Contour fitting using an adaptive spline model

Daniel Rueckert and Peter Burger
Imperial College of Science, Technology and Medicine

Abstract

This paper presents a new segmentation algorithm by fitting active contour models (or snakes) to objects using adaptive splines. The adaptive spline model describes the contour of an object by a set of piecewisely interpolating C^2 polynomial spline patches which are locally controlled. Thus the resulting description of the object contour is continuous and smooth. Polynomial splines provide a fast and efficient way for interpolating the object contour and allow us to compute its internal energy due to bending and elasticity deformations analytically. The adaptive spline model can be represented by its spline control points. The accuracy of the model is gradually increased during the segmentation process by inserting new control points. For estimating the optimal position of the control points, two different relaxation techniques based on Markov–Random–Fields (MRFs) have been combined and evaluated: Simulated Annealing (SA), which is a stochastic relaxation technique, and Iterated Conditional Modes (ICM), which is a probabilistic relaxation technique. We have studied convergence behavior and performance on artificial and medical images. The results show that the combination of both relaxation techniques provides very robust and initialization independent segmentation results.

1 Introduction

Active contour models or *snakes* were first introduced by Kass et al. [10] to model and segment the contour of objects in 2–D images. Snakes are a class of deformable models which try to minimize their associated energy function. A snake in 2–D can be represented as a function of arc length s by $\mathbf{v}(s) = (x(s), y(s))$ with $s \in [0..1]$. The energy of a snake can be written in the following form:

$$E = \int_0^1 E(\mathbf{v}(s))ds = \int_0^1 E_{int}(\mathbf{v}(s)) + E_{image}(\mathbf{v}(s))ds \qquad (1)$$

E_{int} represents the internal energy of the spline due to bending or discontinuities and is given by

$$E_{int}(\mathbf{v}(s)) = \alpha(s)|\mathbf{v}_s(s)|^2 + \beta(s)|\mathbf{v}_{ss}(s)|^2. \qquad (2)$$

The first derivative term $|\mathbf{v}_s(s)|^2$ controls the elasticity of the spline while the second term $|\mathbf{v}_{ss}(s)|^2$ controls the bending of the spline. Setting $\alpha(s)$ to zero would allow the curve to develop gaps; setting $\beta(s)$ to zero would allow the curve to develop corners.

The image energy introduces a coupling of the snake to image features. For example, to drive the snake towards strong edges in the image, the image energy can be expressed by the gradient magnitude, e.g.

$$E_{image}(\mathbf{v}(s)) = -|\nabla I(x,y)|^2. \tag{3}$$

Kass et al. used a gradient descent technique to solve the minimization problem via the Euler–Lagrange method. Other optimization techniques include dynamic programming [1], greedy optimization [17] as well as genetic algorithms (GA) [9] and finite element methods (FEM) [4].

However, the successful application of snakes to image segmentation is limited by the sensitivity of the algorithm towards noise and different initializations. This is especially true for medical image segmentation where robustness and reproducibility is crucial. The sensitivity of the classical snake model is mainly caused by two fundamental problems: First, local optimization techniques will not find the global energy minimum. This makes the approach very sensitive to local minima which can be caused by noise or operator–dependent initialization. Optimization techniques like genetic algorithms (GA) provide an alternative to purely local optimization techniques, however their convergence properties raise some practical problems. Second, discrete polygonal approximations assuming a fixed number of polygon points are usually used to describe the object contour. Such polygonal approximations have several disadvantages: On the one hand, the number of polygon points needed to describe the object contour accurately is not known. On other hand, the estimation of the curvature due to bending and elastic deformations using finite differences can be problematic [17].

