

Fuzzy Techniques in Mammographic Image Processing

Andreas Rick^{1,2}, Sylvie Bothorel², Bernadette Bouchon-Meunier¹, Serge Muller², and Maria Rifqi¹

¹ Université Pierre et Marie Curie
LIP6, Case 169, 4 place Jussieu
75252 Paris CEDEX 05, France

² General Electric Medical Systems
283 rue de la Minière
78533 Buc France

Summary. In this paper we discuss fuzzy techniques for the detection and analysis of potential breast cancer lesions on mammograms. We show how fuzzy measurements can be performed on the images and how this information can be used in the different stages of the processing.

Introduction

Medical decision problems are typical examples for multi-criteria decision problems in which no single criteria is sufficient to arrive at a decision with a comfortable level of certainty. In many cases only the combination of many information sources, like patient interview, patient history and examinations together allow the doctor to pronounce his diagnosis. As different pieces of information acquired and fused for the final decision have different degrees of certainty and precision, the fusion process should be able to reflect these uncertainties. Until recently, all information in mammography was acquired, fused and evaluated by the doctor. Today however, with the increasing performance of image processing techniques, we are at a point where doctors using a computer aided detection (CAD) systems perform better in some aspects, than the doctors alone. The strong points of CAD systems are their precision and repeatability. Logically, efforts are being made to combine the expert knowledge of the doctor with the precision and repeatability of the CAD system. Fuzzy logic is a powerful tool for the formulation of expert knowledge and the combination of imprecise information from different sources. To achieve meaningful results the imprecision in all information used to come to a conclusion should be taken into account.

In this paper we discuss fuzzy techniques for the detection of cancer on mammograms. We show how fuzzy measurements can be performed on the images and how this information can be used in the different stages of the processing.

We show that the use of a fuzzy representation allows the decisions to be taken whenever sufficient information is available in the processing stage and

to propagate the information including its uncertainty level to the next processing stage if the decision cannot be performed with sufficient confidence.

In the first part we discuss techniques for mammographic image processing and in the second part we show some applications of these methods on clinical problems.

Part1: Techniques

The signs of cancer visible on mammograms can be separated in two groups: microcalcifications and fibrous lesions. Figure 1 shows zoomed views of two examples.

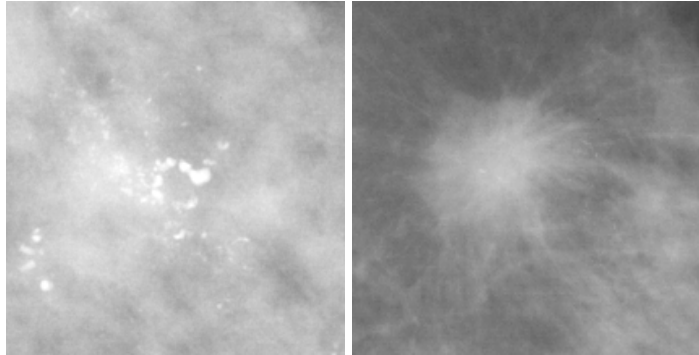


Fig. 1. Examples of microcalcifications (left) and fibrous lesions (right)

The microcalcifications are small depositions of radiologically very opaque materials like $\text{Ca}_3(\text{PO}_4)_2$, CaCO_3 , $\text{Mg}_3(\text{PO}_4)_2$ [17] which can be seen on mammography exams as small bright spots. In section 1 we will discuss how fuzzy contours and measurement of fuzzy attributes can be used for characterization of microcalcifications. In section 2 we describe an adaptive method to determine membership functions for classification with attributes that present significant variability from one image to the other. This technique can be applied to microcalcifications as well as fibrous lesions. In section 3 we will discuss techniques for comparisons of fuzzy measurements which are needed in many parts of mammographic image processing.

1 Fuzzy Contours and Fuzzy Measurements

The first stage of a CAD system is the detection of the microcalcifications on the mammographic images. A number of methods have been proposed in the literature for this purpose. We will not go into details here, interested readers

may refer to [16]. The detection stage gives a number of markers for potential microcalcifications. For the distinction between real microcalcifications and other structures the marked signs have to be analyzed. The attributes that can be calculated on the microcalcifications can be separated into those which are calculated without segmentation and those using a segmentation of the calcification. Information captured by attributes without segmentation tend to be somewhat deluded by the fact that some background is being taken into account for their calculation. On the other hand attributes that use a segmentation have to rely on the segmentation. Segmentation of microcalcification is a difficult task because of their small size, their contour not always well defined and the possibility of overlapping due to the projection. Due to the image acquisition by projection for X-rays, the 3D-structures may superimpose and more than one microcalcification may be projected at the same position on the image. Therefore for some microcalcifications a unique contour can easily be found, but for other microcalcifications more than one contour can potentially be chosen.

