Available online at www.sciencedirect.com
SCIENCE DIRECT®
Medical Image Analysis 8 (2004) 127–138
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Interactive segmentation of abdominal aortic aneurysms in CTA images

Marleen de Bruijne [a,b,*], Bram van Ginneken [a], Max A. Viergever [a], Wiro J. Niessen [a]

[a] *Image Sciences Institute, University Medical Center Utrecht, E01.335, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands*
[b] *Institute of Information and Computing Sciences, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands*

Received 12 February 2003; accepted 12 December 2003

Abstract

A model-based approach to interactive segmentation of abdominal aortic aneurysms from CTA data is presented. After manual delineation of the aneurysm sac in the first slice, the method automatically detects the contour in subsequent slices, using the result from the previous slice as a reference. If an obtained contour is not sufficiently accurate, the user can intervene and provide an additional manual reference contour.

The method is inspired by the active shape model (ASM) segmentation scheme (Cootes et al., 1995), in which a statistical shape model, derived from corresponding landmark points in manually labeled training images, is fitted to the image in an iterative manner. In our method, a shape model of the contours in two adjacent image slices is progressively fitted to the entire volume. The contour obtained in one slice thus constrains the possible shapes in the next slice. The optimal fit is determined on the basis of multi-resolution gray level models constructed from gray value patches sampled around each landmark. We propose to use the similarity of adjacent image slices for this gray level model, and compare these to single-slice features that are more generally used with ASM. The performance of various image features is evaluated in leave-one-out experiments on 23 data sets.

Features that use the similarity of adjacent image slices outperform measures based on single-slice features in all cases. The average number of slices in our datasets is 51, while on average eight manual initializations are required, which decreases operator segmentation time by a factor of 6.

Keywords: Active shape model; Image segmentation; Similarity measure; CT; Abdominal aortic aneurysm; Blood vessels

1. Introduction

An abdominal aortic aneurysm (AAA) is an enlargement of the abdominal aorta, resulting from weakened arterial walls. Once present, AAAs continue to enlarge and become increasingly susceptible to rupture, which usually results in death. Because of the risks associated with surgical AAA treatment, patients with a small AAA are often placed under frequent surveillance until the aneurysm becomes symptomatic or its diameter exceeds 5.5 cm. Worldwide, approximately 100,000 surgical interventions for AAA repair are performed each year, of which at present 30% are endovascular. After endovascular aneurysm repair, in which a synthetic graft is placed inside the aorta, the process of aneurysm shrinkage, ongoing aneurysmal disease, and damage or fatigue of graft material may result in leakage, graft migration, and kinking or buckling of the graft, which can subsequently cause rupture or occlusion. Careful and frequent patient follow-up will likely be needed for the life of the patient after endovascular aneurysm treatment (Fillinger, 1999).

It has been demonstrated that change of aneurysm volume is a good indicator for the risk of aneurysm rupture (Czermak et al., 2001; Pollock et al., 2002; Wever et al., 2000b). Currently, the gold standard for volume assessment is computed tomography angiography (CTA) with subsequent manual aneurysm delineation (Balm et al., 1996). The manual segmentation is a time-consuming process – it takes an experienced

* Corresponding author. Present address: IT University of Copenhagen, Glentevej 67-69, DK-2400, Copenhagen, Denmark.
E-mail address: marleen@itu.dk (M. de Bruijne).

1361-8415/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2004.01.001

operator around 30 min – and suffers from inter- and intra-operator variations. Wever et al. (2000a) reported averaged inter-operator volume errors of 8.3% and intra-operator errors, for two operators, of 3.2% and 5.8%.

To reduce analysis time and to increase reproducibility, automated segmentation would be of great value. However, AAA segmentation in CTA images is a difficult task, and automatic methods are not available. The boundary can be obscured by surrounding tissue of similar gray value, many neighboring structures induce strong edges in close proximity to the aneurysm wall, and the variable aneurysm radius in combination with partial volume effects makes the boundary hard to distinguish even in the absence of neighboring structures. In addition, the size and appearance of aneurysms and neighboring structures vary considerably between patients as well as in one patient over time. Fig. 1 shows some examples of CTA slices, illustrating the diversity of aneurysm and background appearance.

Most publications on computerized AAA segmentation have concentrated on lumen segmentation, either in pre-operative (Bulpitt and Berry, 1998; Fiebich et al., 1997; Giachetti et al., 2002; Magee et al., 2001; Pohle and Toennies, 2001; Rubin et al., 1998; Tek et al., 2001; Weese et al., 2001; Wink et al., 2000) or in post-operative CTA scans (de Bruijne et al., 2003). The more difficult problem of segmenting the aneurysm sac has less frequently been addressed. Wilson et al. (1999) claimed successful aneurysm sac segmentation using an active surface with the central lumen axis as a seed, but did not present an evaluation study. Subasic et al. (2002) segmented the aortic lumen using a three-dimensional level-set approach, but found that the method could not cope with the large regions lacking boundary evidence in delineating the thrombus. They solved this problem using a combination of thresholding, edge detection, morphological operations, and shape restrictions in the form of a distance map, favoring convex cross-sections. A level-set was evolved in the resulting binary image to ensure a smooth object. The average segmentation error per slice was 19.8%, which makes the method suitable for visualization purposes but not for accurate volume measurements.

We present and evaluate an interactive method for aneurysm sac segmentation. After manual segmentation of the first slice, the method automatically detects the contour in subsequent slices, using the result from previous slices as a reference. If an obtained contour is not correct, the user can intervene and provide an additional manual reference contour. The method relies on the fitting of a shape model to points with high correlation with the reference contour.

The method is inspired by the Active Shape Model (ASM) segmentation scheme as put forward by Cootes et al. (1995, 2002), which combines statistical knowledge of object shape and shape variation with local appearance models near object contours. ASMs have been successfully applied to various segmentation tasks in medical imaging (Behiels et al., 2002; Cootes et al., 1994; Duta and Sonka, 1998; Hamarneh and Gustavsson, 2000; Hill et al., 1993; Mitchell et al., 2001; Solloway et al., 1997; van Ginneken et al., 2002). Applications in three-dimensional images have either used a full three-dimensional model (Hill et al., 1993; Kohnen et al., 2002; Paulsen et al., 2002) or a two-dimensional model applied to image slices (Cootes et al., 1994; Mitchell et al., 2001). We have chosen a slice-by-slice approach, which has the advantage that it is easily integrated with the current manual segmentation scheme. Moreover, the high anisotropy of the CTA data makes in-slice gray value models more appropriate, and a three-dimensional model would require resampling to an equal number of slices for all aneurysms, while in our datasets an aneurysm is between 30 and 65 slices long.

