A new 2D segmentation method based on dynamic programming applied to computer aided detectection in mammography.

Sheila Timp, Nico Karssemeijer

Department of Radiology
University Medical Center Nijmegen
The Netherlands

Corresponding author:

Sheila Timp
Department of Radiology
Radboud University Hospital
Geert Grooteplein Zuid 18
6525 GA Nijmegen
The Netherlands
Tel: +31-24-3619811
Fax: +31-24-3540866

Abstract

Mass segmentation plays a crucial role in computer-aided diagnosis (CAD) systems for classification of suspicious regions as normal, benign or malignant. In this paper we present a robust and automated segmentation technique - based on dynamic programming - to segment mass lesions from surrounding tissue. In addition, we propose an efficient algorithm to guarantee resulting contours to be closed.

The segmentation method based on dynamic programming was quantitatively compared with two other automated segmentation methods (region growing and the discrete contour model) on a dataset of 1210 masses. For each mass an overlap criterion was calculated to determine the similarity with manual segmentation. The mean overlap percentage for dynamic programming was 0.69, for the other two methods 0.60 and 0.59 respectively. The difference in overlap percentage was statistically significant.

To study the influence of the segmentation method on the performance of a CAD system two additional experiments were carried out. The first experiment studied the detection performance of the CAD system for the different segmentation methods. FROC (Free Receiver Operator Characteristic) analysis showed that the detection performance was nearly identical for the three segmentation methods. In the second experiment the ability of the classifier to discriminate between malignant and benign lesions was studied. For region based evaluation the area A_z under the ROC (Receiver Operating Characteristic) curve was 0.74 for dynamic programming, 0.72 for the discrete contour model and 0.67 for region growing. The difference in A_z values obtained by the dynamic programming method and region growing was statistically significant. The differences between other methods were not significant.

keywords: dynamic programming, segmentation, classification, breast masses, computer aided diagnosis (CAD)

1 Introduction

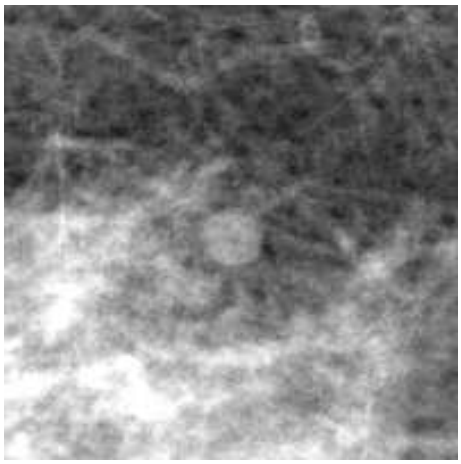
To improve the efficiency of screening mammography CAD methods are being developed ([1], [2], [3]. An important step in CAD algorithms is the segmentation of mammographic masses. Globally we can divide segmentation methods into two main categories: region based and edge based. Methods of both categories have been applied to the segmentation of mammographic masses.

The first category assigns each pixel to a particular object or region. Examples are split-and-merge algorithms and region growing techniques. Region growing is one of the most popular segmentation methods and many different approaches have been proposed. For example, Kupinski and Giger developed two extended region growing techniques, one based on the radial gradient index and another based on simple probabilistic models [4]. They tested these methods against a conventional region growing algorithm using a database of biopsy-proven, malignant lesions and found that the new lesion segmentation algorithms more closely matched radiologists' outlines of these lesions. Guliato et al. proposed fuzzy region growing methods for segmenting breast masses and further classified the segmented masses as benign or malignant based on the transition information present around the segmented regions [5, 6]. Petrick et al. applied object based region growing in combination with a density-weighted contrast enhancement filter to segment all significant structures within the breast [7]. One of the problems of region growing is that small and low-contrast structures have a tendency to grow into the background and become large regions even though the actual mass is quite small. An example is given in figure 1(d). The region growing method fails to find the border of the mass and the resulting segmentation is too large. Another problem is that structures containing internal gradients do not always grow to the correct border but can end up containing only a section of the true object.

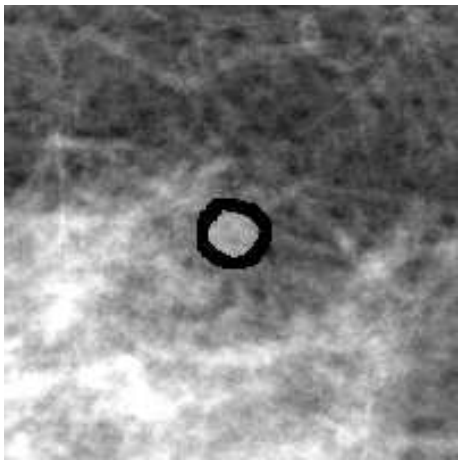
The second category are edge based algorithms. Instead of dividing the image into object and background pixels, the boundary of an object is detected. Most algorithms first construct a so-called *edge image*. In the edge image each pixel is assigned a value according to the edge strength. Based on this image, pixels with strong edges are selected and linked to each other. In most cases the linked pixels represent object boundaries. A disadvantage of the original edge based algorithms was that a closed contour could not be guaranteed. To overcome this problem, an active contour model (snake) was developed for contour detection. Dynamic contour models (snakes) have become en vogue with the snake model of Terzoupolos and coworkers and have, since then, been investigated and applied in various ways ([8]). The snake model builds a deformable contour consisting of connected spline segments and let the contour approximate a desired form by minimizing an energy function containing internal and external energy. The internal energy is the bending energy of the spline, the external energy is calculated by integrating image features, like the presence of lines and edges. Lobregt and Viergever developed a discrete version of the snake model (discrete contour model) and applied the model to medical images [9]. The main drawback of using deformable models for the task of mammographic mass segmentation is that the algorithm heavily depends on being initialized with a contour that is close to the actual boundary. Otherwise the contour may stick to the first strong edge it finds. An example is shown is figure 2(c). The initial estimate of the contour, shown in black, is too far from the mass boundary. As a result the model is not able to find the contour and instead is attracted to the pectoral muscle. Another known problem with deformable models is that the model may shrink owing to internal forces, if the edges are not strong enough or too far from the initial contour.

There are a few studies that compare different segmentation methods [10], [11], [12]. Timp et al compared three segmentation methods with manual segmentation. However, they did not evaluate the effect of the segmentation on the classification performance. Te Brake et al. compared the discrete contour model from Lobregt et al. with the region growing algorithms developed by Kupinski et al. and evaluated the methods by comparing them to manual segmentation ([4], [9], [10]). Furthermore they studied the effect of the segmentation on the detection performance. One of the region growing methods and the discrete contour model performed equally well in the segmentation task. In the detection experiment the discrete contour model had a higher performance in classifying the segmented region as normal or abnormal. Sahiner et al. compared a mass segmentation method based on an active contour model with manual segmentation and studied the effect of the segmentation on the classification accuracy [12]. They found that the A_z values using features based on manual segmentation and automated segmentation were nearly identical.

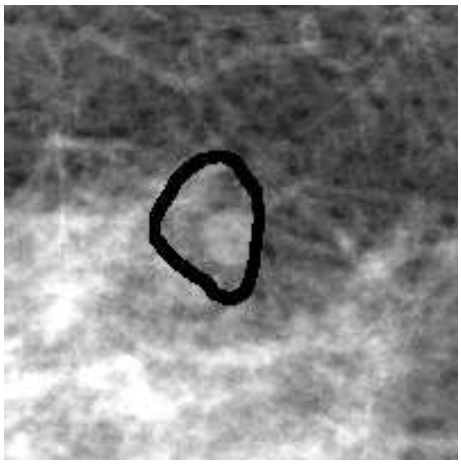
In this study we developed a new segmentation method to overcome the problems of region growing and the discrete contour model. The new method uses both edge based information as well as a priori



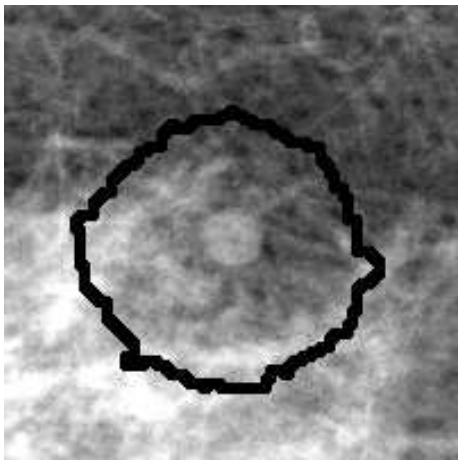
(a)



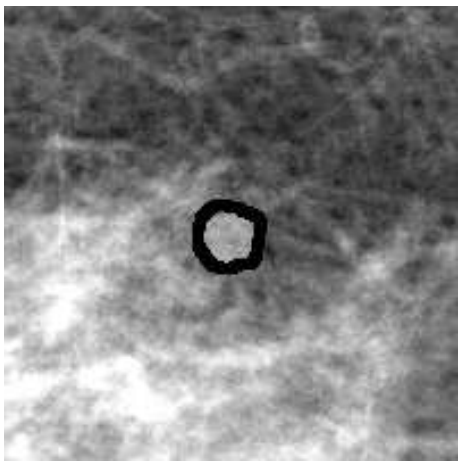
(b)