We propose to use a fuzzy approach to catch the uncertainty of the segmentation and transmit it to the other processing stages.

1.1 Segmentation of Microcalcification and Fuzzy Contours

A set of contour candidates are calculated for each microcalcification by region growing [3]. This ordered set of contours is considered the universe of all possible contours describing the given microcalcification. An example of a zoomed part of a mammogram including the set of possible contours is given in figure 2.

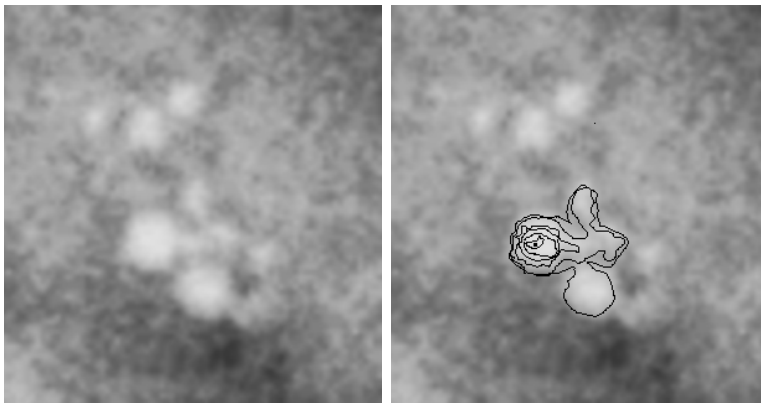


Fig. 2. The set of possible contours for a microcalcification

The prior knowledge about the contours of microcalcifications, which can be resumed by “*microcalcifications are small and have a high contrast*”, is

translated into a fuzzy set description using two attributes namely the *area* and the *gradient* shown in figure 3.

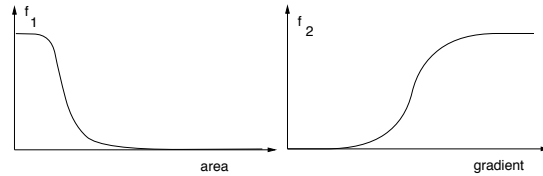


Fig. 3. Translation into fuzzy rules of the a priori information: *microcalcifications are small and have a high contrast*

The two functions express the degree of membership to the class *contour of a microcalcification* as a function of the size (*size is small*) and the image gradient under the contour (*the gradient is high*). Using these membership functions the membership values for each contour are calculated by the conjunction of the membership values for *small* and *high gradient*. Figure 4 show the application of this method for one particular microcalcification: the two graphs at the left show the area and the gradient for each contour obtained by the region growing process. The right graph shows the resulting membership function to the class *contour of the microcalcification* for each contour in the universe.

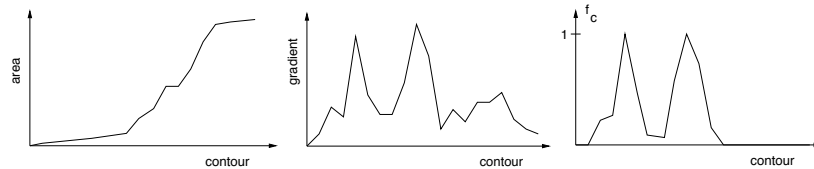


Fig. 4. Measured area and gradient values for all contours of a microcalcification

In this example the bi-lobed membership function can be interpreted in the following way: there are two possibilities for the segmentation, one that corresponds to a small microcalcification and another that corresponds to a bigger one. Each of the two contours with very high membership values are surrounded by some contours with slightly lower degrees of membership which translate the imprecision of the contour position.

Once the membership function for *contour of a microcalcification* on the universe of all contours has been calculated the attributes of this fuzzy contour can be evaluated.

1.2 Measurements of Fuzzy Attributes

Several attributes are calculated on each of the contours with a non-zero membership value. To transform the attribute measurements performed on all the contours of the universe to a fuzzy measure which characterizes the microcalcification we can use the extension principle.

