segmentation results can be obtained, but a hybrid algorithm such as HDAP might be a good approach there too. A variation of HDAP could also be applied to rib segmentation where typical errors made by local methods, such as pixel classification (Loog and van Ginneken, 2006b), might benefit from global methods (shape modeling) and vice versa. A direct application of ASM to the ribs is not possible though as the number of shape components (ribs projecting onto the lung field) are variable.

7. Conclusions

A new method (HDAP) to automatically segment the clavicles in chest radiographs has been presented. HDAP combines three segmentation frameworks: pixel classification applied in two stages and separately for the interior, the order and the head of

the clavicle, followed by active shape model segmentation and, finally, dynamic programming using an optimized cost function. The method is compared with a number of previously described state-of-the-art methods and simplified versions of the proposed method. The large database that was used for training and testing the method is made publicly available to facilitate future comparisons with other methods. Results were analyzed quantitatively with a number of standard measures and compared to two independent human observers.

The main conclusion is that a combination of several existing techniques (multiple pixel classifications, active shape modeling and dynamic programming) leads to an algorithm that handles ambiguous input (due to pathology, projected structures, etc.) well and significantly outperforms previously proposed methods. Dynamic programming is a convenient way to combine information from a number of algorithm components. Yet, the problem cannot be considered solved, as the errors of the automatic methods are larger than the inter-observer variability. Still, results might be sufficient, especially in normal images, to use the segmentation for subsequent steps such as clavicle suppression, measuring rotation of the rib cage and computer-aided detection of abnormalities in the lung apices.

8. Appendix

8.1. Public dataset and challenge

The release of the images and annotations is part of our ongoing effort to improve the quality and progress of medical image analysis research by enabling fair comparisons of algorithms through public datasets and challenges. The dataset can be found at the website of the challenge Chest Radiograph Anatomical Structure Segmentation (CRASS) (http://crass12.grand-challenge.org). The chest radiographs and manual outlines of the clavicles of the training set are provided. For the test set only the radiographs are provided and the outlines are kept secret. New segmentation results on the test set can be uploaded to the website and will be evaluated automatically. Submitted results are automatically evaluated using similar measures as reported in this paper. An automated report and ranking among other methods will be publicly available for each submitted result.

Acknowledgements

The authors gratefully acknowledge Dr. H. Ayles of the ZAMBART Project (University of Zambia, Lusaka, Zambia) and the Department of Clinical Research, London School of Hygiene and Tropical medicine (London, UK) for kindly providing the chest radiographs and T. Dubbelink, N. Snellen and K. Hengstler for outlining the clavicles.

This study was supported by the European and Developing Countries Clinical Trials Partnership (EDCTP), the Evaluation of multiple novel and emerging technologies for TB diagnosis, in smear-negative and HIV-infected persons, in high burden countries (TB-NEAT).

References

Suzuki, K., Abe, H., MacMahon, H., Doi, K., 2006. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). IEEE Trans. Med. Imag. 25, 406–416.

Loog, M., van Ginneken, B., 2006a. Bony structure suppression in chest radiographs. In: Computer Vision Approaches to Medical Image Analysis. Lect. Notes. Comput. Sci., vol. 4241, pp. 166–177.

Freedman, M.T., Benedict Lo, S.-C., Seibel, J.C., Bromley, C.M., 2011. Lung nodules: improved detection with software that suppresses the rib and clavicle on chest radiographs. Radiology 260, 265–273.

Li, F., Hara, T., Shiraishi, J., Engelmann, R., MacMahon, H., Doi, K., 2011. Improved detection of subtle lung nodules by use of chest radiographs with bone suppression imaging: receiver operating characteristic analysis with and without localization. AJR Am. J. Roentgenol. 196, W535–W541.

Quekel, L.G., Kessels, A.G., Goei, R., van Engelshoven, J.M., 1999. Miss rate of lung cancer on the chest radiograph in clinical practice. Chest 115, 720–724.

van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A., 2002. Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Trans. Med. Imag. 21, 139–149.

Yu, T., Luo, J., Ahuja, N., 2005. Shape regularized active contour using iterative global search and local optimization. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005), vol. 2, pp. 655–662.

van Ginneken, B., Stegmann, M.B., Loog, M., 2006. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med. Image Anal. 10, 19–40.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR Am. J. Roentgenol. 174, 71–74.

Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P., 2007. Minimal shape and intensity cost path segmentation. IEEE Trans. Med. Imag. 26, 1115–1129.