(c)

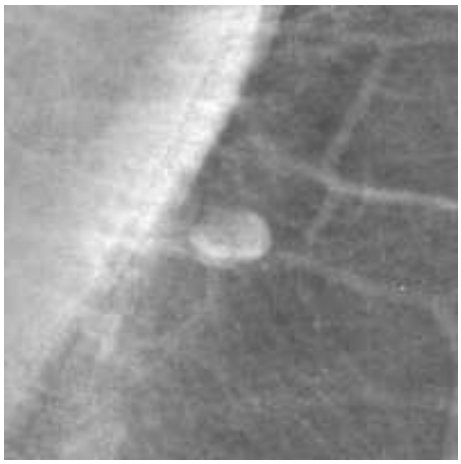


(d)

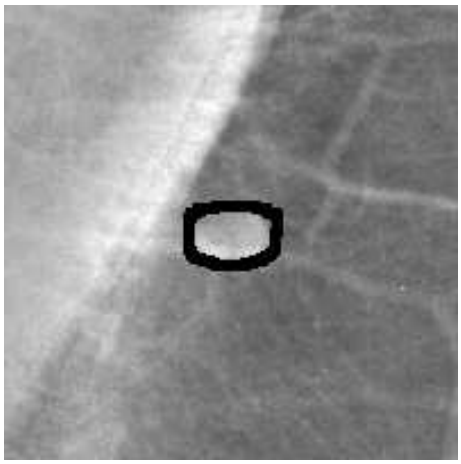


(e)

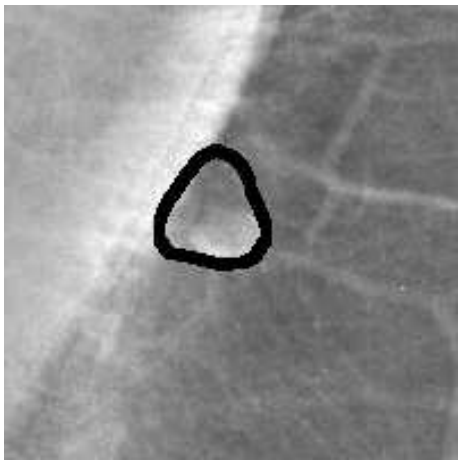
Figure 1: Example of segmentation results for a benign mass: 1(a) original mammogram with benign mass, 1(b) manually segmented mass, 1(c) discrete contour model segmentation, 1(d) region growing method and 1(e) the proposed dynamic programming algorithm.



(a)

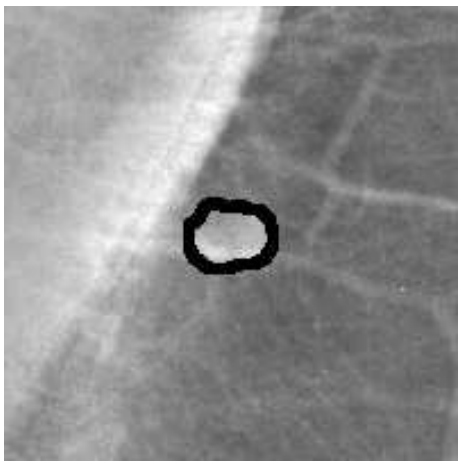


(b)



(c)

(d)



(e)

Figure 2: Example of segmentation results for a benign mass that is located near the pectoral muscle: 2(a) original mammogram with benign mass, 2(b) manually segmented mass, 2(c) discrete contour model segmentation, 2(d) region growing method and 2(e) the proposed dynamic programming algorithm.

knowledge about the gray level distribution of the ROI (region of interest) around the mass. The best contour is found by an optimization technique based on dynamic programming. To test the performance of this method, we compared it with region growing and the discrete contour model using an area overlap criterion. Furthermore we studied the effect of the segmentation on the detection and classification of mammographic masses.

The paper is organized as follows. Section 2 presents an explanation of the different segmentation methods. The segmentation method based on dynamic programming is described in more detail in section 2.1. The experiments are described in section 3. Results are presented in section 4 with a discussion and conclusion in the last sections.

2 Segmentation methods

In this section, the three segmentation methods used in this work are described. The first subsection describes the dynamic programming approach. In the second and third subsection the region growing method and the discrete contour model are briefly reviewed.

2.1 Dynamic Programming

Boundary definition via dynamic programming can be formulated as a graph searching problem where the goal is to find the optimal path between a set of start nodes and a set of end nodes. Typical applications of the use of dynamic programming in boundary tracking problems are tracing borders of elongated objects like roads and rivers in aerial photographs and the segmentation of handwritten characters. Medical applications include the segmentation of spine boundaries and tracing vessel borders.

To apply dynamic programming to find the boundary of a mass we notice that the shape of most masses is approximately circular. This circularity constraint is implemented by carrying out the calculations in polar space. For the transform we use a circular region of interest (ROI) with center (μ_x, μ_y) and radius R. The center of the ROI defines the origin for the coordinate transform and should be within the suspect lesion. The radius should be chosen large enough to allow application of the algorithm to all masses of interest. We choose a radius of 2.4 *cm*. Figure 3(a) and 3(b) show the coordinate transform. All pixels inside a circular ROI in the original image are transformed to the polar ROI 3(b). The x-axis in the polar image represents the angle from $-\pi$ to π, the y-axis the radius from 0 to R. The dynamic programming algorithm is applied to the polar ROI to find the optimal path from one of the pixels in the first column to one of the pixels in the last column.

The optimal path is defined as the minimum cumulative cost path, where the cumulative cost of a path is the sum of the local costs of the pixels on the path. We will now describe the local cost.

2.1.1 Local Cost

Local cost is cost assigned to each pixel in the polar image. This cost should embody a notion of a good boundary: pixels that possess many characteristics of the searched boundary are assigned low cost and vice versa. The local cost components form the following cost function:

$$c(i,j) = w_s s(i,j) + w_d d(i,j) + w_g g(i,j).$$

where s represents the edge strength, d the deviation from an expected size, and g the deviation from an expected gray level. The weights for the components are given by w_s, w_d and w_g.

Edge Strength $s(i,j)$ As most contours exhibit strong edges we want to assign pixels with strong edge features low cost. The edge strength for each pixel is determined by calculating the gradient magnitude in the direction normal to the contour. This corresponds with the gradient in vertical direction in the polar image. Then the relative edge strength is determined by normalizing the gradient values with the maximum gradient $\max(y')$. This normalization ensures that subtle contours with low global but high local edge strength can be found as well. The normalized gradient value is inverted so high gradients produce low costs and vice versa. The gradient component function is

$$s(i,j) = \frac{\max(y') - y'(i,j)}{\max(y')}$$

where y' is the gradient magnitude in vertical direction. For $\max(y')$ we took the 99th percentile of the gradient values measured in the ROI. By taking the 99th percentile it is prevented that one outlier, for instance a very bright micro-calcification, decreases the relative edge strength of all other pixels in the ROI.

Mass size $d(i,j)$ Contours that enclose a mass with a size common for masses are assigned low cost value. On the other hand, very small and very large segmentations are assigned higher cost. Most masses have a radius between 5 mm and 15 mm, with a mean radius of about 9 mm [13]. The following formula was used to incorporate size information in the cost function:

$$d(i,j) = \begin{cases} (j-\mu)^2 & : \quad j < m \\ (m-\mu)^2 & : \quad j >= m \end{cases}$$

where μ is the mean radius of masses. A cost limit m is set to prevent that the size component of the cost function completely determines the value of the cost function for large masses. This limit is set to 15 mm. Alternatively, the cost component for mass size could be obtained by estimating the size distribution of masses. The probabilities for each size could then be used to determine the cost value $d(i,j)$. To use this method a large representative database with benign and malignant masses of known size is needed to accurately determine the probabilities for each mass size. Currently we used the first method as we do not have an independent database that we can use for this purpose.