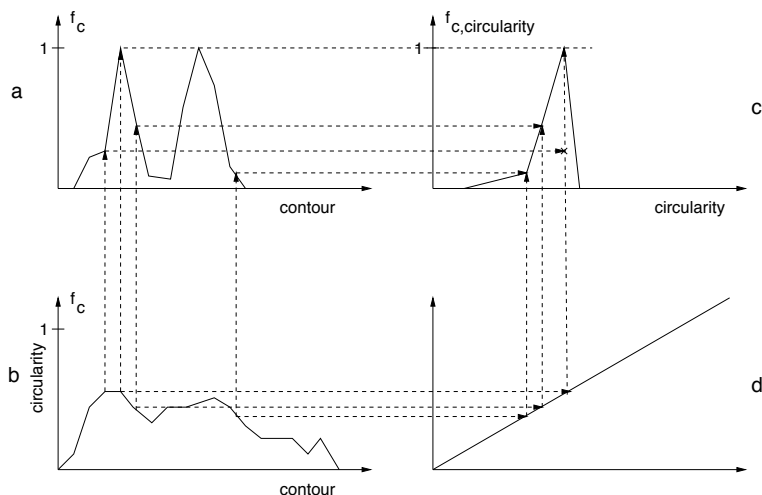


Fig. 5. Fuzzy measurements on contours using the extension principle

Figure 5 show the application of the extension principle to calculate the degree of circularity of a microcalcification. The figure 5a shows the degree of membership to the class *contour of a microcalcification* for each contour obtained by the fuzzy segmentation and the figure 5b shows a crisp circularity measure for each contour. By combining the two graphs we obtain the membership function for the *circularity* of a fuzzy contour which describes a microcalcification. While using the extension principle as shown in figure 5, it is possible to obtain more than one degree of membership for a given attribute value whenever the original membership functions are not monotonous. As we interpret the membership function as a degree of possibility for the microcalcification to have a certain attribute value, we use the maximum operator to obtain a unique degree of membership for each value of the attribute.

In the example shown in figure 5c, we have computed the membership function for the circularity measurement using the example shown in figure 4 which can be interpreted in the following way: the ambiguity about the contours has disappeared in the circularity measure as all possible contours for the microcalcification exhibit a high circularity. Therefore the circularity measurement does not show any ambiguity, but some imprecision remains.

1.3 Conclusion

The fuzzy contours allow to capture the uncertainty and imprecision of the contours. The extension principle makes it possible to forward these uncertainties to any attribute that can be measured on a single contour to obtain fuzzy measurements for any attribute. The fuzzy measurements can be used in many applications like semiological classification of lesion or matching of microcalcifications (see section about applications).

2 Adaptive Membership Functions

A priori, the fuzzy segmentation and measurement methods presented in the previous section should be applicable to fibrous lesions as well as to microcalcifications.

Unfortunately the attributes calculated on mammographic images for the detection or classification of fibrous lesions show a significant variability from one image to the other which are due to the large variability of the breast tissue.

This big variability of the normal structures in mammograms between patients makes it extremely difficult to distinguish between normal and abnormal findings using rules with predefined membership functions applied to the fuzzy measurements.

We propose the construction of adaptive prototypes for the membership functions. The prototypes are adapted to the current image to obtain a more discriminative representation of the class memberships and thereby extract the maximum of information from each attribute. With this approach we can use a single set of rules for the identification of lesions.

2.1 Adaptive Membership Functions for Classification

In the following we consider the fuzzy decision problem that consists in classifying each pixel of the image into one of two classes *normal* or *suspect*. A simple way of creating probabilistic membership functions for the classes is to use attribute histograms as illustrated in figure 6:

The histogram of the attributes over all pixels of each class in the learning database is calculated. The histogram is normalized by the joint histogram of all classes together. If the attribute under consideration exhibits variation coming from other factors than the class membership of the pixels (like exposition parameter setup for the image acquisition) the histogram may be shifted in the universe. The variations will be modeled in the membership functions by an increase in uncertainty. If the variability between images is much bigger than the variations between classes, the membership functions have a very small discrimination power.

