FUZZY SEGMENTATION OF MASSES IN DIGITAL BREAST TOMOSYNTHESIS IMAGES BASED ON DYNAMIC PROGRAMMING

Louis Apffel[†], Giovanni Palma[‡§], Serge Muller[‡] and Isabelle Bloch[§]

[†] *Ecole Centrale Paris, Grande Voie des Vignes F-92 295 Châtenay-Malabry Cedex, France*
[‡] *GE Healthcare, 283, rue de la Minière, 78530 Buc, France*
[§] *Télécom ParisTech - CNRS LTCI, 46, rue Barrault, 75013 Paris, France*
louis.apffel@centraliens.net, {giovanni.palma, serge.muller}@ge.com, isabelle.bloch@enst.fr

Abstract: In this paper we propose a new fuzzy segmentation method to segment lesions in Digital Breast Tomosynthesis (DBT) datasets. In the proposed approach we model a contour as a path in the image. The optimal contour is defined as the path associated with a minimal cost, which is derived from the image content. Using this formalism we present several ways to alter this cost in order to extract several relevant contours from a single image. The set of contours is then used in the fuzzy contour framework to perform mass detection. The method has been tested on synthetic data as well as images containing lesions and provides promising results.

1 INTRODUCTION

Digital Breast Tomosynthesis (DBT) is a new 3D imaging technique (Dobbins III and Godfrey, 2003; Wu et al., 2003) that may potentially overcome some limitations of standard mammography such as tissue superimposition (Gennaro et al., 2008). Such an improvement is gained at the expense of an increased amount of data to be reviewed by the radiologist. In this context the design of a Computer Aided Detection (CAD) system may help the radiologist to keep a high sensitivity in his detection and characterization task. Segmentation is usually a crucial step in CAD systems, and a wrong segmentation can have a disastrous impact on the whole detection scheme. The use of fuzzy sets avoids these drawbacks (Peters, 2007). This can be done by using the fuzzy contour framework, which is suitable to represent several possible contours for one specific structure. Unfortunately, only a small amount of techniques exist in the literature (Bothorel et al., 1997; Peters, 2007; Palma et al., 2008) to extract such fuzzy contours and they suffer from various limitations. In this paper, we introduce a new fuzzy contour extraction procedure by extending a crisp segmentation method originally proposed for standard mammography images, which models contours as paths in an image (Timp and Karssemeijer,

2004).

First, in Section 2 we present the limitations of crisp segmentation methods. Then in Section 3, we recall the formalism of fuzzy contours. In Section 4, we describe the new proposed algorithm and we discuss two aspects: uncertainty and imprecision, which are both handled by our approach. Finally we illustrate results on both synthetic and clinical data.

2 LIMITATIONS OF CRISP SEGMENTATION

In breast imaging, crisp segmentation has many limitations. Indeed, segmenting some masses may be particularly difficult when the contrast between lesion and background of the image is low. Radiologists may have different and various opinions on the exact location of a contour. A segmentation algorithm that proposes a single answer to this kind of problem is usually taking an arbitrary decision on the nature of the contour. Since the shape of the contour is suggestive of the nature of the lesion (benign or malignant), anticipating its nature from a single and inaccurate contour could lead to a classification error.

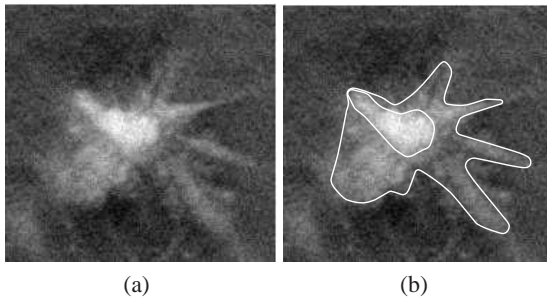The problem is illustrated in Figure 1 where a

Figure 1: Crisp segmentation problem: several contours can fit a given lesion and may lead to different conclusions on its class.

spiculated mass is presented with two different segmentation results.

3 FUZZY SEGMENTATION

The fuzzy set framework seems to be suitable to deal with the mass segmentation problem (Peters, 2007). It allows defining the concept of fuzzy contours, which can overcome several limitations of regular segmentation techniques by allowing not to make a decision too early.

Definition 1. A fuzzy contour is a set of nested crisp contours where each contour is associated with a membership degree to the class *contour*.