Deviation from expected gray level $g(i,j)$ Another characteristic of the mass boundary is the gray level. This gray level should correspond with the edge of the object. A common assumption is that this edge is located at the zero crossing of the second derivative of the edge profile. In projective images however, the real edge is located toward the darker side (background). Consequently, the gray value of the border will have a value close to the background gray level [14]. By estimating the intensity distribution of the mass and the background a preferred gray level for the contour can be determined.

$$g = \alpha\mu_{mass} + (1-\alpha)\mu_{background}$$

where α should be smaller than 1/2 to ensure the edge is located more toward the background level. The gray level component of the cost function is defined as

$$g(i,j) = sqrt(abs(G(i,j) - g))$$

where $G(i,j)$ is the intensity value of the pixel (i,j). To estimate the gray level of the mass and the background we used two methods. The first method estimates μ_{mass} and $\mu_{background}$ as the mean gray level inside and outside an initial estimate of the contour. As initial estimate we took a circle with radius 0.6 cm. In the second method we used histogram analysis to estimate the gray level distributions. We found that most histograms of the ROI can be modeled reasonably well by a mixture of two Gaussian distributions, one narrow Gaussian in the low intensity range representing the fatty tissue, and a broader one in the middle/high intensity range representing dense tissue and/or masses. The parameters for the Gaussian distributions were estimated using the Levenberg-Marquardt method. The first peak in the histogram is used to estimate $\mu_{background}$, and the second peak to estimate μ_{mass}. If the histogram could not be modeled by a mixture of two Gaussians we used the first method to estimate the preferred gray level.

2.1.2 Dynamic programming path finding algorithm

Application of the cost function to all pixels in the polar image results in the so called cost image. This image can be seen as a graph in which the dynamic programming algorithm should find the path with the lowest cost. The first column in the cost image $c(i,0)$ represents the start nodes for the algorithm, whereas the end nodes are represented by the pixels in the last column of the image. The cumulative cost of each path is stored in the cumulative cost matrix. The cumulative cost matrix is constructed in two steps. First the cumulative cost of pixels in the first column are set equal to the cost of these pixels:

$$C(i,0) = c(i,0),$$

where $C(i,j)$ is the cumulative cost and $c(i,j)$ is the cost value for pixel (i,j) in the polar image. For the other pixels the cumulative cost is calculated by a recursive step:

$$C(i,j+1) = \min_{-m \leq l \leq m} C(i+l,j) + c(i,j+1) + h(l). \quad (1)$$

The additional cost of a segment of the path for column j to $j+1$ depends on the cost value of pixel (i,j) and the direction l. The cost of the direction is set according to a function $h(l)$ which we use to control smoothness. $h(l)$ is set to zero for directions outside the interval $[-2,\ldots,2]$. The cumulative cost matrix is shown in figure 3(d). The final contour is found by selecting those pixels that linked together from the boundary with the lowest cost. The endpoint $C(i,\pi)$ of the contour is the pixel in the last column of the cumulative cost matrix with the lowest cost. The optimal path is found by back tracing the path from the end pixel to one of the pixels in the first column (figure 3(e)). The optimal path in the polar image is transformed back to rectangular coordinates in the original image. The resulting segmentation is shown in figure 3(f).

Final contour The dynamic programming algorithm does not guarantee the contour to be closed. In our application we consider a contour as closed if the distance between the start and the end points is less than 3 pixels. This is conform the smoothness function $h(l)$ (1) which is zero outside the interval $[-2,\ldots,2]$. In most cases, especially if the mass is clearly visible, the program will correctly segment the mass by a closed contour. However, if the mass is vague or if other structures obscure the mass boundary, the segmentation program can fail to find a closed contour as the optimal path. The problem is illustrated in figure 4. The small mass in the middle of the image is surrounded by some dense tissue 4(a). Figure 4(b) shows the polar image, with the resulting contour plotted on top of it. In the beginning the path is attracted to an image structure and deviates from the true mass boundary. As a consequence the contour is not closed and contains some extra tissue. Figure 4(d) shows the final contour on the original image. The start and end point have been connected in order to get a closed boundary.

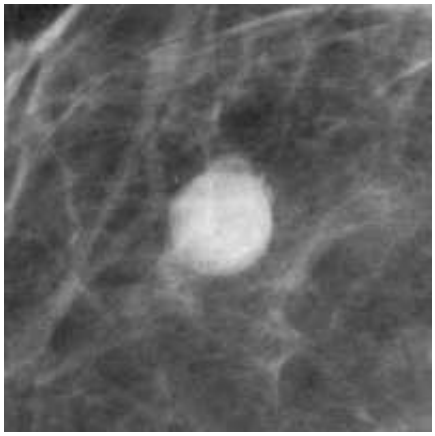
There are some methods to guarantee the contour to be closed. One of the methods is to calculate the optimal path for each radius r under the constraint that the start and the end point are equal. This is accomplished by adding extra cost to all points in the first column of the cost matrix except to point $(r,0)$. Then the optimal path is found by back tracing the path from the endpoint in row r. The extra cost force the path to end at point $(r,0)$. The cumulative cost associated with this path is $C_r(r,\pi)$. For each radius the optimal path and the cost $C_r(r,\pi)$ is calculated. Then from all radii the radius that belonged to the path with the lowest cost C_r is selected. This path represents the final contour. A disadvantage of this method is that the algorithm has to be applied once for each radius which makes it computationally expensive.

We designed a more efficient method to ensure that the resulting contour is closed. Our solution uses an extended cost matrix where costs are plotted in an interval from $-\beta\pi$ to $\beta\pi$. The extension factor β determines the size of the extended cost matrix relative to the original cost matrix. The dynamic programming algorithm is used to find the optimal path in this extended cost matrix. The path from $-\pi$ to π is extracted from the extended cost matrix. This path represents the final contour. In the original cost matrix the final contour depended strongly on the the initial angle of the polar coordinate transform and the resulting segmentation could be different for different initial angles. A disadvantage of this dependency is that image features near the boundaries of the interval from $-\pi$ to π could have undesired effects on the resulting contour. In the new method the dependence on the choosing initial angle is minimized and consequently the image features near the boundaries of the interval have less effect on the final contour. The avoidance of the discontinuities at $-\pi$ and π also leads to more closed contours.

To determine the efficiency of this method we set up the following experiment. For each extension factor β we applied the dynamic programming algorithm to find the optimal contour. Afterward, we calculate for each extension factor the percentage of closed contours. The algorithm is considered efficient if the percentage of closed contours is nearly 100% for a small extension factor.

2.2 Region Growing

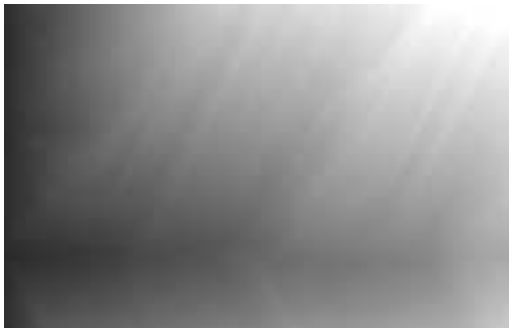
Region growing is one of the most popular segmentation methods. Region growing takes an image and a seed point (μ_x,μ_y) as input. The seed point (μ_x,μ_y) is defined to be within the suspect lesion $\mathcal{L}$. Region