A way to reduce this uncertainty is to extract some information about the current case from the joint histogram of all classes. A simple method to

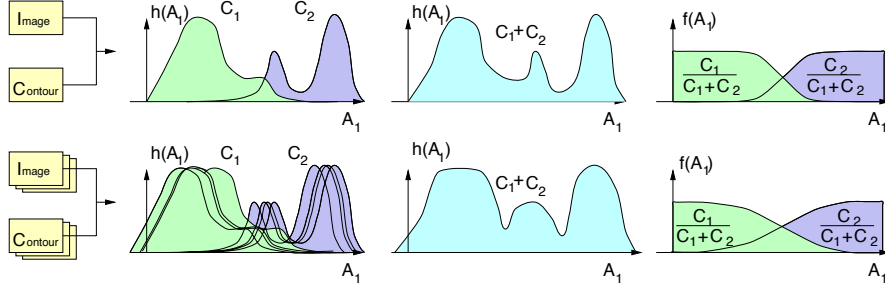


Fig. 6. Classification using varying attributes

do this, would be to adapt the average of the different membership functions to the average of the current image. Nevertheless this will not allow to take advantage of regularities which cannot be described by a simple shift of the histogram.

In order to find more complicated relationships we have to find some parameters to describe the form of the histogram and then find the relationships between these parameters.

2.2 Parametrization of the Histogram

The parametrization of the histogram must satisfy a certain number of conditions in order to be useful for expressing the relationships between the joint histogram of all classes and the histogram of one class.

- **Robustness:** The parametrization must change little if the histogram changes little
- **Completeness:** All histograms which may be encountered for the problem under consideration must be supported by the selected parameters.
- **Compactness:** The number of parameters must be as low as possible.

The parametrization we propose to use is a weighted sum of parametric functions f_k .

$$f(x|p_k, \Theta_k) = \sum_{k=1}^K p_k \cdot f_k(x|\Theta_k) \quad (1)$$

for example a weighted sum of Gaussian:

$$f(x|\Phi) = \sum_{k=1}^K p_k \cdot g_k(x, \mu_k, \sigma_k) \quad (2)$$

with :

$$g_k(x, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2} \quad (3)$$

and :

$\Phi = (\mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma})$
 $\mathbf{p} = (p_1, \dots, p_k)$, with $p_k =$ weight Gaussian k
 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$, with $\mu_k =$ average of Gaussian k
 $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$, with $\sigma_k =$ variance of Gaussian k

2.3 Parameter Estimation with the EM-Algorithm

There exists a variety of algorithms for the parameter estimation of a Gaussian mixture model. Most of them are iteration methods based on the *EM-algorithm* proposed by Demster [8]. The idea of the EM-algorithm is to start with an initial estimate $(\mathbf{p}_0, \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)$ for the parameters of the component functions and estimate a membership value λ_{ki} for each example i to each component k . During the iterations two steps are performed alternately:

- E-step: Update the λ_{ki} based on the last set of parameters $(p^{(m-1)}, \mu^{(m-1)}, \sigma^{(m-1)})$.

$$\lambda_{ki}^{(m)} = \frac{p_k^{(m-1)} \cdot g(x_i, \mu_k^{(m-1)}, \sigma_k^{(m-1)})}{\sum_{j=1}^K p_j^{(m-1)} \cdot g(x_i, \mu_j^{(m-1)}, \sigma_j^{(m-1)})} \quad (4)$$

- M-step: Optimize the parameters of the Gaussian with the examples weighted by the λ_{ki} .

$$p_k^{(m)} = \sum_{i=1}^N \lambda_{ki}^{(m)} \quad \mu_k^{(m)} = \frac{\sum_{i=1}^N \lambda_{ki}^{(m)} x_i}{p_k^{(m)}} \quad (5)$$

$$\sigma_k^{(m)} = \sqrt{\frac{\sum_{i=1}^N \lambda_{ki}^{(m)} (x_i - \mu_k^{(m)})^2}{p_k^{(m)}}} \quad (6)$$

The problem of the standard EM-algorithms is the necessity of an initialization close to the optimal state. Based on the initialization, the obtained result corresponds to a local minimum close to the starting position which in general is not the global minimum. As an illustration we present a histogram approximated by 8 or 12 Gaussian which were initially equally spaced in the universe (figure 7).

The main lobe in the histogram is approximated with one or two components, depending on the initialization. This method of parameter estimation is not well adapted for variance modeling, as the parameters depend not only on the histogram itself but also on the initialization.