2   Contour fitting by adaptive splines

We propose a new adaptive spline model similar to *Geometrically Deformable Models* (GDM) [12] to overcome these problems. The adaptive spline model represents the object contour by a set of C^2 spline patches which are locally controlled by their defining control points. Its accuracy is gradually increased by resampling the model and adding new control points. Using this notation the model is defined within the framework of Markov–Random–Fields (MRFs). Two different relaxation techniques are used during the contour fitting process: In the initial stage Simulated Annealing (SA) is used to escape local minima and to find a global minimum. In later stages Iterated Conditional Modes (ICM) is used to achieve a fast and efficient convergence. Both techniques have been successfully used for the model based segmentation of ventricular contours [7], [6].

The adaptive spline model has several advantages: Its representation is compact, analytical and continuous as well as locally controlled. Using this model, the

energy due to elastic and bending deformations of the contour can be computed analytically. Moreover, most objects of interest in medical images have a smooth shape which does not vary very rapidly. These objects can be usually described accurately with only a few control points. Thus the size of the resulting MRF is computationally feasible. Another important feature of this representation is the fact that the contour shape is locally controlled: Changing a control point of the adaptive spline model influences only the contour in the local neighborhood of the control point. Other analytical shape representation models like fourier descriptors [14] or active shape models [5] do not provide any local control and can therefore not easily be defined within the framework of MRFs.

2.1 The adaptive spline model

An adaptive spline model can be described by an ordered set of n points, $P = \{p_1, \ldots, p_n\}$, where p_i denotes the ith control point of the interpolating spline. We generate a smooth curve by using a set of curve segments whose coefficients depend only on a few points. The contour Q described by the adaptive spline model P is given by the set of curve segments $Q = \{Q_1, \ldots, Q_n\}$ which are defined by the control points. Each curve segment $Q_i(\tau)$ describes a curve starting at point p_i and ending at point p_{i+1} with its parameter τ ranging between 0 and 1. Since we want to be able to compute the internal energy of the curve due to elastic and bending deformations, we require the curve to be twice differentiable, thus being C^2 continuous.

For example, natural cubic splines are C^2 continuous but do not provide any local control element which is important to keep the computational complexity small. Instead we will use a fifth–order polynomial function allowing us to easily compute an interpolating curve which is C^2 continuous and provides local control. The derivation of the coefficients for a C^1 polynomial spline is given in [3]. In order to extend the concept for C^2 polynomial splines, two new conditions about the second derivative of the curve segment Q_i at the joint points $\tau = 0$ and $\tau = 1$ have to be added. Using these six different conditions for the spline it is possible to find exactly one fifth–order polynomial function $K(\tau)$ which satisfies all six conditions.

$$K(\tau) = a\tau^5 + b\tau^4 + c\tau^3 + d\tau^2 + e\tau + f \qquad 0 \leq \tau \leq 1 \qquad (4)$$

With these six conditions we obtain a system of linear equations which can be written in matrix form as

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 5 & 4 & 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 20 & 12 & 6 & 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} p_i \\ p_{i+1} \\ p'_i \\ p'_{i+1} \\ p''_i \\ p''_{i+1} \end{pmatrix} \qquad (5)$$
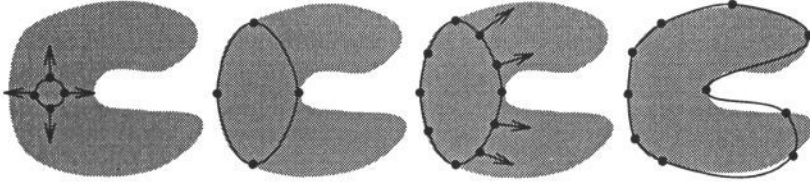
Figure 1: Deformation and resampling process

Inverting the matrix yields:

$$
\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} -6 & 6 & -3 & -3 & -0.5 & 0.5 \\ 15 & -15 & 8 & 7 & 1.5 & -1 \\ -10 & 10 & -6 & -4 & -1.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} p_i \\ p_{i+1} \\ p'_i \\ p'_{i+1} \\ p''_i \\ p''_{i+1} \end{pmatrix}
\tag{6}
$$