Mitchell et al. (2001) proposed the use of a two-dimensional hybrid active shape model/active appearance model, fitted to several cardiac MR image slices independently. Cootes et al. (1994) reported successful slice-by-slice ASM segmentation of the ventricles in 3D MR datasets, using the fitted shape in one slice as the initialization in the next slice. We extend this scheme by modeling two successive image slices together. The shape obtained in the previous slice is not only used as an initialization but constrains the shape in the next slice, resulting in a segmentation where the transition

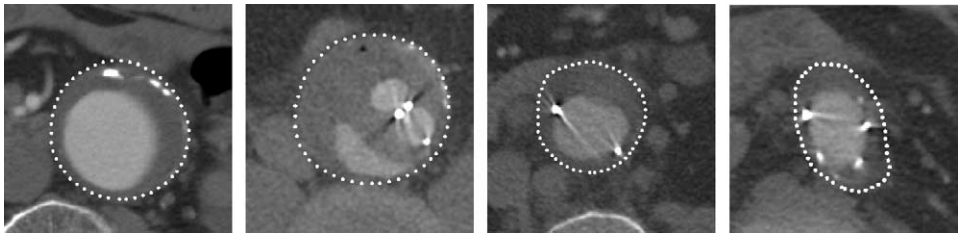

Fig. 1. Four CTA slices of different patients, with manually delineated aneurysms. The left image is taken before surgery, the others are post-operative. The interior of the contour contains thrombus, contrast enhanced blood in the lumen and possibly in endoleaks, calcification, and – post-operatively – metal graft markers and attachment systems. Neighboring structures are the spinal column, blood vessels and intestines. All CTA images in this paper have been contrast stretched for better visibility of the thrombus.

between slices is in accordance with the shapes in the training set.

Conventional ASMs use a linear model generated from gray value patches in training images to fit the shape model to the image. AAAs have such a diverse range of possible surrounding structures, which vary in location, shape, intensity and texture, that a linear model of gray level structure about the contours may not be able to find the true contour in a new image. An important indicator for the image structure in one image slice is the structure in the adjacent slices. We propose a modified ASM scheme in which optimal landmark positions are defined by maximum gray value correlation with adjacent slices rather than by correlation with the training data. In this study, several similarity measures are evaluated and compared to features that were shown to perform well in combination with ASMs (Behiels et al., 2002; Cootes and Taylor, 2002).

The shape model is briefly discussed in Section 2.1. Section 2.2 presents various features for aneurysm boundary localization, and Section 2.3 explains how the shape model is fitted to the image using these features. Section 3 describes a series of leave-one-out experiments on 23 datasets. Several robustness tests are presented in Sections 3.4 and 3.5. The performance of the interactive procedure is measured by the required amount of user interaction in a simulated interactive segmentation scheme, as presented in Section 3.6. Discussion and conclusions are given in Sections 4 and 5.

2. Methods

This section describes the segmentation scheme. Shape variations in the training set are described using a Point Distribution Model (PDM) (Cootes et al., 1992). The shape model is used to generate new shapes, similar to those found in the training set, which are fitted to the data using a model of local gray value structure.

2.1. Point distribution models

In PDMs, a statistical model of object shape and shape variation is derived from a set of s training examples. Each training example is described by a shape vector $\mathbf{x}$ containing the coordinates of n landmark points that correspond between shapes. Variations in the coordinates of these landmark points describe the variation in shape and pose across the training set.

To maximize the specificity of the model, the shapes can be aligned by rotation, translation and scaling, thus reducing non-linearities in the shape distribution. However, any relation between the object's shape and its pose or scale is then lost. Alternatively, one could align the shapes using only translation and rotation and construct a so-called size-and-shape model, in order to retain the relationship between shape and scale while removing the pose.

Principal component analysis (PCA) is applied to the (aligned) shape vectors. To this end, the mean shape $\overline{\mathbf{x}}$, the covariance matrix $\mathbf{S}$, and the eigensystem of $\mathbf{S}$ are computed. The eigenvectors $\boldsymbol{\phi}_i$ of $\mathbf{S}$ provide the modes of shape variation present in the data. The eigenvectors corresponding to the largest eigenvalues λ_i account for the largest variation; a small number of modes usually explains most of the variation. Each shape $\mathbf{x}$ in the set can then be approximated by

$$\mathbf{x} \approx \overline{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}, \tag{1}$$

where $\boldsymbol{\Phi}$ consists of the eigenvectors corresponding to the t largest eigenvalues, $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1|\boldsymbol{\phi}_2|\ldots|\boldsymbol{\phi}_t)$, and $\mathbf{b}$ is the model parameter vector that weighs the contribution of each of the modes.

The total variance in the dataset is given by $\sum \lambda_i$. The number t of modes in the model is often chosen such that the model captures a certain proportion f_v of the total variance observed:

$$\sum_{i=1}^{t} \lambda_i \geqslant f_v \sum_{i=1}^{2n} \lambda_i. \tag{2}$$

2.2. Gray value model

A rectangular gray value patch of $k_n \times k_t$ pixels is sampled symmetrically around each landmark, with the length k_n normal to the contour and the width k_t tangential to the contour. The patches are sampled at multiple resolutions, to enable coarse-to-fine fitting of the model. The finest resolution uses the original image and a sample spacing of 1 voxel, the next resolution is the image convolved with a Gaussian kernel of width one and sampled with a spacing of 2 voxels, and subsequent resolutions are obtained by doubling both the kernel width and the sample spacing.

For a successful segmentation procedure, the ability of an image feature to distinguish between correct and false boundary positions is crucial. We evaluate the performance of the following features.

2.2.1. Edges

An object edge likely corresponds to image positions with a strong intensity gradient. We applied the gradient magnitude in the direction normal to the contour, GM.

2.2.2. Linear model of training data

The boundary appearance can be learned from a training set. In the original ASM formulation, Cootes and Taylor (2002) use the Mahalanobis distance from the sampled gray value patch $\mathbf{g}_s$ to the mean of training patches $\overline{\mathbf{g}}$ as a boundary measure:

$$\mathrm{MAH}(\mathbf{g}_s) = (\mathbf{g}_s - \overline{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \overline{\mathbf{g}}). \tag{3}$$

The effect of global intensity changes can be reduced by sampling the first derivative and normalizing the patch. We have applied MAH to intensity (I) and first order derivative (I') patches, and the normalized versions thereof. The patches are normalized such that the sum of the absolute intensity values equals 1.

2.2.3. Linear model of training data combined with slice similarity

The linear model of training patches that is used in ASM can be combined with information from adjacent slices. The Mahalanobis distance from the sample patch $\mathbf{g}_s$ to the reference patch $\mathbf{g}_r$ expresses the difference between the two patches, taking the covariance across the training set into account:

$$\mathrm{MAHR}(\mathbf{g}_s) = (\mathbf{g}_s - \mathbf{g}_r)^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \mathbf{g}_r). \quad (4)$$

The reference patch is sampled at the corresponding landmark in the adjacent slice. In the experiments, MAHR is applied to (normalized) intensity and first order derivative patches.