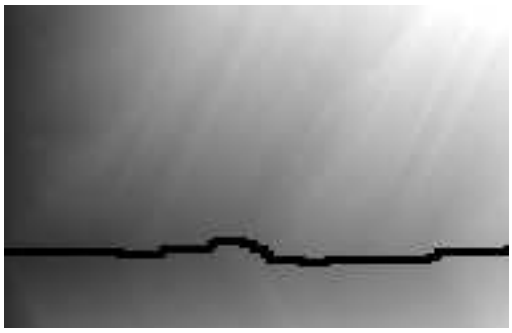
(a) Original ROI

(b) Polar ROI

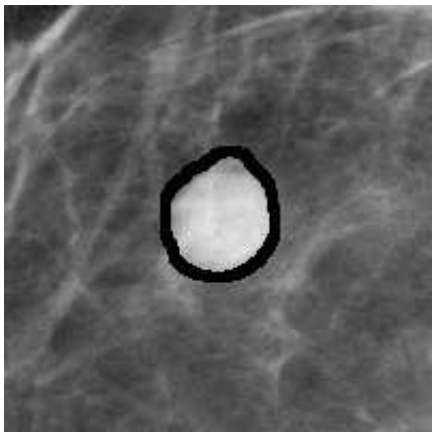
(c) Cost matrix



(d) Cumulative cost matrix

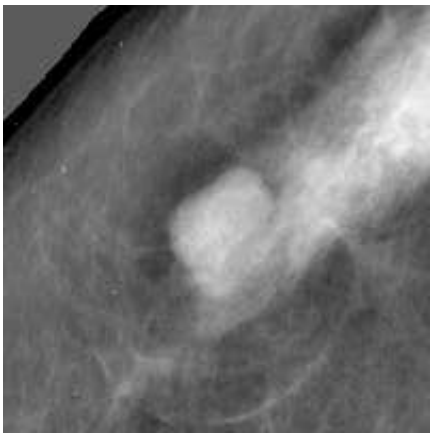


(e) Path found by dynamic programming

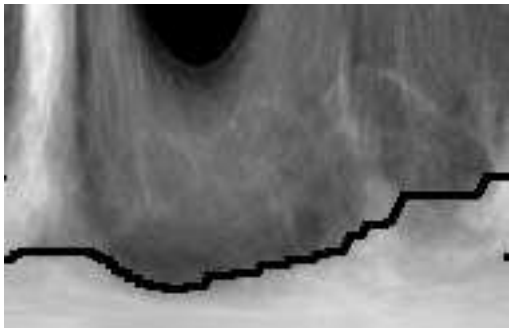


(f) Segmented mass

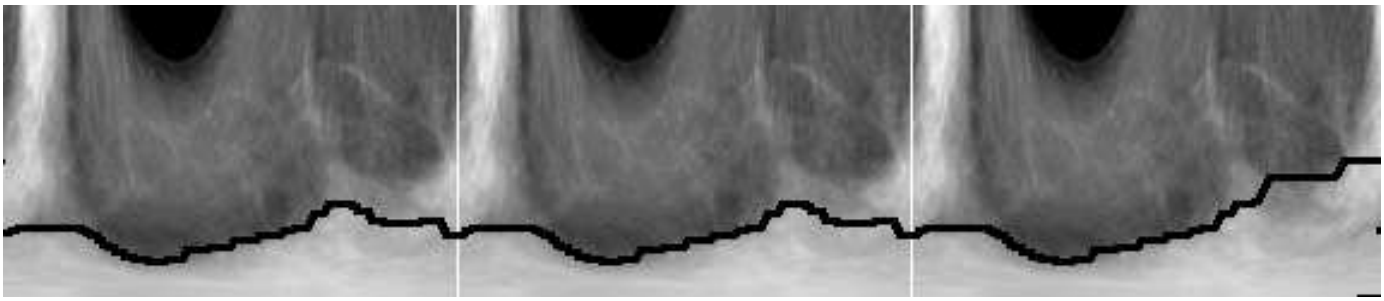
Figure 3: Implementation of the proposed dynamic programming algorithm for the segmentation of masses. 3(a) shows the ROI (region of interest) with a benign mass, 3(b) the ROI in polar coordinates. The cost of all pixels form the cost matrix (3(c). The dynamic programming algorithm calculates the optimal path in the cumulative cost matrix (3(d) en 3(e). The final contour is shown in 3(f).



(a) Original image



(b) Polar image from $-\pi$ to π



(c) Polar image from -3π to 3π

(d) Contour as extracted from the polar image from $-\pi$ to π

(e) Contour extracted from the extended polar image

Figure 4: The original dynamic programming segmentation algorithm does not guarantee a closed contour. This problem is illustrated with the mass in 4(a). Without the constraint of a closed contour the optimal path is attracted toward some dense tissue and the resulting contour is not closed. 4(b) shows the polar image with the contour plotted on top of the image and 4(d) and the resulting contour. The new algorithm extracts the contour from an extended cost matrix. 4(c) shows the extended contour plotted on top of the polar ROI and 4(e) shows the final contour.

growing then grows the seed regions in an iterative fashion. At each iteration the pixels that border the growing regions are examined. Conventional region growing defines several lesion partitions $\mathcal{L}_i$ based solely on gray level information in the image

$$\mathcal{L}_i^{(rg)} = \{G(i,j) > t_i\}$$

where G(i,j) is the pixel gray level and t_i is the gray level threshold. For each partition features are calculated such as circularity and size. Based on these features the partition that best characterizes the lesion is selected as the final contour.

We implemented an extended version of the algorithm developed by Kupinski and Giger [4]. In this method the partitions are created using gray level information as well as prior knowledge about the shape of typical mass lesions. To include information about the shape of the lesion, the region is pre-processed by multiplication with a Gaussian centered at (μ_x, μ_y). The partitions returned by thresholding are now more compact than before because distant pixels are suppressed. To determine which partition best delineates the lesion a likelihood measure was used. This measure estimates the gray level distribution for gray levels inside and outside the lesion for each partition. The partition that maximizes this likelihood is selected as the final contour.

2.3 Discrete contour model

The active contour model (or snake) formulates the boundary detection issue as an energy function minimization problem [8]. We implemented a discrete version developed by Lobregt and Viergever, the discrete contour model [9]. Starting from an initial shape the discrete contour model actively modifies its shape approximating some desired contour. The driving force behind the shape deformation is calculated from internal forces, derived from the shape of the contour model itself, and external forces, derived from some image feature energy distribution.

The internal force is based on the local shape of the contour, and aims at minimizing local curvature. Local curvature for vertex v_i is defined as the difference between the directions of the two edge segments that join at that location.

The external force is based on the image gradient magnitude. This force moves the vertices to locations in the image with strong gradients: the edges of the mass. Computation of the external force is done in the radial direction, as this prevents vertices from moving along the contour. The total force f_i acting on a vertex is a weighted combination of external and internal forces. As a result of this force the vertex v_i will start to move and change its position p_i. This position vector, together with the vertex velocity and acceleration vectors v_i and a_i, describes the dynamic state of the vertex. The deformation process stops if the condition $v_i = a_i = 0$ is met by all vertices. To ensure stability of the deformation process, a small damping factor is added to the force that is applied to vector v_i.

3 Experiments

Four experiments were done to evaluate the performance of the dynamic programming method. The first experiment was done to estimate the value of the extension factor needed to guarantee a closed contour in the dynamic programming method. The second experiment quantitatively analyzed the segmentation performance of the new method compared with the other two methods – region growing and the discrete contour model – using an overlap criterion. The other two experiments were conducted to evaluate effects of the segmentation performance on the classification accuracy. The third experiment was done to study the ability to discriminate malignant lesions from suspiciously looking normal tissue (mass detection). In the fourth experiment the ability to classify lesions correctly as either benign or malignant was studied (benign/malignant classification).

3.1 Database

The mammograms used in this study all came from the Dutch Breast Cancer screening Program. All women aged 50-70 are invited biannually to participate in this program. Two mammographic views – mediolateral oblique (MLO) and craniocaudal (CC) – are obtained at the initial screening in this program, where only mediolateral views are obtained at subsequent screenings, unless there is an indication that

additional craniocaudal views would be beneficial. The mammograms were digitized with a Canon laser scanner at a pixel resolution of $50\mu m$ x $50\mu m$, and were averaged to a resolution of $200\mu m$ maintaining the original gray value resolution of 12 bits.

The total dataset consisted of 1427 two view and four view mammograms, resulting in a total of 4295 images. Images with only microcalcifications were excluded for further research. The remaining set consisted of 1152 images with at least one biopsy proven mass (mass dataset) and 2822 normal images without pathology (normal dataset). The mass dataset contained a total 1210 masses of which 551 were malignant and 659 were benign. The set included spiculated, circumscribed and vague masses, ranging from obvious to very subtle. The resulting 1210 masses have been manually segmented by an expert radiologist. The radiologist used specially designed software to outline the masses on a dedicated mammographic review station. We used these annotations as the ground truth for our experiment. The center of mass of this annotation (μ_x, μ_y) was used as a reference point for our segmentation method. The detection experiment was done with the set of malignant masses and the normal dataset. The mass dataset was used in the other three experiments.

3.2 Extension factor for closed contours

The first experiment was done to estimate the value of the extension factor needed to guarantee a closed contour. As the original dynamic programming algorithm does not guarantee resulting contours to be closed, we proposed a solution that makes use of an extended cost matrix. The extension factor defines the size of the extended cost matrix. To estimate an appropriate value for the extension factor we calculated the percentage of closed contours for each extension factor for all images that contained at least one mass lesion.

3.3 Segmentation

The segmentation performance is evaluated for all three segmentation methods using an area overlap criterion. The used dataset consisted of all images that contained at least one mass lesion (mass dataset). All three segmentation methods need a seed point. As seed point for the algorithms we used the center of mass of the ground truth. For the discrete contour model, a circular region with radius 0.6 cm was used to initialize the algorithm. The performance of the segmentation methods has been evaluated by an overlap criterion $O = S \cap T / S \cup T$, where S is the segmented area and T is the segmentation made by the radiologist. An overlap percentage close to one means a good match between the two regions. The two-sided Wilcoxon test with confidence level 0.95 test was used to asses the difference in overlap percentage between two segmentation methods.

3.4 Mass detection

The third experiment was done to study the influence of the segmentation method on the detection of masses. First a pixel level mass detection method was used to assign a value of suspiciousness to all pixels in the image. The most suspicious sites were selected for further processing ([15], [16]). If a selected site was found closer than 1 cm to another selected site, it was considered to belong to the same suspicious region and the least suspicious site was removed. A lesion was considered detected if a selected site was located inside the manual segmenation. The method was applied at a high sensitivity level to detect most of the masses in the set. The average number of selected sites for each image was 10.

The coordinates of the sites were used as seed points (μ_x, μ_y) for the segmentation algorithms. After segmentation of the suspicious regions, features were calculated for each region to classify the region as normal or malignant. The dataset used for this experiment consisted of all normal images and all images that contained at least one malignant lesion.