Using this matrix it is now possible to compute the coefficients of the polynomials $x(\tau)$ and $y(\tau)$. The curve segment Q_i can be described as a parametric vector of two functions $x(\tau)$ and $y(\tau)$ with $Q_i(\tau) = (x(\tau), y(\tau))$. The tangent vector $\mathbf{t}(\tau)$ and the normal vector $\mathbf{n}(\tau)$ of a curve in parametric form are given by

$$
\mathbf{t}(\tau) = (\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}) \quad \text{and} \quad \mathbf{n}(\tau) = (\frac{-\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}).
\tag{7}
$$

2.2 Contour fitting

The concept of contour fitting is similar to the concept for GDMs in [12], [15]: The contour fitting starts with an initial model which is placed by the user inside the object. The contour fitting algorithm then performs alternately a deformation and resampling process (Figure 1):

- *The deformation process*: The deformation process can be viewed as minimizing the energy function of the model. The energy during the initial deformation process is minimized by SA while following minimizations are carried out by ICM. The deformation process can take place only along the direction of the normal vector $\mathbf{n}(\tau)$ of the contour at the control points.

- *The resampling process*: The resampling process allows the spline model to adapt to the number of control points which are necessary to describe the object. The length and the curvature of each curve segment is used as criteria to adapt the number of control points: If the length or the curvature of a curve segment exceeds a certain threshold, the curve segment is subdivided by adding a new control point.

The contour fitting process stops if no further subdivision is necessary.

3 Energy function

The adaptive snake model is composed of n different curve segments $Q_i(\tau)$, where the parameter τ ranges between 0 and 1. Consequently, the energy of the model becomes

$$E = \sum_{i=1}^{n} E(Q_i) \tag{8}$$

where $E(Q_i)$ denotes the energy of curve segment Q_i. The internal energy of a curve segment Q_i due to elastic and bending deformation is given by

$$E_{int} = E_1(Q_i) + E_2(Q_i) = \int_0^1 |\dot{Q}_i(\tau)|^2 d\tau + \int_0^1 |\ddot{Q}_i(\tau)|^2 d\tau. \tag{9}$$

Since the curve is represented in analytical form one can directly compute the energy due to elastic deformations via the definite integral:

$$
\begin{aligned}
\int_0^1 |\dot{Q}(\tau)|^2 d\tau \quad = \quad & \frac{25}{9}(a_x^2 + a_y^2) + 5(a_x b_x + a_y b_y) + \frac{16}{7}(b_x^2 + b_y^2) + \frac{30}{7}(a_x c_x + \\
& a_y c_y) + 4(b_x c_x + b_y c_y) + \frac{9}{5}(c_x^2 + c_y^2) + \frac{10}{3}(a_x d_x + a_y d_y) + \\
& \frac{16}{5}(b_x d_x + b_y d_y) + 3(c_x d_x + c_y d_y) + \frac{4}{3}(d_x^2 + d_y^2) + 2(a_x e_x + \\
& b_x e_x + c_x e_x + d_x e_x + a_y e_y + b_y e_y + c_y e_y + d_y e_y) + \\
& e_x^2 + e_y^2 \quad ,
\end{aligned}
\tag{10}
$$

where the x and y coefficients of the polynomial are denoted with subscripts. The energy of a curve segment due to bending can be computed similarly by the definite integral of $|\ddot{Q}(\tau)|^2$:

$$
\begin{aligned}
\int_0^1 |\ddot{Q}(\tau)|^2 d\tau \quad = \quad & \frac{400}{7}(a_x^2 + a_y^2) + 80(a_x b_x + a_y b_y) + \frac{144}{5}(b_x^2 + b_y^2) + 48(a_x c_x + \\
& a_y c_y) + 36(b_x c_x + b_y c_y) + 20(a_x d_x + a_y d_y) + 16(b_x d_x + \\
& b_y d_y) + 12(c_x^2 + c_y^2 + c_x d_x + c_y d_y) + 4(d_x^2 + d_y^2)
\end{aligned}
\tag{11}
$$