2.2.4. Slice similarity

Several features based only on similarity between adjacent slices are considered:

- Sum of squares of intensity differences between sample and reference

$$\mathrm{SD} = \sum_{x=0}^{k_n \times k_t} (I_s(x) - I_r(x))^2, \quad (5)$$

  where I_s is the image intensity in the sample patch and I_r the intensity in the reference patch.
- Sum of absolute intensity differences between sample and reference

$$\mathrm{AD} = \sum_{x=0}^{k_n \times k_t} |I_s(x) - I_r(x)|. \quad (6)$$

- Normalised cross-correlation between sample and reference

$$\mathrm{NCC} = \frac{\sum_{x=0}^{k_n \times k_t} I_s(x) \cdot I_r(x)}{\sqrt{\sum_{x=0}^{k_n \times k_t} I_s(x)^2 \cdot \sum_{x=0}^{k_n \times k_t} I_r(x)^2}}. \quad (7)$$

- Correlation coefficient between sample and reference

$$\mathrm{CC} = \frac{\sum_{x=0}^{k_n \times k_t} (I_s(x) - \overline{I_s}) \cdot (I_r(x) - \overline{I_r})}{\sqrt{\sum_{x=0}^{k_n \times k_t} (I_s(x) - \overline{I_s})^2 \cdot \sum_{x=0}^{k_n \times k_t} (I_r(x) - \overline{I_r})^2}}. \quad (8)$$

  All slice similarity features are tested on raw intensity patches.

GM, NCC and CC are maximized, all other measures are minimized.

2.3. Fitting procedure

Starting with one manually drawn contour, the segmentation proceeds through the dataset, fitting the two-slice model to all pairs of adjacent slices. Each time the result from the previous slice is used as a reference for determining the contour in the new slice. From coarse to fine resolution, the model is fitted to a pair of slices in an iterative procedure. The process initializes with the shape from the reference slice, in both slices. For all landmarks in the new slice, n_s possible new positions are evaluated along the line perpendicular to the contour, on both sides. The distance between the positions to evaluate depends on the level of resolution and is equal to the sample spacing of the corresponding gray level model: at the finest resolution the distance is 1 voxel, at each subsequent resolution it is doubled. The optimal new position is determined by one of the features that were introduced in the previous section. The shape model parameters $\mathbf{b}$ that minimize the squared distance between the landmarks and the optimal positions are computed using

$$\mathbf{b} = \mathbf{\Phi}^T (\mathbf{x} - \overline{\mathbf{x}}). \quad (9)$$

Hard limits are applied to constrain $\mathbf{b}$ to plausible values:

$$|b_i| \leqslant f_c \sqrt{\lambda_i}, \quad (10)$$

in which the parameter f_c is typically chosen between 2 and 3. The landmarks of the new slice are moved to their new positions according to the model parameters, while the landmarks in the reference slice are kept fixed. This process is repeated a fixed number of N times, whereupon it is repeated at the next level of resolution.

3. Experiments

A series of leave-one-out experiments is performed on 23 CTA images.

3.1. Data

All experiments are carried out on 23 routinely acquired CTA images from 23 different patients. Of these, 2 are pre-operative scans and all others are taken after endovascular stenting, 9 directly after surgery and 12 at follow-up ranging from 1 to 12 months. The scan resolution is $0.488 \times 0.488 \times 2.0$ mm. Each image consists of circa 125 slices of 512×512 voxels, of which 34–63 slices contain aneurysmal tissue. The total number of slices to be segmented is 1175.

3.2. Landmarking and alignment

A PDM is built of all pairs of adjacent slices. The original CT-slices, which are perpendicular to the body axis and therefore always give approximately perpendicular cross-sectional views of the aorta, are used. Contours are drawn manually by an expert between the

first and last slice in which thrombus is visible. In the absence of well-defined anatomical landmark points, landmarks are equidistantly placed along the object contours. The starting point of a contour is the posterior point with the same x-coordinate as the center of mass.

If the relation between position, size, orientation and shape is not important, the shapes can be aligned with respect to these transformations prior to model construction, in order to increase specificity and compactness of the model. To make the model even more compact, the slices could be aligned separately so that translation, rotation and scaling of a shape with respect to its reference is removed from the model. However, we keep the transformations from slice to slice in the model, as there is a correlation between these transformations and the shape. For instance, contours in the top and bottom of the aneurysm, where the largest size changes between slices occur, are usually more elliptic than in the rest of the aneurysm.

To investigate the relations between size, orientation, and shape, the shapes are aligned using Procrustes analysis and a new PDM is constructed. Fig. 2 shows scatter plots of the first principal components versus the scale and the rotation angle. We find no evidence of a relation between shape and size, but there is a clear correlation between the second mode of variation and the orientation. In the remainder, the shapes are aligned only by translation, such that the centers of the reference contours coincide.

Fig. 3 shows the distribution of shape vectors by pairwise scatter plots of the first 10 principal components. From this, the assumption that all shapes that lie within a box of $2.5\sqrt{\lambda_i}$ are valid model instances seems reasonable.

Fig. 4 shows the first eight modes of shape variation of the two-slice model trained with all 23 datasets. The largest variation in the shapes in the training set, the first mode, is a size difference. The second and third mode show that a strong elongation in x direction is correlated with a slightly larger upper contour, while an elongation in the y direction corresponds to a smaller upper contour. These are the types of shapes that are found in the bottom and the top of the aneurysm. The next mode describes a less pronounced deviation from the mean shape, where both slices deviate in the same way. Only from the fifth mode, a larger difference between the two adjacent slices is observed, with first a shape difference, then translations in the x and y directions, followed by a difference in size.

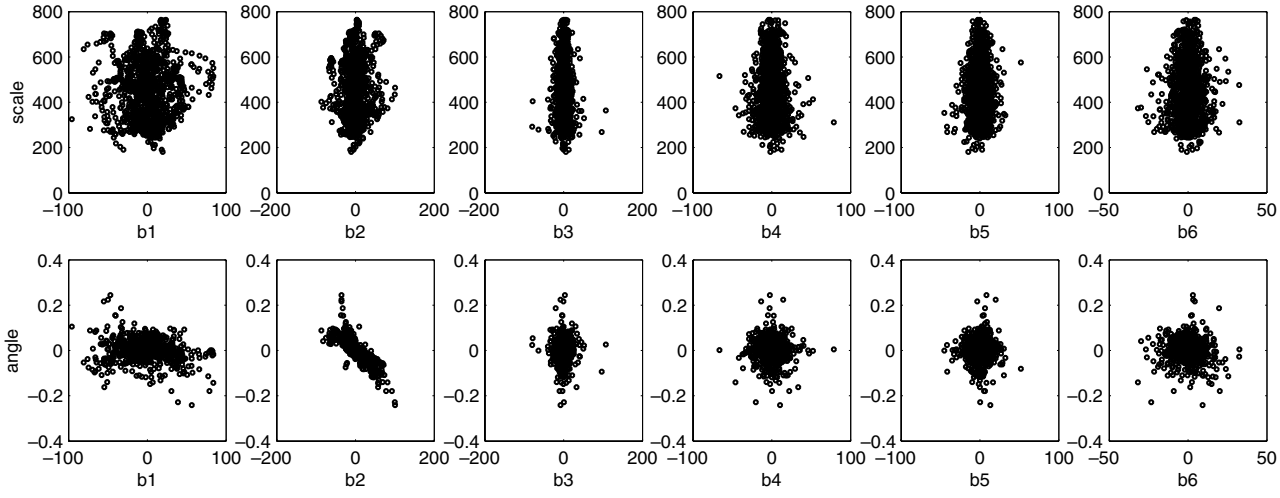

Fig. 2. Relation between the first six modes of shape variation of an aligned model and the global transformations. Top row: scale. Bottom row: rotation angle.