3.1 Imprecision and Uncertainty

In DBT images, masses may present two types of imperfections. First, a lesion may be hard to define (i.e. there are several possible contours that can fit the lesion boundaries) even if we are sure of its presence. This is called imprecision. Secondly, it may sometimes be difficult to detect and locate a mass. In this case we are dealing with uncertainty. A single fuzzy contour can handle the notion of imprecision because it considers several potential contours. On the other hand, uncertainty can be handled by considering several fuzzy contours.

3.2 Existing Schemes to Extract Fuzzy Contours

In the literature, several techniques to extract fuzzy contours have been proposed. Originally, a multi-thresholding approach was designed in order to detect microcalcifications in mammography (Bothorel et al., 1997). This approach has also been used in order to segment masses in projected images used to reconstruct DBT volumes (Peters, 2007). Unfortunately, because of potential superimpositions of tissues in projection images, this approach is not well suited for masses in 2D projection images.

More recently, a segmentation scheme based on the level set framework (Osher and Fedkiw, 2002) has been proposed (Peters, 2007; Palma et al., 2008). This approach is taking advantage of the side effects introduced on the Lipschitz function, which implicitly represents a contour, during the minimization stage of the contour energy. The idea is then to slice this function at several levels after convergence to build a fuzzy contour.

Unfortunately, tuning parameters of this segmentation approach is a pretty difficult task. In fact, the problem comes from the large variability of structures that may be encountered in mammography. Furthermore, slicing the function representing the contour at several levels is something that may be hard to interpret.

4 DYNAMIC PROGRAMMING AND FUZZY SEGMENTATION

4.1 Original Approach

Recently, an alternative to conventional approaches was proposed (Timp and Karssemeijer, 2004). In this approach, a contour is modeled as a path in an image converted into the polar domain. This path is associated with a cost that depends on the image content. This formulation allows using well known dynamic programming (DP) techniques in order to efficiently solve the problem. Several improvements have been proposed based on this approach (Rojas and Nandi, 2007).
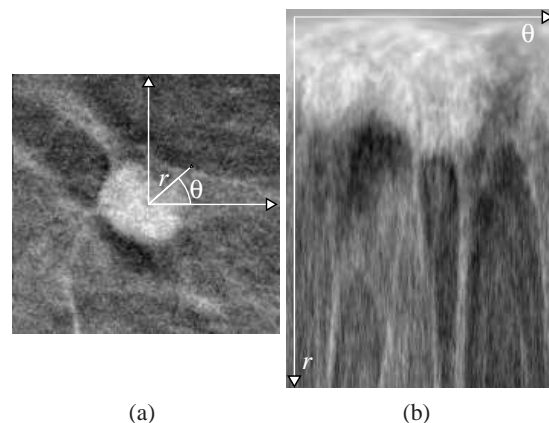

Figure 2: Polar coordinates representation (b) of a circumscribed lesion (a).

The authors originally designed this algorithm on standard mammography images. We eliminated the use of the shape coefficient they propose since the masses used in our database were of numerous shapes and sizes and the research of fuzzy contours implies contours of different radii. Concretely, the image to segment is first converted in the polar domain. The rows and the columns of the new image represent respectively the distance to the lesion center and the angle around it. This representation is illustrated in Figure 2. A cost matrix is then defined on the same domain leading to the definition of a cost function for any path linking the first to the last column:

$$C(c) = \sum_{\theta \in [[0;\theta_{max}]]} M(\theta, c(\theta))$$

with M a cost matrix, θ_{max} the index corresponding to the maximum angle (2π), and c a path, which associates any angle in $[0; 2\pi[$ with a radius. In our implementation we used the measures based on the image gradient along the radial direction and the optimal gray value of the contour, which were proposed in the original approach (Timp and Karssemeijer, 2004).

Now to introduce some smoothness in the contours we may want to consider, we can restrict the search space to the following set of contours:

$$\mathcal{P} = \{c / \forall \theta \in [[0; \theta_{max}]] \quad |c(\theta) - c(\theta - 1)| \le f\}$$

with $f \in \mathbb{N}$ the parameter allowing to tune the contour smoothness.

The segmentation problem is then solved by searching for a contour $\hat{c}$ such that:

$$\hat{c} = \underset{c \in \mathcal{P}}{\arg\min} \, C(c)$$

This method was initially designed to segment circumscribed lesions in standard mammography. Our implementation, which is similar as the original one except for the lesion size prior, has been tested on spiculated and circumscribed lesions in DBT as it will be presented in Section 5.