Classification was done in two independent steps, feature selection and classification. Both procedures are completely independent. Cross-validation was used to randomly partition the dataset into a trainset and a testset on a 10:1 ratio under the constraint that the images from the same patient are grouped into the same subset. For every partition the training set is used for feature selection and the test set is used for the classifier validation. Every image is grouped into one of the 10 different test sets.

For feature selection we used a k-nearest neighbor (KNN) algorithm in a leave-one-out basis to select the most useful features from the entire feature space. Features were selected from each training set in

a sequential forward procedure. New features were included if they increased the performance of the classifier. The classifier was then applied to all regions in the testset. In that way every lesion in each view received a probability of malignancy. Results are presented in FROC curves in which the true positive fraction is plotted as a function of the average number of false positives per image.

3.5 Benign/malignant classification

The set of segmented masses was used to study classification performance. For each segmented mass several features were calculated including spiculation measures, contour features, texture features, and some simple morphological features. These features are described in more detail in [15],[16] and [17].

Classification was performed as above: KNN based feature selection and classification. Evaluation was performed in two ways, film based and region based. In the film based method, the likelihood of malignancy provided by the classifier was used as a discrimination score to classify the lesion as malignant or benign. In the region based method, the purpose is to classify each region as malignant and benign using the information from all available images. The discrimination score for each case is the maximum of the discrimination scores of the lesions belonging to it. This discrimination score was defined as the average of the classifier output for all different views of a lesion. In both evaluation methods, the discrimination scores were analyzed using ROC methodology using the LABROC program [18]. To evaluate the statistical significance between the different methods the CLABROC program was used.

4 Results

4.1 Percentage of closed contours

To determine an optimal value for the extension factor β we applied the dynamic programming algorithm to the region from $-\beta\pi$ to $\beta\pi$ for several values of β. For each extension factor β we calculated the percentage of closed contours. A contour was considered closed if the distance between the first and the last pixel was less than 3 pixels, according to the definition of the cost function. The result is shown in figure 5. About 40% of the contours is immediately closed. This percentage increased and reached 98% for extension factor 2. The maximum number of closed contours was reached for extension factor 3. At that time there was only one image that was not closed according to our definition. Considering these results it seems appropriate to apply the algorithm with extension factor 3 and use an extended cost matrix from -3π to 3π. The contour from $-\pi$ to π should be extracted and defines the final contour. For the one mass that was not closed with this method we applied the first method (section 2.1.2) to find a closed contour. There was one image that had two optima. The final contour depended on the initial angle of the coordinate transform and alternated between the two states for the different values of β. Figure 6 shows the two cases where the dynamic programming algorithm was unable to find a closed and stabile contour. Figure 6(a) shows the manual segmentation of a benign mass embedded in dense tissue. The two states of the contour are given in figure 6(b) and 6(c). The contour in figure 6(b) is too large and contains some dense tissue around the mass. The other state, figure 6(c), gives a correct segmenation. The case where the dynamic programmig algorithm was unable to reach closure is shown in figure 6(d). The resulting segmentation is shown in figure 6. For this mass we applied the first method to obtain a closed contour. This contour is shown in figure 6(f).

4.2 Segmentation performance

For each mass the overlap percentage with manual segmentation was calculated. The average overlap percentage for dynamic programming was 0.69, for discrete contour model 0.60 and for region growing 0.58 (table 1 This indicates that the dynamic programming method is more suited to segment mammographic masses than the other two methods. Figure 7 displays the overlap percentages for the different methods. The figure shows that not only the mean overlap percentage is higher, but also that the percentage of masses with poor overlap is smaller for dynamic programming than for the other two segmentation methods.

Table 2 shows the results of the Wilcoxon statistic. The difference in overlap percentage between the proposed method and the other two methods is statitically significant (p-value ¡¡ 0.05). It is clear from

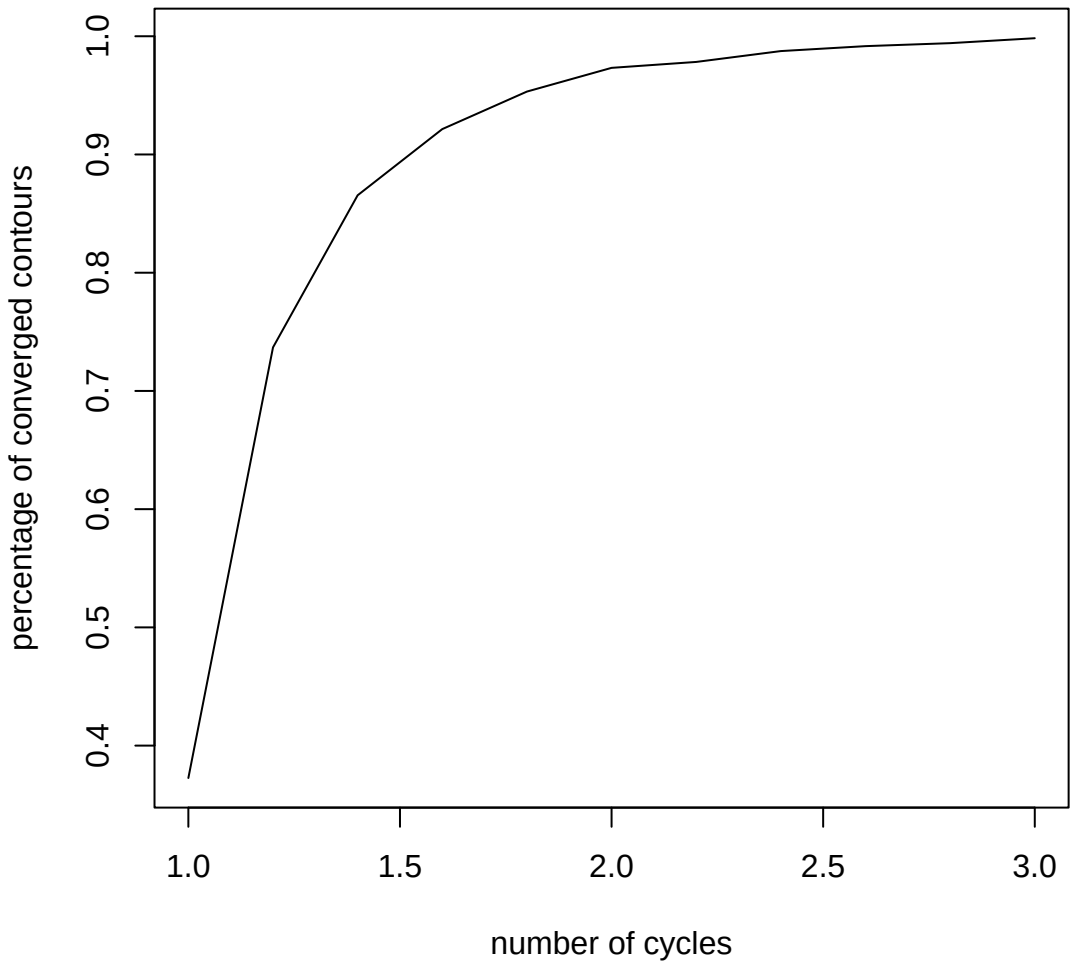
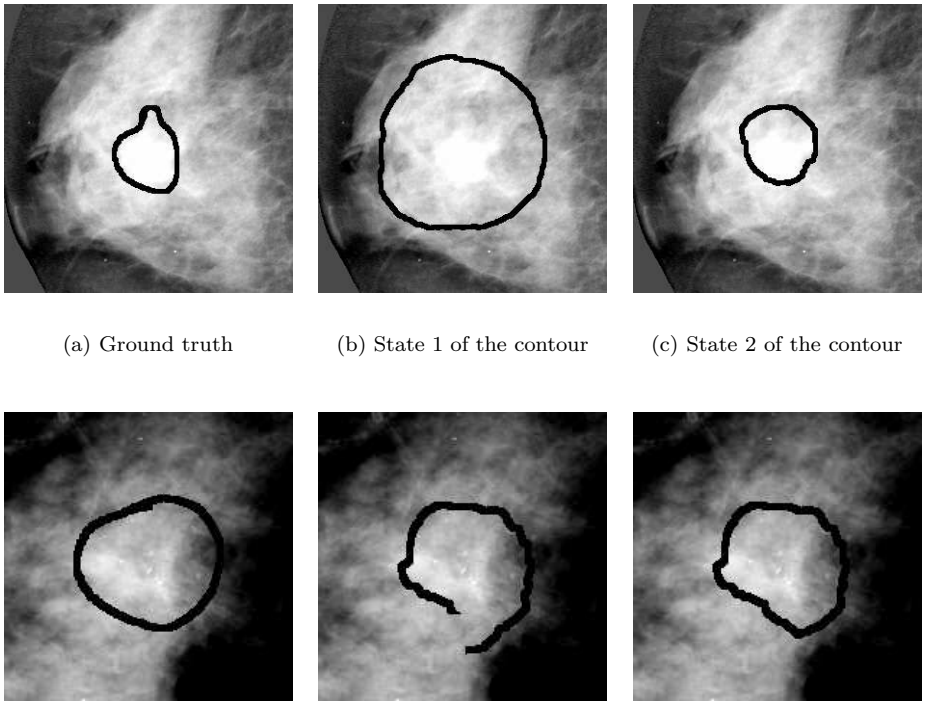


Figure 5: The percentage of closed contours is plotted against the extension factor β. In the original dynamic programming algorithm (with $\beta = 1.0$) 40% of the contours is closed. In the improved algorithm where the path is calculated over a larger area (extension factor 3.0), more then 99% of the contours is closed.