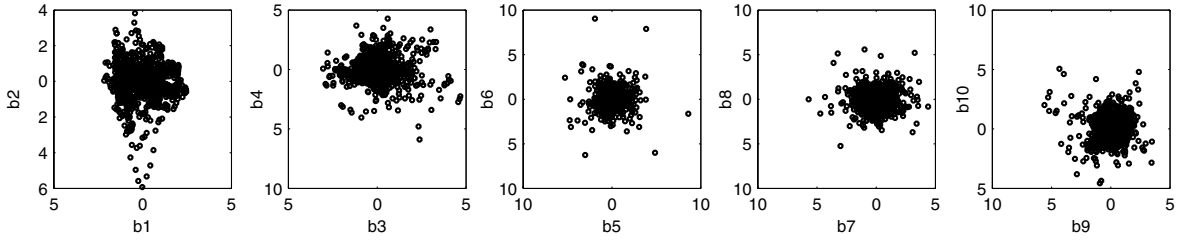

Fig. 3. Pairwise plots of the first 10 principal components. Units are given in $\sqrt{\lambda_i}$.

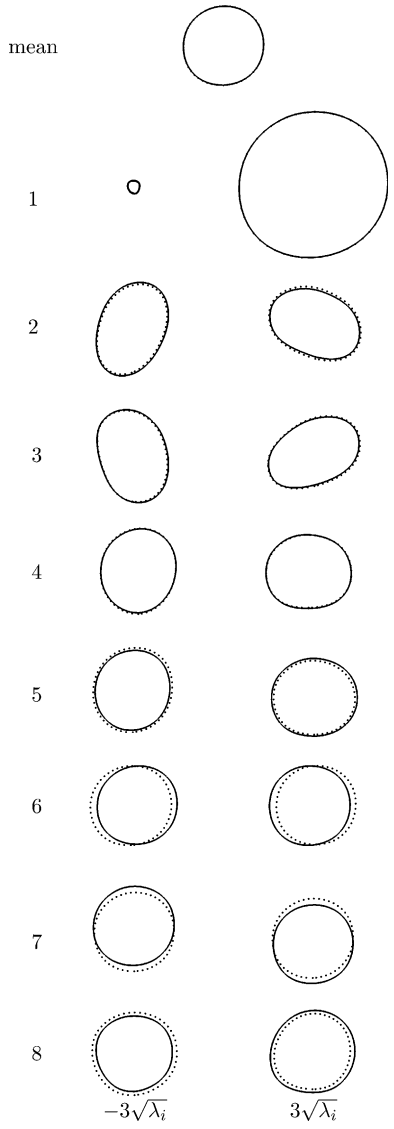

Fig. 4. The effect of varying each of the first eight shape parameters individually in a two-slice model built from all 23 datasets. The continuous contour denotes the lower slice, the dotted contour the upper slice.

3.3. Parameter settings

Table 1 lists the parameter values that are used in all experiments, unless mentioned otherwise. Since the shapes are not size normalized, the proportion of the variance to retain in the model, f_v, is chosen relatively large (0.99).

3.4. Shape model evaluation

The amount of variance explained by the model is plotted as a function of the number of modes in Fig. 5. Over 99% of the total amount of variation present in the training set is described by the first 11 modes.

The validity of the shape model and the applied constraints are tested by fitting the model directly to the contours that were manually drawn, which gives an upper bound for the accuracy that can be obtained when the model is fitted to new image data. Fig. 6 shows the

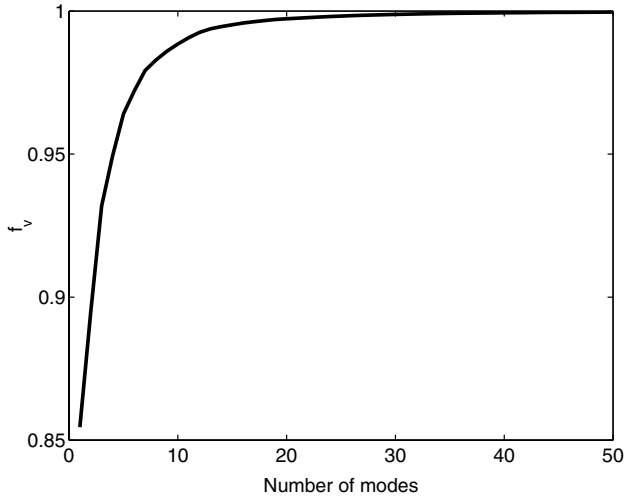

Fig. 5. The amount of variance explained by the model as a function of the number of modes contained in the model.

Table 1
Parameters of the segmentation scheme

Shape model (Section 2.1)		
n	50	Number of landmark points per slice
f_v	0.99	Part of shape variance to be explained by the model, controlling the number of modes t (Eq. (2))
f_c	2.5	Bound on eigenvalues λ_i (Eq. (10))
Gray value model (Section 2.2)		
k_n	7	Patch length
k_t	1	Patch width
Fitting algorithm (Section 2.3)		
n_s	2	Number of new landmark positions to examine per iteration on either side of the current position
L	3	Number of resolution levels
N	5	Number of iterations per resolution level

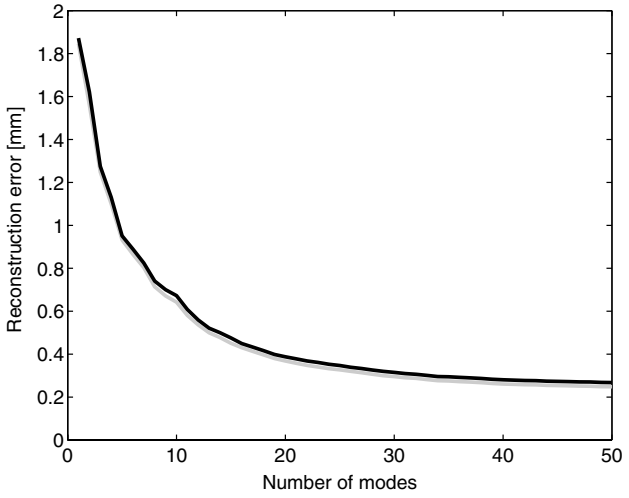

Fig. 6. Root mean squared error as a function of the number of modes, for a model built of all datasets (gray line) and the leave-one-out case (black line).

root mean squared error in millimeters as a function of the number of modes retained. The averaged leave-one-out reconstruction error, where a model built of 22 training shapes is fitted to the remaining shape, almost coincides with the reconstruction error for a model built of all 23 shapes, indicating that sufficient training data are available to represent all shapes.