We will now present our main contribution over the original work by proposing means to extract fuzzy contours from an image.

4.2 Penalization

To obtain several contours, the first idea was to prevent the algorithm from using the pixels that have already been crossed by a path in the cost matrix by setting their costs to an infinite value. This operation can be iterated while non infinite cost paths exist in the cost matrix.

When looking for a path in the cost matrix, the DP-based algorithm is constrained by the parameter f along the vertical axis whose role was previously detailed. Thus, if we consider a pixel (θ, r) in the cost matrix, the path can reach $2f + 1$ pixels from one column to another. This parameter allows radial variations in the polar coordinates system described in (Timp and Karssemeijer, 2004). Thus, a simple penalization of the pixels through which goes the first path does not prevent contours from crossing each other.

Contours obtained with such a method are often similar to the first contour. To force finding a contour with a significantly different shape, it is necessary to generate a lot of contours, which might be hard to handle in further processing. The information conveyed by the repetition of quasi-identical contours is related to the imprecision contained in the lesion we want to segment. This correlation has some meaning but we may miss information related to the uncertainty implicitly present in some images.

In order to satisfy the inclusion criterion required by Definition 1, it is possible to penalize a band nearby the path found in the cost matrix. Instead of setting a one-pixel band to an infinite value, we penalize a band of $2f + 1$ pixels. Thus, the algorithm forces the contours to be nested.

4.3 Contour Selection and Elimination

In this section, we present an approach to model uncertainty rather than imprecision. The goal is to obtain a set of contours different from each other. To guarantee that the contours will be different, we define a distance on the space of contours represented in polar coordinates. Let $c_1(\theta)$ and $c_2(\theta)$ be two contours, we define the distance $d(r_1, r_2)$ as follows:

$$d(c_1, c_2) = \max_{\theta}(|c_1(\theta) - c_2(\theta)|)$$

We can now extract representative contours using the penalization technique previously presented with no need to use forbidden bands of pixels but skipping contours that are too close from each other. This is done iteratively by discarding all the contours whose distance with the former representative contour is less or equal to a given threshold. This approach allows modeling uncertainty and produces contours that are not nested but that exhibit differences for at least few points. Segmentation results obtained with this method are presented in Section 5.

4.4 Extraction of Several Fuzzy Contours

In order to capture both uncertainty and imprecision contained in a given image, we can combine the band penalization and the distance-based approaches. Obviously, we can extract for each contour c obtained with the second approach, a fuzzy contour by applying the first method on a matrix $M' = M$ where the path c is set to ∞ on a large band. Doing this enables us to have a set (uncertainty) of nested contours (imprecision).

4.5 Membership Degree

To complete the extraction of fuzzy contours, we need to assign a membership degree to each contour constituting the fuzzy set. This membership degree should be representative of the confidence we can have in the contour.

The dynamic programming algorithm provides us with a direct measurement of confidence linked to the values of the gradient along the contour and its gray levels: the total cost of the path ($\mathcal{C}$). This cost can be used to derive a membership value for the contour if we normalize it considering the whole set of contours obtained for an image. For instance, if we obtain n contours with the associated costs $C_i, i \in [[1;n]]$, we can define the membership degree μ_i as follows:

$$\mu_i = \frac{C_{\max} - C_i}{C_{\max} - C_{\min}} \qquad (1)$$

where $C_{\max}$ and $C_{\min}$ are $\max_i\{C_i\}$ and $\min_i\{C_i\}$, respectively. Thus, we get a value ranging between 0 and 1, with 1 being the smallest cost (hence highest membership value) and 0 being the highest cost.

Another type of membership degree can be defined. Since we extract a lot of contours to construct a pool from which we extract the contour with the distance method, we are able to create a map of contours density. Each time a path is derived in polar coordinates (in the cost matrix), we set to 1 the corresponding pixels in the contour map. A pixel cannot be crossed twice because of the penalization. So, we obtain a map as represented in Figure 3(a). By averaging the image vertically with a Gaussian filter, we obtain a blurred version as represented in Figure 3(b). This step allows us to take into account the spatial density of the contours, thus the resulting density map is representative of the possibility of having a contour crossing given pixels. Summing up the values of the density map D along a path enables us to obtain a representative value of the confidence we can have in the contour. The higher this value, the more confident we