(a) Ground truth (b) State 1 of the contour (c) State 2 of the contour

(d) Ground truth (e) Contour found by dynamic programming (f) Contour found by the original algorithm

Figure 6: Some images where the dynamic programming algorithm did not succeed in finding a closed and stabile contour. The first set of images show the case where the final contour alternated between two states (6(b) and 6(c)). The second set of images shows the case where the dynamic programming algorithm was unable to find a closed contour. Figure 6 shows the manual segmenation, 6(e) the optimal path found by dynamic programming and 6(f) the segmentation obtained by the original algoritm with constraints.

the test that a significant increase in area overlap has been obtained with the dynamic programming method compared with the other two segmentation methods. Although the discrete contour model had better results than the region growing method, these results were not statistically significant.

Method	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
dynamic programming	0.005	0.608	0.747	0.687	0.825	0.940
discrete contour model	0.006	0.494	0.633	0.599	0.743	0.913
region growing	0.034	0.460	0.641	0.586	0.748	0.914

Table 1: Summary statistics for the performance of the three segmentation methods based on an area overlap criterion measuring the overlap between the automated segmentation and the manual segmentation.

Method	W+	p-vale	conf interval
dynamic programming - discrete contour model	519821	$<< 0.05$	0.081 – 0.11
discrete contour model - region growing	362020	0.6507	-0.01 – 0.01
dynamic programming - region growing	517192	$<< 0.05$	0.08 – 0.11

Table 2: Results of the Wilcoxon's test for the statistical difference in overlap percentage between the existing methods and the proposed method. The second column displays the sum of the positive ranks, the third column gives the p-value and the last column the 95% confidence interval for the difference in the means.

4.3 Mass detection performance

Results of the mass detection experiment are shown in figure 8 and 9. In FROC curves, horizontally the number of false positive detections per image is shown, vertically the sensitivity that is achieved at this specificity level. A tumor was considered detected if the center of a detection was inside the segmentation made by the radiologist. Detections outside the annotated areas were counted as false positive signals. The film based performance is given in figure 8. The case based performance is given in figure 9. In the case based evaluation a lesion is considered detected if it is detected on either view. Both figures show that the detection performance for the segmentation methods is nearly identical.

4.4 Benign/malignant classification accuracy

The second experiment studied the ability to discriminate between benign and malignant lesions for the three segmentation methods. The average classification accuracy has been determined by calculating the area under the ROC curve (A_z) using the LABROC program [18]. The average test results for film based performance corresponded to an area under the ROC curve of 0.67 for region growing, 0.71 for the discrete contour model and 0.74 for dynamic programming. The average A_z values for a region based evaluation, where different views of the same lesion were combined by the classifier, were 0.74 for dynamic programming, 0.72 for the discrete contour model, and 0.67 for the region growing method. The region based performance is shown in figure 11. We used the CLABROC program to evaluate the statistical significance of differences in classification performance. We found that the difference in A_z values between the region growing and the dynamic programming method was statistically significant (two tailed p-level 0.0044). No statistical significance was found between the other methods. The two tailed p-level for dynamic programming verus the disrete contour model was 0.20, for the discrete contour model versus regions growing 0.08. The results are summarized in table 3.

5 Discussion/Conclusion

In this work, we developed a segmentation method based on dynamic programming to more accurately extract mammographic mass contours. In addition, we developed a method to obtain closed contours.

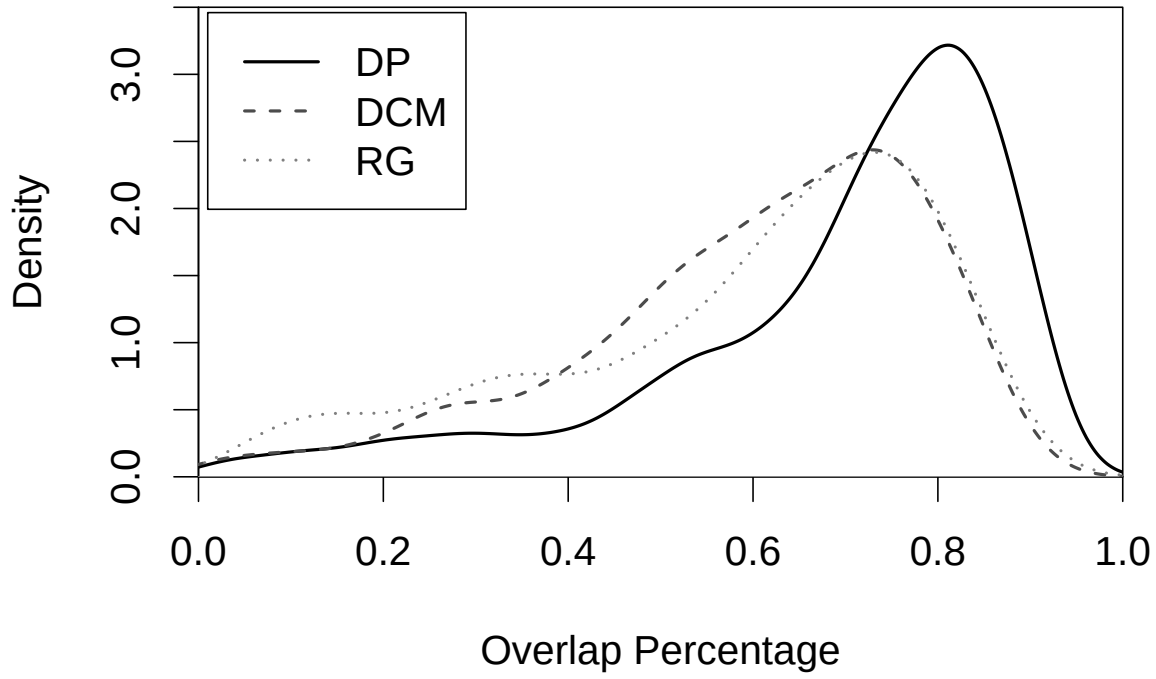


Figure 7: Distribution of the overlap percentages for the different segmentation methods. The average overlap percentage for dynamic programming (DP) is higher then for the other segmentation methods (regions growing (RG) and the discrete contour model (DCM))

Methods	film based A_z	region based A_z	p-value
dynamic programming	0.72	0.74	
discrete contour model	0.71	0.72	0.20
region growing	0.67	0.67	0.0044

Table 3: Film based and region based performance for classification of malignant versus benign masses. Masses were segmented with the proposed dynamic programming method, the region growing method and the discrete contour model. The last column gives the results of the tow-tailed Student t test for the differences in areas under the ROC curve between the proposed method and the existing methods.