3.5. Feature evaluation

The following experiments compare boundary localization performance of the different features that were introduced in Section 2.2. First, feature performance is evaluated independent of the rest of the segmentation process. Starting from the landmarks on the manually drawn contour, the optimal landmark positions according to the feature are determined, for varying n_s. The distances from these points to the landmark positions provided by the manual tracings are measured. The resulting root mean squared distances as a function of the size of the search region are shown in Fig. 7. Features using patches of length $k_n = 7$ and width $k_t = 1$ and 5, with a sample spacing of 1 or 4 voxels, are considered. Initial experiments revealed that smaller errors are obtained with the Mahalanobis distance based measures when the patches are normalized. For the sake of clarity of the graph, we here only show the results on normalized intensity and derivative patches, but the final results on non-normalized patches are given in Section 3.6 for comparison. The dashed black line starting at the origin corresponds to the expectation for random landmark selection.

At a small scale and in a small search region, all features perform similarly, but at larger scale or in a larger region the errors obtained using slice similarity measures are much smaller than those obtained using the gradient magnitude or the linear statistical model. Both Mahalanobis distance based features perform much better on derivative than on intensity patches. The Mahalanobis distance to the average intensity patch is even worse than random landmark selection. Among the slice similarity features, AD and SD yield better results than NCC and CC. The robustness of the slice similarity features increases with increasing scale, while the robustness of the Mahalanobis distance and gradient measures decreases. The same holds for enlarging the patch width.

The segmentation process is less sensitive to random outliers than to several neighboring landmarks that are consistently placed at a wrong position, which can be the case when a neighboring structure produces misleading boundary evidence. The result of fitting the shape model to the optimal points of the previous test is given in Fig. 8. The minimum error has increased, since it cannot be smaller than the reconstruction error from Fig. 6, and all maximum errors have decreased. Again, both Mahalanobis distance based features perform better on derivative than on intensity patches. The Mahalanobis distance to the reference patch gives errors similar to those obtained using the other slice similarity measures. GM is one of the best features if the search regions is within 3 voxels of the true contour, but errors quickly increase with larger search areas. At the lowest resolution, NCC and CC perform better than AD and SD, but at higher resolutions AD gives the smallest errors. Enlarging the patch width does not increase the robustness for the slice similarity measures as was the case before fitting the shape model, and robustness decreases for all Mahalanobis distance based measures. In the subsequent experiments we use $k_t = 1$.

3.6. Simulated interactive segmentation

To evaluate all features evenly in an interactive segmentation setting, we simulated observer interaction using manually segmented datasets. Starting with the bottom contour of the manual segmentation as a reference, successive slices are segmented. After each segmented contour, the similarity to the manually drawn shape is computed. If the similarity is sufficient, the segmentation proceeds, otherwise, a new contour from the available manual segmentation is added, so as to simulate operator intervention. The performance is measured by the number of manual initializations needed to segment an entire volume.

The similarity of the two contours is described in terms of the distance from each landmark to the contour drawn manually. Since the main clinical objective of AAA segmentation is volume measurement, the mean signed distance, which is proportional to the error in the measured area in a slice, should be small (1 mm in our experiments). To ensure similarly shaped and positioned

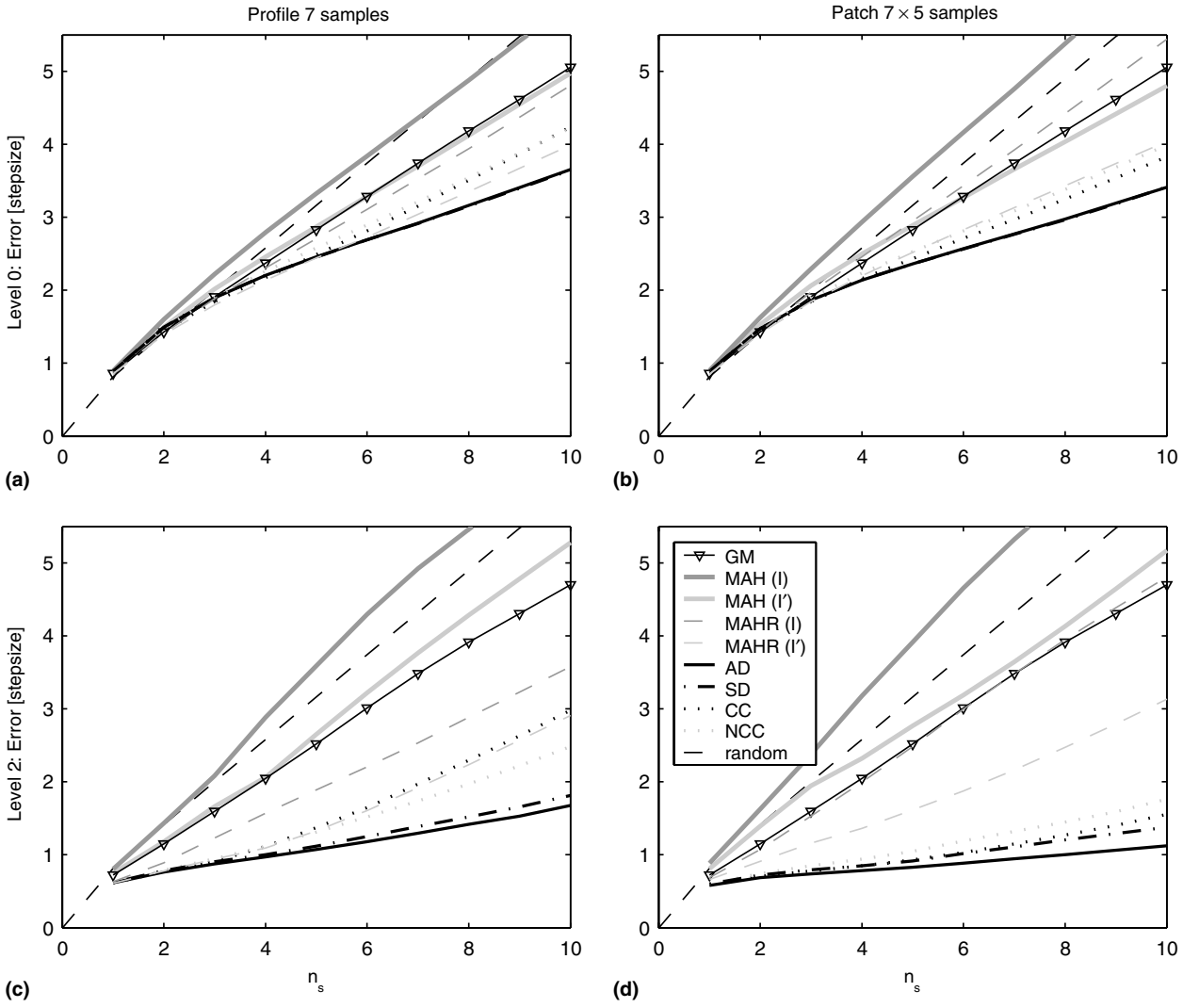

Fig. 7. Root mean squared error of optimal landmark position detection as a function of the number of points evaluated on either side of the contour, for (a) a patch of 7×1 voxels, sampled at scale 0; (b) a patch of 7×5 voxels, sampled at scale 0; (c) a patch of 7×1 voxels, sampled at scale 2; (d) a patch of 7×5 voxels, sampled at scale 2. The units are step sizes in the position evaluation procedure, i.e. 1 voxel (0.488 mm) for the profiles sampled at scale 0 and 4 voxels at scale 2.

contours the mean and maximum absolute distance are constrained as well, to 2 and 4 mm, respectively. Four randomly chosen examples of segmentations fulfilling these requirements are shown in Fig. 9.