The second part of the energy function is designed to detect image features like edges. An edge can be characterized by its derivatives, namely by its first and second derivative. The first derivative or the magnitude of the gradient vector, ∇I, can be used to measure the strength of an edge. The zero–crossings of the second derivative indicate a local optimum of the first derivative which is equivalent to the centre of an edge. In order to maximize the gradient and to minimize the distance D to the nearest zero–crossing of each pixel along the curve, the image energy can be written as

$$E_{image} = E_3(Q_i) + E_4(Q_i) = \sum_{(x,y) \in Q_i} (-|\nabla I(x,y)|^2 + |D(x,y)|^2) \tag{12}$$

The total energy is computed as a linear combination of the energy terms E_i. The influence of the each energy term is controlled by constant weighting factors α_i:

$$E(Q_i) = \alpha_1 E_1(Q_i) + \alpha_2 E_2(Q_i) + \alpha_3 E_3(Q_i) + \alpha_4 E_4(Q_i) \tag{13}$$

4 Optimization by stochastic and probabilistic relaxation

The deformation of the adaptive spline model is controlled by minimizing its associated energy function. Thus the robustness of the minimization is essential in order to achieve good segmentation results. The size of the configuration space makes any direct optimization impossible, even for very small MRFs. We have consequently implemented Simulated Annealing (SA) as a minimization technique, which is able to escape local energy minima. However, our contour fitting process requires several optimizations which makes the use of SA with its high computational demands not practical. We have implemented a second minimization technique, Iterated Conditional Modes (ICM), which is an efficient local minimization technique. ICM is very robust concerning noise if given a reasonable estimate of the object contour. Using the final estimate of the previous deformation process as an initial estimate for the following deformation process, the combination of SA and ICM provides a very robust and efficient minimization technique.

4.1 Stochastic relaxation

Simulated Annealing [11], [16] is a stochastic relaxation technique which is based on the analogy to the physical process of annealing a metal: At high temperatures the atoms are randomly distributed. With decreasing temperatures they tend to arrange themselves in a crystalline state which minimizes their energy. Using this analogy, the algorithm generates randomly new configurations by sampling from the probability distribution of the system.

If the adaptive spline model is defined as a 1–D Markov–Random–Field (MRF), which assumes that the position of a control point depends only on itself and on its neighbors, the probability distribution is given by a Gibbs distribution:

$$P(X = \omega) = \frac{1}{Z(T)} \exp[-\frac{E(\omega)}{T}] \tag{14}$$

where $Z(T)$ is a normalization factor. T is a control parameter, called temperature, which influences the form of the probability distribution. New configurations are accepted with a certain acceptance probability $H(T)$ depending on the temperature:

$$H(T) = \exp[-\frac{\Delta E}{T}] \tag{15}$$

Since increases of energy can be accepted, the algorithm is able to escape local energy minima. It has been shown [8] that the algorithm converges to a global energy minimum, if the temperature at iteration k is

$$T(k) \geq \frac{c}{\log(1 + k)} \quad , \tag{16}$$

where c is a constant depending on the amount of energy which is necessary to escape local minima.
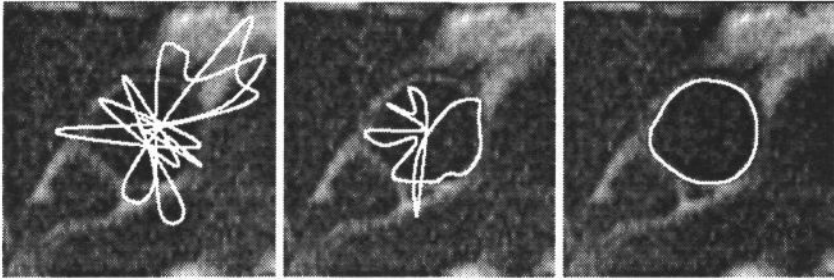
Figure 2: Evolution of the adaptive spline model during the SA optimization

4.2 Probabilistic relaxation

In contrast to stochastic relaxation, which makes random changes, probabilistic relaxation makes only deterministic changes and converges much faster. Besag [2] proposed an iterative algorithm, called Iterative Conditional Modes (ICM), which maximizes the conditional probability based on a provisional estimate: Suppose $\hat{\omega}$ denotes a provisional estimate of our adaptive spline model and that our goal is to update the position of the control point v_i in the context of all available information. This means that we want to maximize the conditional probability of ω_i given our current estimate $\hat{\omega}$. Since our system is described by a MRF, the probability $P(\omega_i|\hat{\omega})$ is given by the Gibbs distribution in eq. (14).