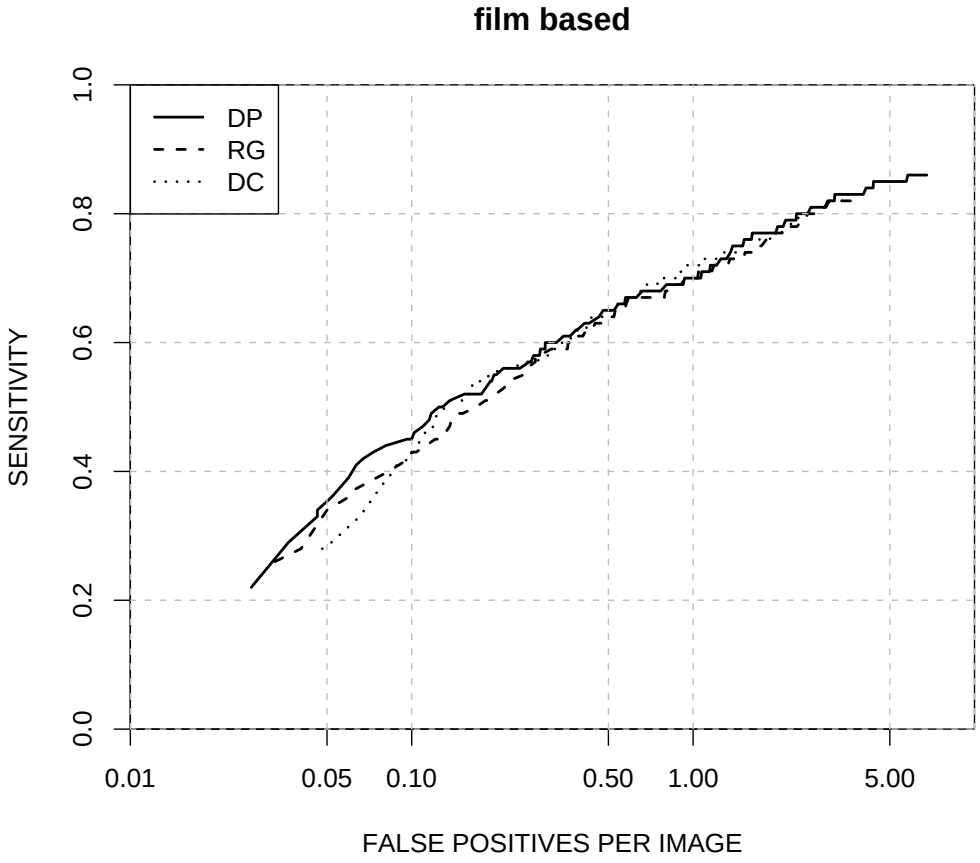


Figure 8: FROC curve film based

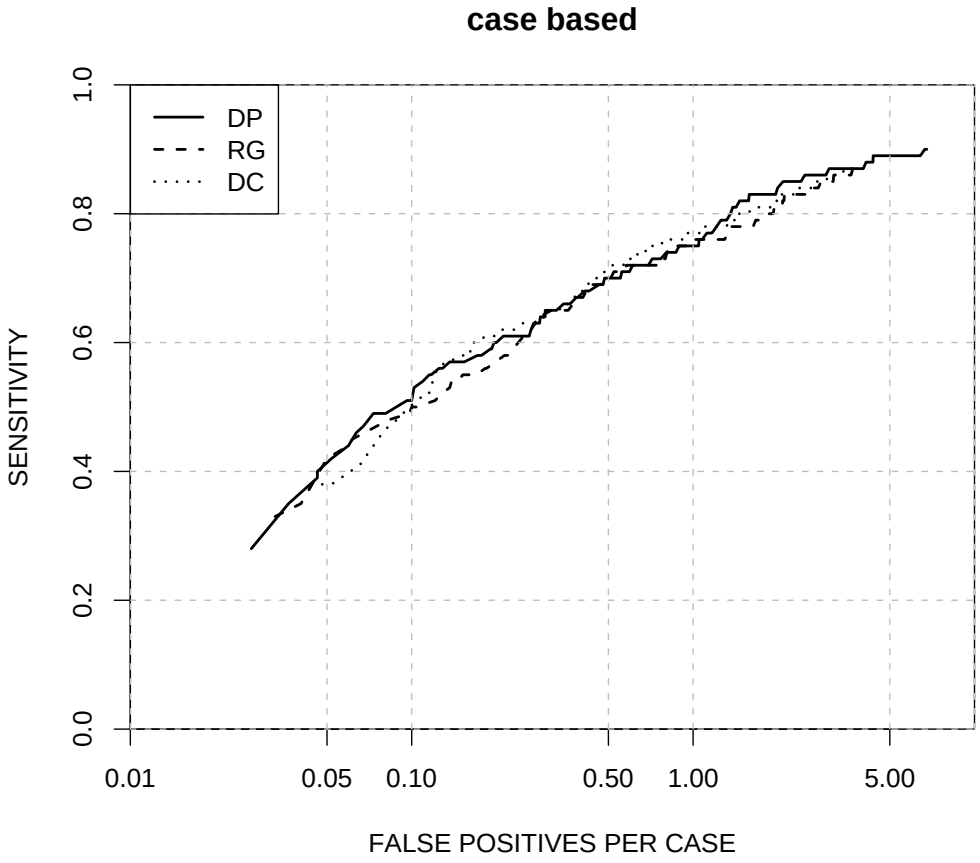


Figure 9: FROC curve case based

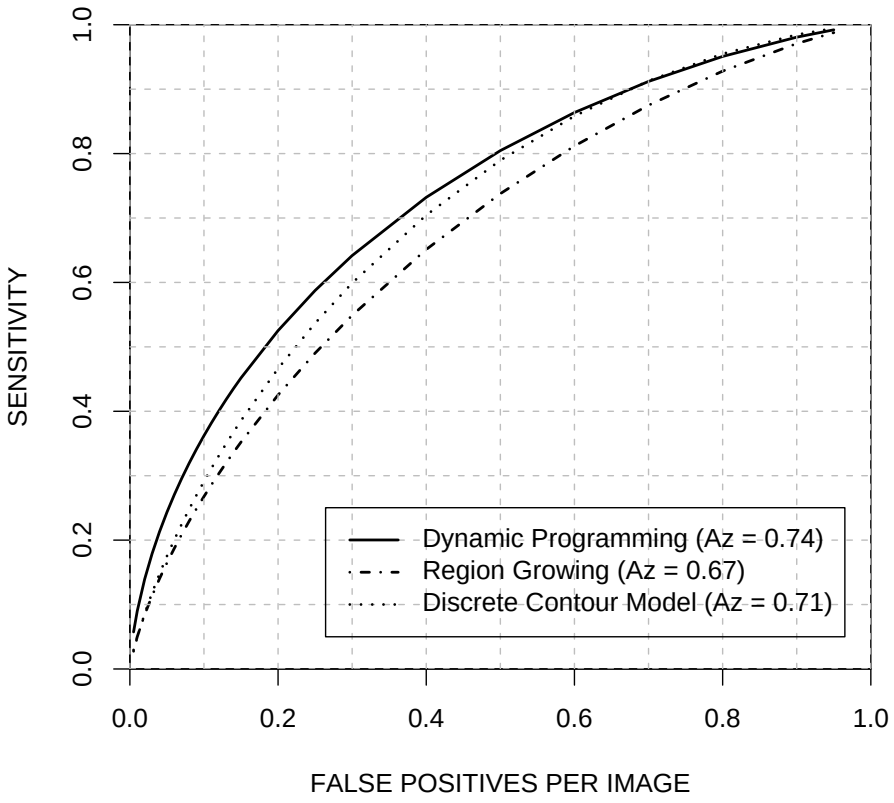


Figure 10: Film based performance for classification of malignant versus benign masses. Masses were segmented with the proposed dynamic programming method, the region growing method and the discrete contour model.

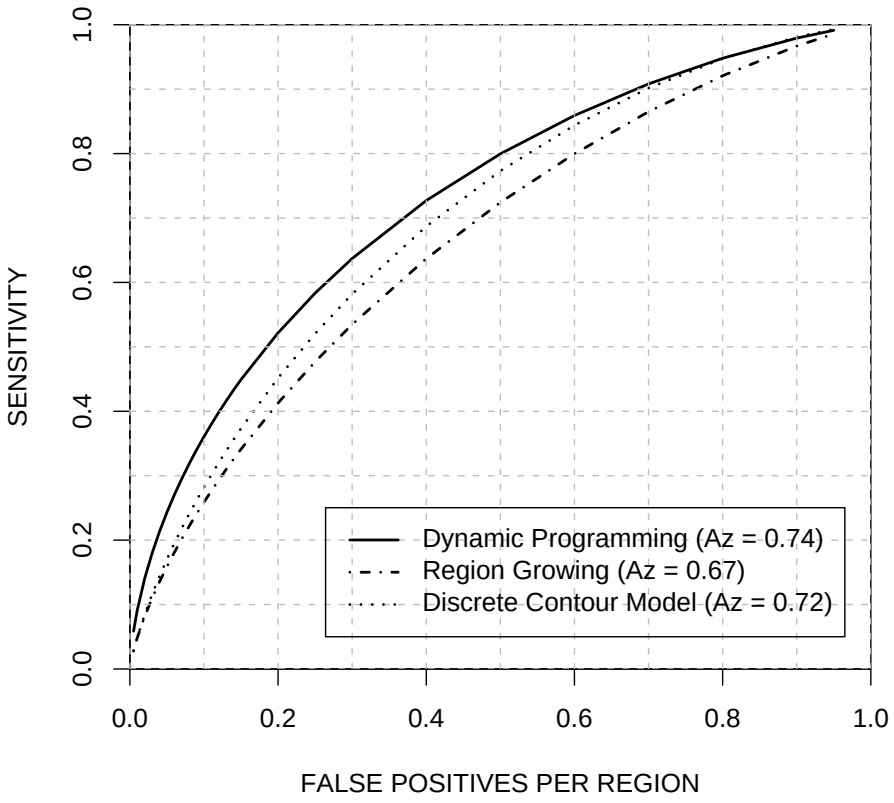


Figure 11: Region based performance for classification of malignant versus benign masses. Masses were segmented with the proposed dynamic programming method, the region growing method and the discrete contour model.

We found that 99.9% of the contours was closed and stabile for an extension factor of 3.0. As this percentage did not change for larger values of beta, we choose to use a beta of 3.0 in our algorithm. Another option would be to use a beta of 2.0 and increase beta to 3.0 for the contours that were not closed. A disadvantage of this method is that the algorithm will also find a number of closed contours, that have not converged yet. We consider these contours as suboptimal. It depends on the application to make a balance between optimality and speed. In our application, as the computational burden of applying the algorithm with beta 3.0 is minimal, we choose for optimality.