Fig. 10 shows, for each feature and each dataset, the average number of slices that was correctly segmented using one initialization. Table 2 lists these results averaged over all datasets. In all cases, the smallest number of interactions was needed when one of the slice similarity measures was used. The average number of automatically obtained contours was largest when using AD, but the difference with SD was not significant ($p = 0.75$ in a paired t-test). Both AD and SD performed significantly better than all other tested features ($p = 0.06$ as compared to CC, $p < 0.00001$ for all other features). If MAHR was used, thus combining the information from the reference slice with the statistical model, more user interactions were needed, and using only the model (MAH) again increased the number of interactions ($p = 0.003$). Overall, Mahalanobis distance based measures perform best on normalized first order derivative patches. A segmentation process based on MAH(I') on normalized patches needs less re-initializations than one based on GM ($p = 0.011$).

3.7. Expert supervised interactive segmentation

In the simulated experiments the manual tracings are used as the undisputed gold standard. However, inter-observer errors in aneurysm volume measurements are as high as 8.3% (Wever et al., 2000a). In case of clear image evidence the automatic segmentation should indeed be similar to the manual contour, but when no boundary is visible the criteria of Section 3.6 may be relaxed.

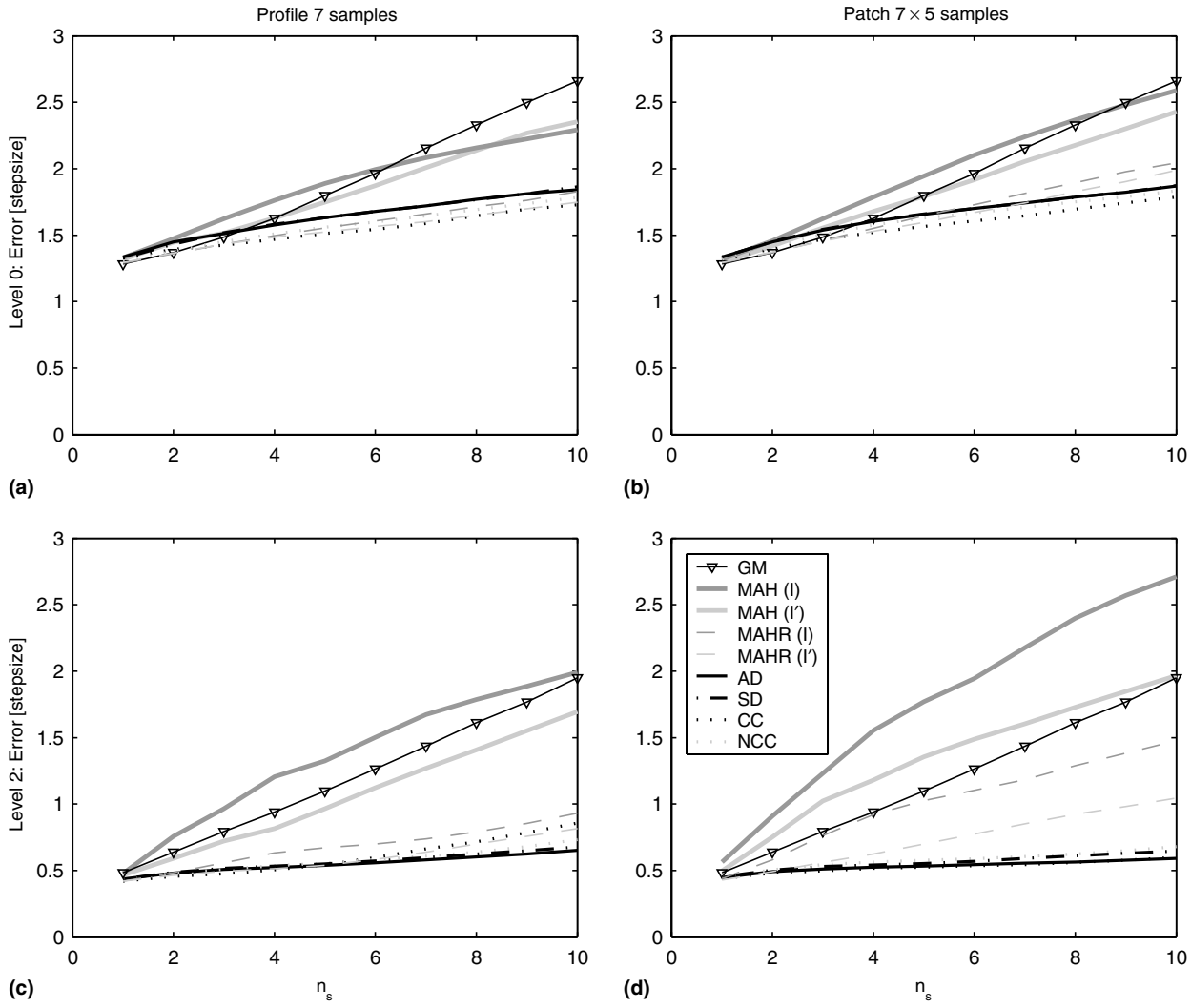

Fig. 8. Root mean squared error of landmark positions as a function of the number of points evaluated on either side of the contour, after fitting the shape model to the optimal positions, for (a) a patch of 7×1 voxels, sampled at scale 0; (b) a patch of 7×5 voxels, sampled at scale 0; (c) a patch of 7×1 voxels, sampled at scale 2; (d) a patch of 7×5 voxels, sampled at scale 2. The units are step sizes in the position evaluation procedure, i.e. 1 voxel (0.488 mm) for the profiles sampled at scale 0 and 4 voxels at scale 2.

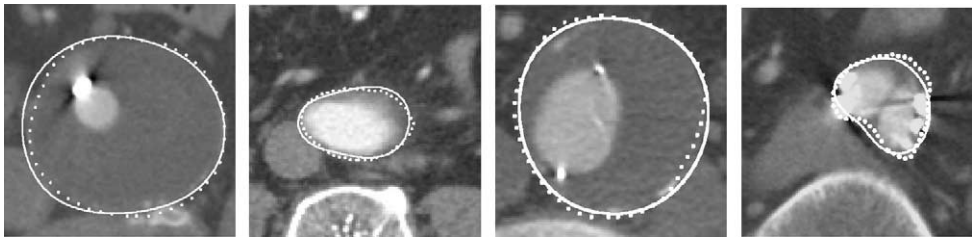

Fig. 9. Examples of accepted segmentations. The continuous line depicts the contour that was obtained automatically; the dotted line is the manual segmentation.

To get an idea of the true number of interactions that will be needed for segmentation, segmentation with feature AD was repeated, but this time a human observer decided whether re-initialization was needed. This resulted in a smaller number of interactions, especially in the top and bottom part of the aneurysm where the boundary is hard to distinguish owing to partial volume effects. The number of interactions used is given in Fig. 10 and Table 2. This number may still be an overestimation of the true number of interactions required, since the observer assesses segmentations that are based on reference contours drawn by another observer.