It should be noted that the ICM algorithm is equivalent to the SA algorithm with instantaneous freezing; it converges therefore much faster. However, the ICM algorithm is not a global minimization technique but rather a local one, as the results are dependent on the initial estimate.

5 Results

We have tested the segmentation algorithm on medical images. These medical images are magnetic resonance (MR) images of the cardiovascular system. The task was to segment the ascending and descending aorta to measure regional aortic compliance. This is a very challenging task since MR images of the cardiovascular system are characterized by motion artefacts and a low signal-to-noise ratio which makes the accurate and reproducible segmentation even for expert cardiologists difficult. The parameters of the algorithm have been fixed empirically. The same set of parameters has been used for all tests. It should be noted that the initial estimate of the contour does not influence the final estimate of the contour. This is a result of the SA optimization which at high temperatures makes the system behave chaotic while at lower temperatures the systems becomes more stable and finally converges to the global optimum. The evolution of the adaptive spline model during the SA optimization is shown in Figure 2.
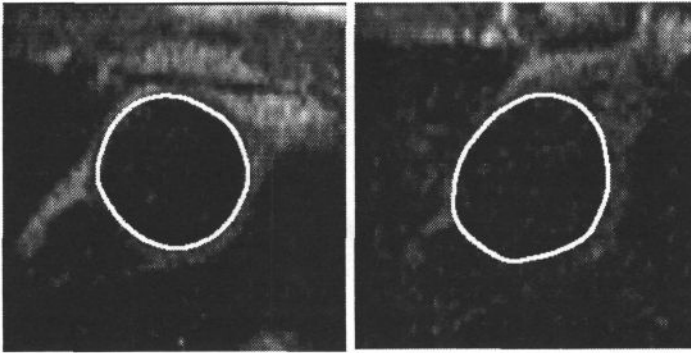
Figure 3: Segmentation results of the ascending aorta in spin echo MR images

In the first study we have compared manual and automated segmentation results of the ascending and descending aorta of 12 individuals. The images are spin-echo MR images with a resolution of 256 x 256 and are acquired at the systole and diastole of the heart cycle. Figure 3 shows the segmentation results for these images. To evaluate the accuracy and reliability of the segmentation, we have compared the intra- and interobserver variability of expert cardiologists with the performance of our segmentation algorithm. It is important to note that the standard procedure for measuring the aortic area requires four different manual segmentations by the same cardiologist. The average area of all four segmentation results is then taken as the *standard area*. The intraobserver variability indicates the variation of these measurements with respect to the standard area. Due to the low signal-to-noise ratio and the size of the aorta the intraobserver variability is about 3.1 % for the ascending aorta and about 2.7 % for the descending aorta. Another important evaluation factor is the interobserver variability which indicates the variation of the standard area measured by different cardiologists. The interobserver variability lies between 3.9 % for the ascending and 3.3 % for the descending aorta. Compared to these results, our segmentation algorithm performs very well: The segmentation error, which is defined as the deviation between the standard manual segmentation and our algorithm, lies between 3.2 % for the ascending and 2.9 % for the descending aorta. Thus the segmentation error of the computer–based segmentation is only slightly larger than the deviation of different segmentations by the same observer but smaller than the deviation between different observers. As a consequence our algorithm allows the precise and objective measurement of the aortic compliance in spin echo MR images.