can be in the contour. This total density can also be normalized to create a membership degree:
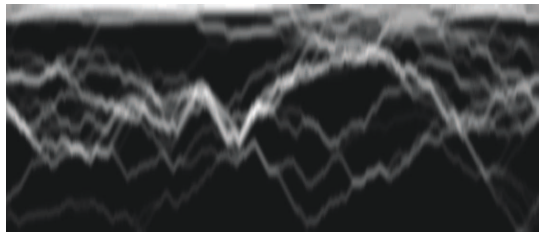
$$\mu'_i = \frac{\min\limits_{c' \in C} \sum\limits_{\theta=0}^{\theta_{\max}} D(\theta, c'(\theta)) - \sum\limits_{\theta=0}^{\theta_{\max}} D(\theta, c_i(\theta))}{\max\limits_{c' \in C} \sum\limits_{\theta=0}^{\theta_{\max}} D(\theta, c'(\theta)) - \min\limits_{c' \in C} \sum\limits_{\theta=0}^{\theta_{\max}} D(\theta, c'(\theta))} \qquad (2)$$

with C the set of extracted contours and c_i the considered contour.

The behavior of both approaches will be discussed in Section 5.
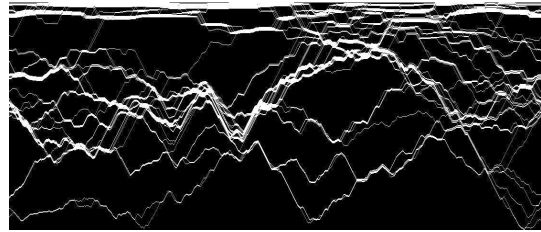
(a)

(b)

Figure 3: Computation of a contour density map. (a) Contour map extracted from an image. 100 contours have been marked on the image. (b) Density map after smoothing.

5 RESULTS

The database we used for evaluation contains 52 regions of interest extracted from DBT slices. Sixteen of them contain a circumscribed mass, while remaining ones contain a spiculated lesion. The algorithm was first validated on these images for crisp segmentation. The results are presented in Section 5.1. In Section 5.2 we interpret the results obtained with our new fuzzy segmentation approach. We focus on the results obtained with both the penalization and the distance and discuss the combination of the two approaches. Finally we compare the two ways of constructing membership degrees.

5.1 Crisp Segmentation on DBT Slices

In order to validate the approach, we performed an assessment of the crisp approach on two subsets of

lesions (spiculated and circumscribed). This assessment was done using ground truth contours built by a human reader and a criterion P, which is the average of 3 measures (Rojas and Nandi, 2007):

$$P_1 = \frac{|A \cap R|}{|A \cup R|} \qquad \text{similarity}$$

$$P_2 = 1 - \frac{|R \setminus (A \cap R)|}{|R|} \qquad \text{under-segmentation}$$

$$P_3 = 1 - \frac{|A \setminus (A \cap R)|}{|A|} \qquad \text{over-segmentation}$$

where R and A are the reference contour and the one to be evaluated, respectively.

The evaluation was done using the leave-one-out method in order to learn without bias the algorithm parameters. The best results were obtained for the segmentation of circumscribed lesion with a learning step using the same type of lesions ($P = 0.87$). The performance obtained for spiculated lesions with a learning step using the same population was a bit lower ($P = 0.71$). Finally testing the approach on circumscribed lesions while learning on spiculated data results in a small performance decrease compared to the ideal case.

5.2 Fuzzy Segmentation

Figure 4 illustrates an example of contour extraction using the penalization of a band on each side of the contours to force them to be nested. Let us notice that three of the contours are really close to each other. This could be interesting because this models the imprecision we have on the edges. Let us also remark that it may be interesting to get a common portion between several contours since imprecision may take place only in a part of the lesion. Unfortunately, the penalization method prevents such a behavior. Nonetheless, with a high enough radius quantization step in the polar domain, the obtained contours can be close enough to finally overlap when expressed in the Cartesian domain.
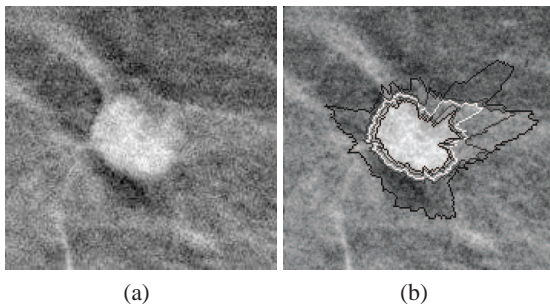

(a)   (b)

Figure 4: Fuzzy segmentation (b) of a cicumscribed lesion (a) using the penalization method.