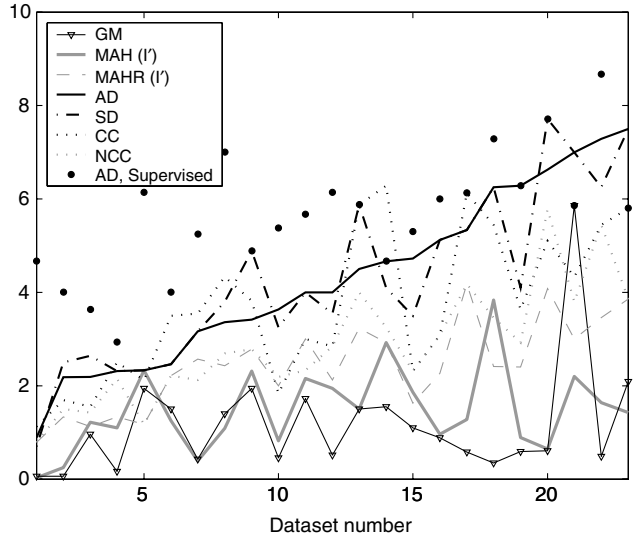

Fig. 10. Average number of correct contours obtained with one initialization, shown for all datasets.

Table 2
Average number of correct contours obtained using one initialization

		Normalized
GM	1.82	
MAH(I)	1.32	1.16
MAH(I')	1.11	2.25
MAHR(I)	2.86	2.96
MAHR(I')	2.61	3.23
AD	4.76	
SD	4.72	
CC	4.24	
NCC	3.59	
AD, Supervised	6.39	

The resulting segmentations had an average volume overlap of 95.8% (minimum 93.0%) with the manual segmentation. The average relative volume difference is 1.5% (maximum 3.9%), well below the 3.5% average intra-observer relative volume difference that was measured in a study on comparable datasets (Wever et al., 2000a).

4. Discussion

The presented method uses a shape model from two adjacent slices for segmenting a three-dimensional structure, while a full three-dimensional approach – allowing gray value modeling perpendicular to the object's surface – could be more appropriate in some applications. In the case of CTA images, which are in general highly anisotropic (in our images the voxels are over four times larger in the z-direction), we do not expect a significant improvement in boundary localization if the profiles would be sampled in three dimensions. An advantage of the two-slice approach is that it can easily be integrated in the current clinical practice of manual slice-by-slice segmentation. The use of a three-dimensional model would require a much larger number of training shapes, and all datasets would have to be described by an equal number of landmarks. The latter is not desirable for objects that can vary largely in length. Moreover, the use of slice similarity features in combination with the current ASM optimization scheme would be computationally infeasible for a model consisting of a large number of slices. Performance of the presented method may be improved, whilst avoiding the drawbacks of a complete three-dimensional approach, by adding one or more extra slices to the model.

It was shown that the original ASM scheme that uses local intensity models does not perform well in AAA segmentation. One of the shortcomings of the gray value model that is used in ASMs is that only the appearance of the correct boundary is learned from the training set, wherefore it cannot distinguish the true contour from other contours with similar gray values. Another problem is the underlying assumption of a normal intensity distribution. In the presence of distinct background structures of varying shape and brightness this assumption does not hold. The use of a non-linear gray value model may be more suitable. In this work we effectively solved this problem by using gray value patches from the reference slice.

A consequence of using slice similarity measures is that errors propagate through the dataset and therefore performance deteriorates with distance from the reference contour. We previously showed that a refinement step, in which the model is allowed to adjust to nearby edges, can increase contour tracking robustness (de Bruijne et al., 2002) in the middle part of the aneurysm. However, the model can then also fit to wrong edges. Fig. 8 shows that the errors quickly become large when fitting on edge strength. In segmentation of the entire aneurysm, including the more difficult top and bottom part, we have not observed a significant improvement when such a refinement was applied.

The proposed method could obtain accurate segmentations while only in one of six slices manual intervention is needed. Currently, the user must redraw the entire contour, but more sophisticated methods of user interaction can easily be incorporated. Since errors often occur at a small part of the contour while the larger part is correct, dragging one landmark point to the correct boundary and iteratively fitting the model to this point may be sufficient to obtain a correct contour (van Ginneken et al., 2003).

We have evaluated the application of this method to AAA segmentation, but the idea of propagating a two-slice shape model based on slice similarity measures can be used in other areas as well. Especially time-sequences and anisotropic medical images form a promising application. The dependency on image slicing and orien-

tation restricts the applicability to segmentation tasks where the shape change between slices is not too large and the object is always imaged in approximately the same direction. However, many medical images are acquired with fixed scan protocols which satisfy these conditions. In case the object axis is known, for instance through a – more easily automated – vessel lumen segmentation (Bulpitt and Berry, 1998; de Bruijne et al., 2003; Fiebich et al., 1997; Giachetti et al., 2002; Magee et al., 2001; Pohle and Toennies, 2001; Subasic et al., 2002; Tek et al., 2001; Weese et al., 2001; Wilson et al., 1999; Wink et al., 2000), or by manual definition, the proposed method could also be used for segmentation in reformatted slices orthogonal to that axis.

5. Conclusions

A new approach to the semi-automatic delineation of abdominal aortic aneurysms, based on shape model fitting in sequential slices, is reported. In a simulated interactive segmentation process, the benefit of using slice correlation in shape model fitting was demonstrated. The best performing features for aneurysm boundary localization are the absolute and squared difference with a reference image (AD and SD). Using a segmentation process steered by slice correlation, on average 3.76 slices could be segmented following one manual initialization, with mean signed error less than 1 mm, mean absolute error less than 2 mm and maximum absolute error at most 4 mm. An expert guided segmentation experiment showed that the simulated interaction slightly overestimates the number of re-initializations required.

Using the proposed method and the optimal slice correlation feature, the time required for expert segmentation may be reduced by a factor of 6.

Acknowledgements

This research was funded by the Netherlands Organization for Scientific Research (NWO).

We would like to thank our colleagues M. Prinssen and J.D. Blankensteijn from the Department of Vascular Surgery for providing the datasets and expert segmentations.

References

Balm, R., Kaatee, R., Blankensteijn, J., Mali, W., Eikelboom, B., 1996. CT-Angiography of abdominal aortic aneurysms after transfemoral endovascular aneurysm management. European Journal of Vascular and Endovascular Surgery 12 (2), 182–188.

Behiels, G., Maes, F., Vandermeulen, D., Suetens, P., 2002. Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models. Medical Image Analysis 6 (1), 47–62.

Bulpitt, A., Berry, E., 1998. Spiral CT of abdominal aneurysms: comparison of segmentation with an automatic 3D deformable model and interactive segmentation. In: Hanson, K. (Ed.), Medical Imaging: Image Processing. In: Proceedings of SPIE, vol. 3338. SPIE Press, pp. 938–946.

Cootes, T., Taylor, C., 2002. Statistical models of appearance for computer vision, Tech. rep., University of Manchester.

Cootes, T., Cooper, D., Taylor, C., Graham, J., 1992. Training models of shape from sets of examples. In: Hogg, D., Boyle, R. (Eds.), Proceedings of the British Machine Vision Conference. Springer, Berlin, pp. 9–19.