The second set of tests was performed on cine gradient echo MR image sequences of the heart. These images have a resolution of 256 x 256 and contain 16 frames in time and cover the whole heart cycle. As before, the ascending as well as the descending aorta have been located successfully and their boundaries have been correctly segmented (Figure 4). In both cases we have found that the segmentation algorithm provides an accurate and automated method for reproducible
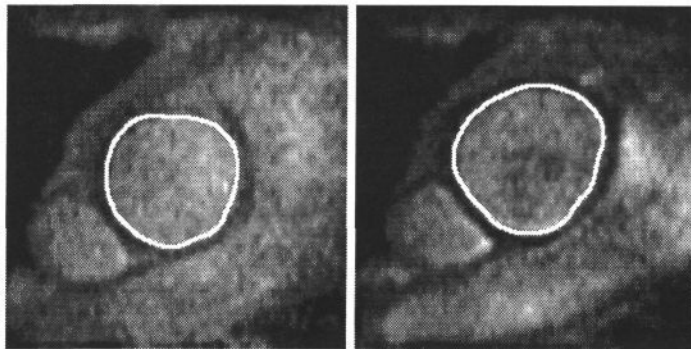
Figure 4: Segmentation results of the ascending aorta in gradient echo MR images

quantitative studies of regional aortic compliance.

6 Conclusion and future work

We have presented a new algorithm for active contour models using a contour representation based on piecewisely interpolating polynomial splines. This compact and analytic representation allows for a fast and efficient computation of the internal energy due to elastic and bending deformations avoiding the problems of discrete approximations. The use and combination of stochastic and probabilistic relaxation methods yields very robust segmentation results which overcomes most of the problems concerning noise and initialization using the classical snake model.

As the presented results are very encouraging and have demonstrated the robustness for 2D images, there is need to further investigate the influence of the energy function and its parameters. A critical point is concerned with the a-priori constraints made in the energy function. For example, the smoothness (or minimal curvature) constraint may not be true for object with more complex shapes, e.g. the human brain. A possibility of avoiding the problems, e.g. matching contour and image curvature, is given in [13].

References

[1] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855–867, September 1990.

[2] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Stat. Soc. B*, 48(3):259–302, 1986.

[3] P. Burger and D. Gillies. *Interactive Computer Graphics*, pages 280–282. Addison–Wesley, 1989.

[4] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[5] T. F. Cootes and C.J. Taylor. Active shape models - smart snakes. In *British Machine Vision Conference*, pages 267–275, 1992.

[6] M. T. de Figueiredo and J. M. N. Leitão. Bayesian estimation of ventricular contours in angiographic images. *IEEE Transactions on Medical Imaging*, 11(3):416–429, September 1992.

[7] N. Friedland and D. Adam. Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing. *IEEE Transactions on Medical Imaging*, 8:344–353, December 1989.

[8] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):721–741, 1984.

[9] A. Hill, T.F. Cootes, and C.J. Taylor. A generic system for image interpretation using flexible templates. In *British Machine Vision Conference*, pages 276–285, 1992.

[10] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proceedings of the first international conference on Computer Vision*, 1987.

[11] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, May 1983.

[12] J. V. Miller, D. E. Breen, W. E. Lorensen, R. M. O'Bara, and M. J. Wozny. Geometrically deformed models: A method for extracting closed geometric models from volume data. *Computer Graphics*, 25(4):217 – 226, July 1991.

[13] Julia A. Schnabel and Simon R. Arridge. Active contour models for shape description using multiscale differential invariants. In *British Machine Vision Conference*, 1995. These proceedings.

[14] L. H. Staib and J. S. Ducan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11):1061 – 1075, November 1992.

[15] K. D. Toennies and D. Rueckert. Image segmentation by stochastically relaxing contour fitting. In *SPIE Conference on Medical Imaging, Image Processing*, volume 2167, pages 18–27, Newport Beach, USA, February 1994.

[16] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications*. Reidel, 1987.

[17] D. J. Williams and M. Shah. A fast algorithm for active contours and curvature estimation. *CVGIP: Image Understanding*, 55(1):14–26, 1992.