We now present the results obtained using the distance-based contour selection method. Figure 5 illustrates a clinical example with contours extracted using the distance method. As we can see, the advantage of this method is that it captures the uncertainty: there is a lot of variations in the proposed contours. Let us remark that most of these contours make sense according to the lesion to segment. Nonetheless some are less likely to represent the lesion. However this is not an issue if we can associate those contours with a low membership degree. In this example, the gray levels of the contours are proportional to their membership degree: the lower the membership degree (contour gray levels), the less possible the contour.

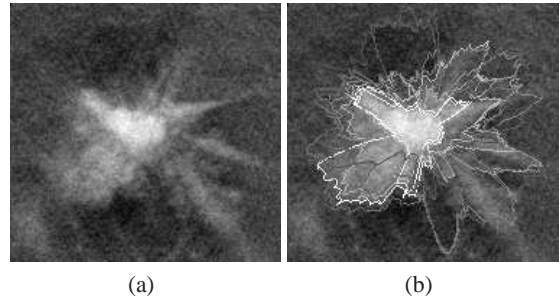(a)   (b)

Figure 5: Fuzzy segmentation (b) of a spiculated lesion (a) using the distance-based contour selection method.

Figure 6 shows how the fuzzy contours extraction method relying on the two contours selection method performs on a synthetic image. The distance based method allows retrieving four contours representing the uncertainty in the image. Those contours appear in black boxes, corresponding to the two disks, their union and intersection. For each of those contours, a set of nested contours is then extracted using the band penalization method allowing to build a fuzzy contour. Here we have 4 fuzzy contours in Figure 6(b), 6(c), 6(d) and 6(e). Note that for the sake of clarity, slight variations of the presented contours (part of the imprecision) that are also extracted are not shown. Let us also remark that fuzzy contours 6(b) and 6(e) are almost the same. Such a redundancy may appear because we want to extract as much main contours as possible in the first stage. Nonetheless such a behavior should not be a problem in real applications like automatic detection/characterization of cancer because no information is discarded.

5.3 Comments on the Membership Degree Definition

The comparison between the two techniques of membership degree computation is illustrated in Figures 7 and 8. These figures present the ranking of the contours in a fuzzy contour according to their membership degrees. It is interesting to notice the position of
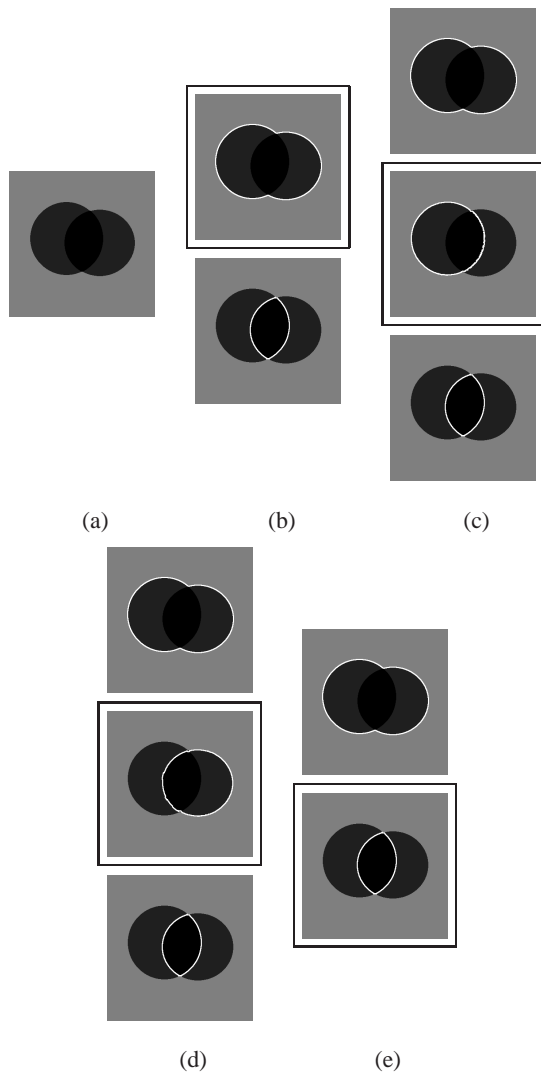
Figure 6: Fuzzy contours extraction. (a) Synthetic image to segment. (b) First fuzzy contour. (c) Second fuzzy contour. (d) Third fuzzy contour. (e) Fourth fuzzy contour. For each sub-figure, the black box designates the main contour retrieved by the distance based method, while the remaining contours constitute the fuzzy contour.