Cootes, T., Hill, A., Taylor, C., Haslam, J., 1994. The use of active shape models for locating structures in medical images. Image and Vision Computing 12 (6), 355–366.

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

Czermak, B., Fraedrich, G., Schocke, M., Steingruber, I., Waldenberger, P., Perkmann, R., Rieger, M., Jaschke, W., 2001. Serial CT volume measurements after endovascular aortic aneurysm repair. Journal of Endovascular Therapy 8 (4), 380–389.

de Bruijne, M., van Ginneken, B., Niessen, W., Maintz, J., Viergever, M., 2002. Active shape model based segmentation of abdominal aortic aneurysms in CTA images. In: Sonka, M., Fitzpatrick, M. (Eds.), Medical Imaging: Image Processing. Proceedings of SPIE, vol. 4684. SPIE Press, pp. 463–474.

de Bruijne, M., Niessen, W., Maintz, J., Viergever, M., 2003. Localization and segmentation of aortic endografts using marker detection. IEEE Transactions on Medical Imaging 22 (4), 473–482.

Duta, N., Sonka, M., 1998. Segmentation and interpretation of MR brain images: An improved active shape model. IEEE Transactions on Medical Imaging 17 (6), 1049–1067.

Fiebich, M., Tomiak, M., Engelmann, R., McGilland, J., Hoffman, K., 1997. Computer assisted diagnosis in CT angiography of abdominal aortic aneurysms. In: Hanson, K. (Ed.), Medical Imaging: Image Processing. Proceedings of SPIE, vol. 3034. SPIE Press, pp. 86–94.

Fillinger, M., 1999. Postoperative imaging after endovascular AAA repair. Seminars in Vascular Surgery 12 (4), 327–338.

Giachetti, A., Tuveri, M., Zanetti, G., 2002. Reconstruction and web distribution of measurable arterial models. Medical Image Analysis 7 (1), 79–93.

Hamarneh, G., Gustavsson, T., 2000. Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images. In: Proceedings IEEE Computers in Cardiology, vol. 27, pp. 115–118.

Hill, A., Thornham, A., Taylor, C., 1993. Model-based interpretation of 3D medical images. In: Illingworth, J. (Ed.), Proceedings of the British Machine Vision Conference, The British Machine Vision Association, pp. 339–348.

Kohnen, M., Mahnken, A., Kesten, J., Koeppel, E., Günther, R., Wein, B., 2002. A three dimensional knowledge based surface model for segmentation of organic structures. In: Sonka, M., Fitzpatrick, M. (Eds.), Medical Imaging: Image Processing. In: Proceedings of SPIE, vol. 4684. SPIE Press, pp. 485–494.

Magee, D., Bulpitt, A., Berry, E., 2001. Combining 3D deformable models and level set methods for the segmentation of abdominal aortic aneurysms. In: Cootes, T., Taylor, C. (Eds.), Proceedings of the British Machine Vision Conference, The British Machine Vision Association, pp. 119–126.

Mitchell, S., Lelieveldt, B., van der Geest, R., Bosch, H., Reiber, J., Sonka, M., 2001. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. IEEE Transactions on Medical Imaging 20 (5), 415–423.

Paulsen, R., Larsen, R., Laugesen, S., Nielsen, C., Ersbøll, B., 2002. Building and testing a statistical shape model of the human ear canal. In: Dohi, T., Kikinis, R. (Eds.), Medical Imaging Computing & Computer-Assisted Intervention. Lecture Notes in Computer Science, vol. 2489. Springer, Berlin, pp. 373–380.

Pohle, R., Toennies, K., 2001. A new approach for model-based adaptive region growing in medical image analysis. In: Skarbek, W. (Ed.), Computer Analysis of Images and Patterns. Lecture Notes in Computer Science, vol. 2124. Springer, Berlin, pp. 238–246.

Pollock, J., Travis, S., Whitaker, S., Davidson, I., Gregson, R., Hopkinson, B., Wenham, P., MacSweeney, S., 2002. Endovascular AAA repair: Classification of aneurysm sac volumetric change using spiral computed tomographic angiography, Journal of Endovascular Therapy 9 (2).

Rubin, G., Paik, D., Johnston, P., Napel, S., 1998. Measurement of the aorta and its branches with helical CT. Radiology 206 (3), 823–829.

Solloway, S., Hutchinson, C., Waterton, J., Taylor, C., 1997. The use of active shape models for making thickness measurements of articular cartilage from MR images. Magnetic Resonance in Medicine 37 (6), 943–952.

Subasic, M., Loncaric, S., Sorantin, E., 2002. 3D image analysis of abdominal aortic aneurysm. In: Sonka, M., Fitzpatrick, M. (Eds.), Medical Imaging: Image Processing. Proceedings of SPIE, vol. 4684. SPIE Press, pp. 1681–1689.

Tek, H., Comaniciu, D., Williams, J., 2001. Vessel detection by mean shift-based ray propagation. In: Staib, L., Rangarajan, A. (Eds.), IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. IEEE Computer Society Press.

van Ginneken, B., Frangi, A., Staal, J., ter Haar Romeny, B., Viergever, M., 2002. Active shape model segmentation with optimal features. IEEE Transactions on Medical Imaging 21 (8), 924–933.

van Ginneken, B., de Bruijne, M., Loog, M., Viergever, M., 2003. Interactive shape models. In: Sonka, M., Fitzpatrick, M. (Eds.), Medical Imaging: Image Processing. Proceedings of SPIE, vol. 5032. SPIE Press, pp. 1206–1216.

Weese, J., Kaus, M., Lorenz, C., Lobregt, S., Truyen, R., Pekar, V., 2001. Shape constrained deformable models for 3D medical image segmentation. In: Insana, M., Leahy, R. (Eds.), Information Processing in Medical Imaging. Lecture Notes in Computer Science, vol. 2082. Springer, pp. 380–387.

Wever, J., Blankensteijn, J., van Rijn, J., Broeders, I., Eikelboom, B., Mali, W., 2000a. Inter- and intra-observer variability of CTA measurements obtained after endovascular repair of abdominal aortic aneurysms. American Journal of Roentgenology 175 (5), 1282–1297.

Wever, J., Blankensteijn, J., Mali, W., Eikelboom, B., 2000b. Maximal aneurysm diameter follow-up is inadequate after endovascular abdominal aortic aneurysm repair. European Journal of Vascular and Endovascular Surgery 20 (2), 177–182.

Wilson, L., Brown, S., Young, J., Li, R., Brandt, L., 1999. Three-dimensional computer models of abdominal aortic aneurysms by knowledge-based segmentation. In: Lemke, H., Vannier, M., Inamura, K., Farman, A. (Eds.), Computer Assisted Radiology and Surgery. Excerpta Medica International Congress Series, vol. 1191. Elsevier Publishers, Amsterdam, pp. 213–217.

Wink, O., Niessen, W., Viergever, M., 2000. Fast delineation and visualization of vessels in 3-D angiographic images. IEEE Transactions on Medical Imaging 19 (4), 337–346.