The proposed method has been compared with two other methods: region growing and the discrete contour model. The accuracy of the methods have been determined by comparing the mass segmentation with the manual outline of the contour drawn by an expert radiologist. The mean overlap percentage for dynamic programming was 69%, for the other two methods 60% and 58% respectively. These differences were statistically significant ($p << 0.05$).

In a case review we found that in general segmentations with an overlap percentage of at least 70% were visually acceptable. The percentage of cases with a reasonable overlap percentage, i.e. more than 70% is 52% for dynamic programming, 39% for region growing and 38% for discrete contour model. These results demonstrate that the segmentations obtained with dynamic programming, more closely match manual segmentation than the other two automated segmentation methods. All three methods rarely achieved more than 90% overlap. One reason for this is that the accuracy of the manually segmented regions is limited. Often the manual segmentations are somewhat large to make sure the whole tumor is inside the annotation. Another reason is that when the radiologists' annotation and the automatically segmented area are not identical the chosen overlap criterion quickly decreases.

In this study we only had one radiologist to do the manual segmentations. If another radiologist would have done the segmentations, both the ground truth and the center of mass would change. Small changes in the seed point will give similar segmentation results. Both region growing and the discrete dynamic contour model are not very sensitive for small changes in seed point ([4], [9]). To determine the sensitivity of the dynamic programming method for changes in seed point, we applied the method with two different types of seed points: the center of mass of the radiologists' segmentation and the most suspicious site in a neighborhood of the center of mass. The results for both seed points were similar.

Several measures were taken to minimize uncertainty of the ground truth due to intra-observer variation. All manual segmentations were done with special software on a dedicated mammographic review station. The radiologist was given clear instructions how to outline certain mass types. For architectural distortions and spiculated masses, only the central tumor was annotated, not the spiculations. The main reason for this was that the outlining of spicules is very subjective. By taking these measures we expect that the inter-observer variability will be much smaller then the differences between our automated methods and the radiologist's segmentation. Figure 7 shows the overlap percentages for the three methods. There is a considerable number of cases where the overlap percentage is less then 50%. In a case review we found that in cases where the overlap percentage is lower than 50%, the segmentation method often failed to find the right contour. Examples are shown in figure 2(c), 2(d) and 1(c), 1(d). In our experience segmentation differences caused by inter-observer variability will be of a more subtle nature, and mainly concern slight variations in outlining tumors with vague boundaries and architectural distortions. Therefore, we believe that our results would be rather unaffected by variability of the manual segmentations.

Figures 1 and 2 show examples where the two existing methods fail to find the contour of the mass. In the first figure the contour is attracted toward the strong edge of the pectoral muscle. In the second figure the contrast of the mass is low and the resulting segmentation is too large. In both cases the dynamic programming method correctly segments the mass. The advantage of the dynamic programming method is that both global and local cost are combined with different weights to determine the optimal contour.

The detection experiment showed that the improvement in segmentation performance did not result in a better detection of malignant masses. This may be understandable as we did not yet focus our attention on the design of special contour features for mass detection. Instead we used the contour features that were developed for benign/malignant classification. These contour features in particular measure the sharpness of the mass boundary which is a good feature in discriminating between benign and malignant masses. The boundaries of malignant masses and patches of normal tissue that are initially detected by our method, often display more similar boundary characteristics. Further studies are needed to develop contour features that capture relevant aspects of the mass boundaries to discriminate between malignant masses and false positive detections.

The proposed dynamic programming method improved the classification of malignant versus benign masses. When using the KNN-classifier, the A_z value for region based performance was 0.74 for the proposed segmentation method, 0.67 for the region growing method and 0.72 for the discrete contour model. The difference in A_z values was between the proposed method and the region growing method was statistically significant. Differences between other methods were not statistically significant. The classification results are in agreement with the ranking of the segmentation results. Region growing, the method with the lowest segmentation performance - as measured with the area overlap criterion - showed the lowest classification performance. The best segmentation method - the dynamic programming algorithm - also performed best in classification.

These results contradict the study from Sahiner et al ([12]). They compared a mass segmentation method based on an active contour model with manual segmentation by two expert radiologists. Even when the radiologist and the computer had high disagreement they observed no difference in classification accuracy. Our experiments suggest that a more accurate segmentation may result in improved classification of benign versus malignant masses. With better contour features, we expect that the mass detection performance will also benefit from a more accurate segmentation.

6 Conclusion

In this study we compared a new segmentation method based on dynamic programming. The new segmentation algorithm turned out to be robust and accurate with a high overlap in a large number of cases. The mean overlap percentage for dynamic programming was 69%, for discrete contour model 60% and for region growing 58%. The differences between the new method and the existing methods were statistically significant. This improvement in segmentation performance resulted in better benign/malignant classification. The detection performance was nearly identical for the three methods.

We will concentrate our future work on improving our contour dependent features, especially contour features that discriminate between suspiciously looking normal tissue and malignant masses. We expect that our CAD program will benefit from a very accurate segmentation method in combination with these new features.

References

[1] N. Karssemeijer, J.D. Otten, and A.L. Verbeek et al. Computer-aided detection versus independent double reading of masses on mammograms. *Radiology*, 227:192–200, 2003.

[2] Y.L. Jiang, R.M. Nishikawa, R.A. Schmidt, C.E. Metz, M.L. Giger, and K. Doi. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiology*, 6(1):22–33, 1999.

[3] H-P. Chan, B. Sahiner, M.A. Helvie, N. Petrick, M.A. Roubidoux, T.E. Wilson, D.D. Adler, C. Paramagul, J.S. Newman, and S. Sanjay-Gopal. Improvement of radiologists's characterization of mammographic masses by using computer-aided diagnosis: An ROC study. *Radiology*, 212:817–827, 1999.

[4] M.A. Kupinski and M.L. Giger. Automated seeded lesion segmentation on digital mammograms. *IEEE Trans Med Imaging*, 17(4):510–517, Aug 1998.

[5] D. Guliato, R.M. Rangayyan, and W.A. Carnielli et al. Segmentation of breast tumors in mammograms by fuzzy region growing. In *Proc. 20th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, pages 1002–1004, 1998.

[6] D. Guliato, R.M. Rangayyan, J.A. Zuffo, and J.E. Leo Desautels. Detection of breast tumor boundaries using iso-intensity contours and dynamic thresholding. In N. Karssemeijer, M.A.O. Thijssen, J.H.C.L. Hendriks, and L.J.T.O. van Erning, editors, *Digital Mammography*, pages 253–260. Kluwer, Dordrecht, 1998.

[7] N. Petrick, H-P. Chan, B. Sahiner, and M. Helvie. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. *Med Phys*, 26(8):1642–1654, 1999.

[8] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Med Image Anal*, 1(2):91–108, Jun 1996.

[9] S. Lobregt and M. Viergever. A discrete dynamic contour model. *IEEE Trans Med Imaging*, 14:12–24, 1995.

[10] G.M. te Brake, M.J. Stoutjesdijk, and N. Karssemeijer. A discrete dynamic countour model for mass segmentation in digital mammograms. In *Proc SPIE Medical Imaging*, volume 3661, pages 911–919, 1999.

[11] S. Timp, N. Karssemeijer, and J.H.C.L. Hendriks. Comparison of three different mass segmentation methods. In H. Peitgen, editor, *IWDM proceedings 2002*, pages 218–222. Springer-Verlag, 2002.

[12] B. Sahiner, N. Petrick, and H-P. Chan et al. Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. *IEEE Trans Med Imaging*, 20(12):1275–1284, Dec 2001.

[13] S. Timp, N. Karssemeijer, and J.H.C.L. Hendriks. Analysis of changes in masses using contrast and size measures. In H. Peitgen, editor, *IWDM proceedings 2002*, pages 240–242. Springer-Verlag, 2002.

[14] E. Claridge and J.H. Richter. Characterisation of mammographic lesions. In A.G. Gale et al., editor, *Digital Mammography*, pages 241–250. Elsevier, Amsterdam, 1994.

[15] G.M. te Brake and N. Karssemeijer. Single and multiscale detection of masses in digital mammograms. *IEEE Trans Med Imaging*, 18:628–639, 1999.

[16] N. Karssemeijer and G.M. te Brake. Detection of stellate distortions in mammograms. *IEEE Trans Med Imaging*, 15:611–619, 10 1996.

[17] C. Varela, N. Karssemeijer, J.M. Muller, and P.G. Tahoces. Classification of breast tumors in digitized mammograms. In *IWDM proceedings 2002*, Bremen, 2002.

[18] C.E. Metz, B.A. Herman, and J.H. Shen. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med*, 17(9):1033–1053, 1998. Available http://www-radiology.uchicago.edu/krl/rocstudy.htm.