the smallest contour corresponding to the kernel of the lesion. When we use Equation 1 to compute the membership degrees, this contour is the worst one while when Equation 2 is used, this contour is the best. Actually, the membership degrees of the three best contours according to Equation 2 are almost the same, which is coherent with the visual quality of these contours. Let us also remark that this method could still be improved because the second and third contours may appear more relevant, but again since the difference between the membership degrees is low, this would be a minor improvement.
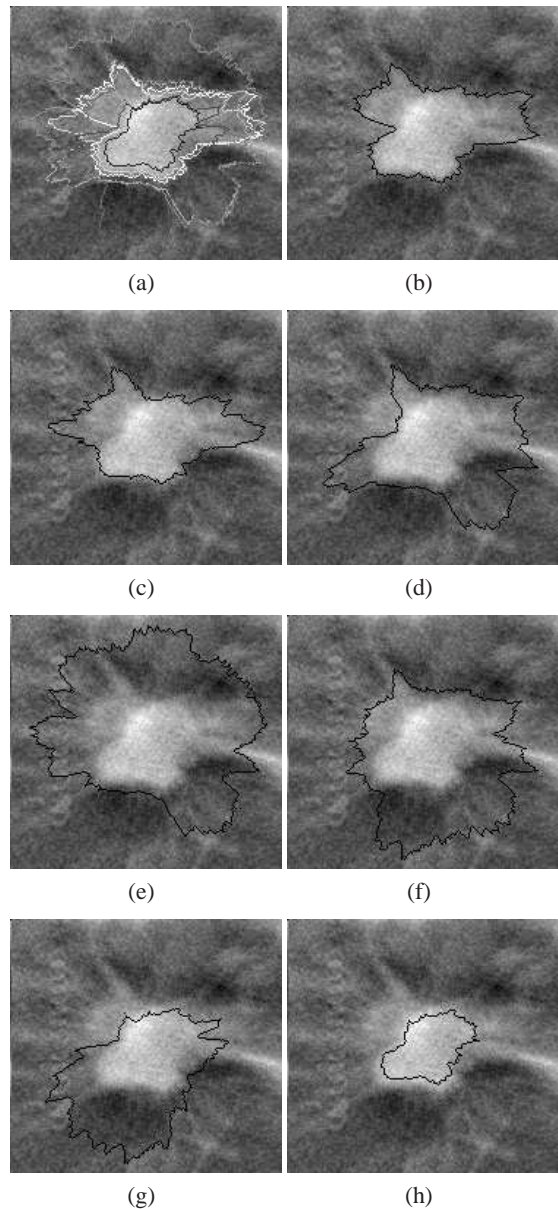

Figure 7: Membership degree computation using Equation 1. (a) Fuzzy contour to be considered. The remaining images are ranked in decreasing order according to the membership degree of the contour they contain. Due to this ranking, the first contour also corresponds to the result of a crisp segmentation.