Simkó, G., Orbán, G., Máday, P., Horváth, G., 2009. Elimination of clavicle shadows to help automatic lung nodule detection on chest radiographs. In: Sloten, Jos, Verdonck, Pascal, Nyssen, Marc, Haueisen, Jens, Magjarevic, Ratko (Eds.), IFMBE Proceedings: 4th European Conference of the International Federation for Medical and Biological Engineering, vol. 22, pp. 488–491.

Cootes, T.F., Taylor, C.J., Cooper, D., Graham, J., 1995. Active shape models – their training and application. Comput. Vis. Image Underst. 61, 38–59.

Jain, A.K., Duin, R.P.W., Mao, J., 2000. Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. 22, 4–37.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. John Wiley and Sons, New York.

Florack, L.M.J., ter Haar Romeny, B.M., Viergever, M.A., Koenderink, J.J., 1996. The Gaussian scale-space paradigm and the multiscale local jet. Int. J. Comput. Vis. 18, 61–75.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: Wells, William, Colchester, Alan, Delp, Scott (Eds.), Med. Image Comput. Comput. Assist. Interv., Lect. Notes Comput. Sci., vol. 1496. Springer, Berlin/Heidelberg, pp. 130–137.

van Rikxoort, E.M., van Ginneken, B., Klik, M., Prokop, M., 2008. Supervised enhancement filters: application to fissure detection in chest CT scans. IEEE Trans. Med. Imag. 27, 1–10.

van Ginneken, B., 2006. Supervised probabilistic segmentation of pulmonary nodules in CT scans. In: Med. Image Comput. Comput. Assist. Interv. Lect. Notes Comput. Sci., vol. 4191, pp. 912–919.

Cootes, T.F., Taylor, C.J., 2001. Statistical models of appearance for computer vision. Technical report.

Loog, M., van Ginneken, B., 2006b. Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification. IEEE Trans. Med. Imag. 25, 602–611.

Vittitoe, N.F., Vargas-Voracek, R., Floyd Jr., C.E., 1999. Markov random field modeling in posteroanterior chest radiograph segmentation. Med. Phys. 26, 1670–1677.

Bellman, R.E., 1962. Applied Dynamic Programming. Princeton University Press.

Montanari, U., 1971. On the optimal detection of curves in noisy pictures. Commun. ACM 14, 335–345.

Blum, H., 1967. A transformation for extracting new descriptors of shape. Models for the perception of speech and visual form, 362–380.

Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y., 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. J. ACM 45, 891–923.

Pudil, P., Novovicova, J., Kittler, J., 1994. Floating search methods in feature selection. Pattern Recognit. Lett. 15, 1119–1125.

Gerig, G., Jomier, M., Chakos, M., 2001. Valmet: a new validation tool for assessing and improving 3D object segmentation. In: Med. Image. Comput. Comput. Assist. Interv. Lect. Notes. Comput. Sci., pp. 516–523.

McGill, R., Tukey, J.W., Larsen, W.A., 1978. Variations of box plots. Am. Stat. 32, 12–16.

Timp, S., Karssemeijer, N., 2004. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Med. Phys. 31, 958–971.

Cui, J., Sahiner, B., Chan, H., Nees, A., Paramagul, C., Hadjiiski, L.M., Zhou, C., Shi, J., 2009. A new automated method for the segmentation and characterization of breast masses on ultrasound images. Med. Phys. 36, 1553–1565.

van Rikxoort, E.M., de Hoop, B., Viergever, M.A., Prokop, M., van Ginneken, B., 2009. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Med. Phys. 36, 2934–2947.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imag. 23, 903–921.

Artaechevarria, X., Munoz-Barrutia, A., Oritz de Solórzano, C., 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imag..

Kuncheva, L.I., 2002. A theoretical study on six classifier fusion strategies. IEEE Trans. Pattern Anal. Mach. Intell. 24, 281–286.

Brejl, M., Sonka, M., 2000. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. IEEE Trans. Med. Imag. 19, 973–985.

Deriche, R., 1993. Recursively implementating the Gaussian and its derivatives. Rapport de, recherche.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: Proceedings IEEE Conf. on Computer Vision and, Pattern Recognition, vol. 1, pp. I-511–I-518.

Dam, E.B., Loog, M., 2008. Efficient segmentation by sparse pixel classification. IEEE Trans. Med. Imag. 27, 1525–1534.