6 CONCLUSIONS

We have proposed a new technique to extract fuzzy contours from images containing breast masses. This method relies on previous works that model a contour as a path in the image to segment, which has been converted in the polar domain (Timp and Karssemei-
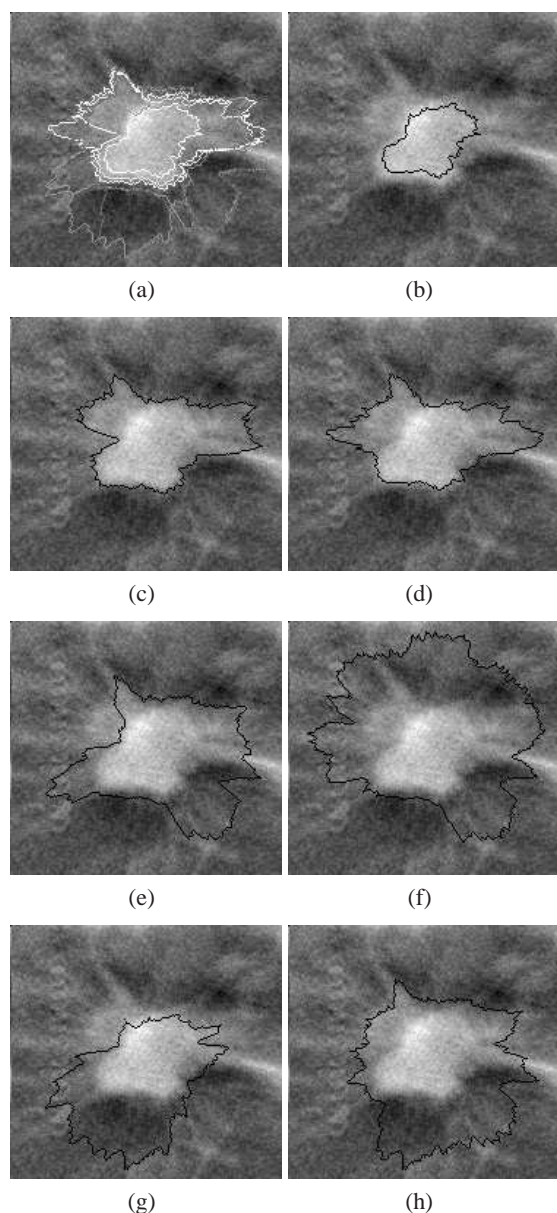
Figure 8: Membership degree computation using Equation 2. (a) Fuzzy contour to be considered. The remaining images are ranked in decreasing order according to the membership degree of the contour they contain.

jer, 2004; Rojas and Nandi, 2007). We extended this method in order to capture imprecision as well as uncertainty of lesions in mammography images or DBT slices. This is performed through two distinct contour extraction schemes, which we finally combine in order to get several fuzzy contours for a given structure to be segmented.

We also proposed a new technique to compute the membership degree of each contour contained in a fuzzy contour. This approach relies on the idea that

location where we can extract a lot of contours are likely to represent good contours. Using a contour density map, we thus derived a way to compute such degrees.

Because building a ground truth database for fuzzy contours is a pretty difficult task, as of today we only evaluated our approach visually on the lesions of our whole database. One next step may be to work with clinical experts to review images in order to get several contours for these lesions. Of course to be relevant all the segmentations for each lesion should be performed independently.

REFERENCES

Bothorel, S., Bouchon Meunier, B., and Muller, S. (1997). A fuzzy logic based approach for semiological analysis of microcalcifications in mammographic images. *International Journal of Intelligent Systems*, 12(11-12):819–848.

Dobbins III, J. T. and Godfrey, D. J. (2003). Digital X-ray tomosynthesis: current state of the art and clinical potential. *Physics in Medicine and Biology*, 48(19):R65–R106.

Gennaro, G., Baldan, E., Bezzon, E., Grassa, M. L., Pescarini, L., and di Maggio, C. (2008). Clinical performance of digital breast tomosynthesis versus full-field digital mammography: Preliminary results. In *International Workshop on Digital Mammography (IWDM)*, pages 477–482, Berlin, Heidelberg. Springer-Verlag.

Osher, S. J. and Fedkiw, R. (2002). *Level Set Methods and Dynamic Implicit Surfaces*. Springer.

Palma, G., Peters, G., Muller, S., and Bloch, I. (2008). Masses classification using fuzzy active contours and fuzzy decision trees. In *SPIE Symposium on Medical Imaging*, number 6915, pages 691509.1–691509.11, San Diego, CA, USA.

Peters, G. (2007). *Computer-aided Detection for Digital Breast Tomosynthesis*. PhD thesis, Ecole Nationale Supérieure des Télécommunications.

Rojas, A. and Nandi, A. K. (2007). Improved dynamic-programming-based algorithms for segmentation of masses in mammograms. *Medical Physics*, 34:4256–4269.

Timp, S. and Karssemeijer, N. (2004). A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *Medical Physics*, 31:958–971.

Wu, T., Stewart, A., Stanton, M., McCauley, T., Phillips, W., Kopans, D. B., Moore, R. H., Eberhard, J. W., Opsahl-Ong, B., Niklason, L., and Williams, M. B. (2003). Tomographic mammography using a limited number of low-dose cone-beam projection images. *Medical Physics*, 30(3):365–380.