2.4 Parameter Estimation with the Hierarchical EM-Algorithm

A modified version of the EM-algorithm is proposed by Huang et al. [31]. This approach adds the components one by one to the model and estimates their parameters under the assumption of uniform noise. Uniform noise is

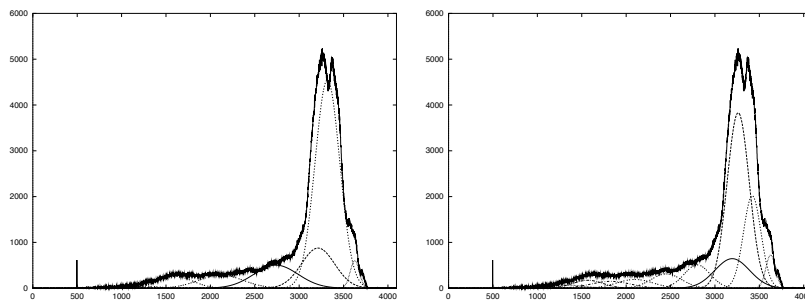


Fig. 7. Approximation of a distribution by 8 or 12 components

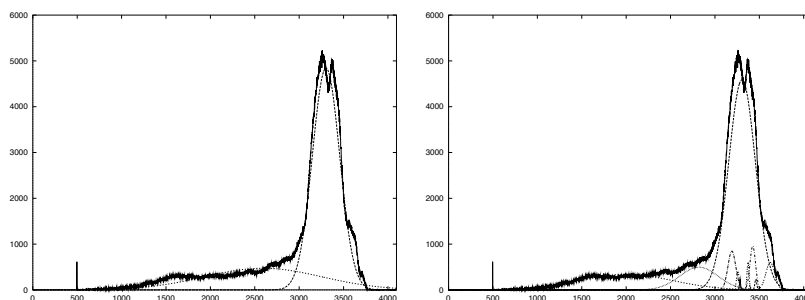


Fig. 8. Hierarchical approximation by 2 or 12 components

used as a simplification for all components which are to be identified later. This hierarchical decomposition is much more stable. In figure 8 the same histogram is used to identify 2 or 12 components.

The principal components are always identified the same way and are independent of the number of total components that are used. In the following we will call this algorithm *H-EM*.

2.5 Modeling of Variability

Once the histogram has been parametrized the links between the joint histogram of all classes and the membership functions for each class must be established. Figure 9 summarizes the learning phase in which the parameter models for two classes are identified.

The regression fits some function to the parameters of the joint histogram to predict the membership functions for each class. In some practical applications the estimated parameters by means of the hierarchical EM algorithm are not sufficiently stable for a regression (e.g. with a linear model). One of the reasons of this can easily be understood for a histogram with two equally high and well separated components. The H-EM algorithm will detect the two components, but the order of the components is undetermined.

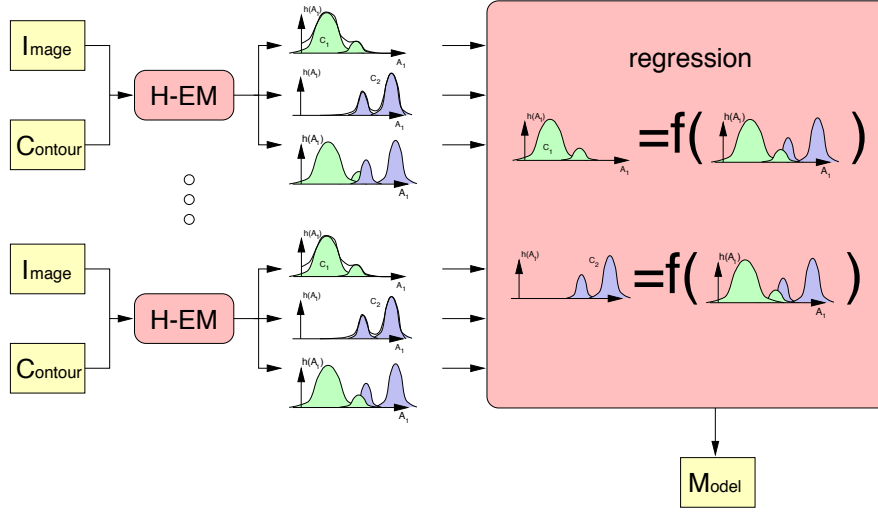


Fig. 9. Regression of membership function parameters

2.6 Generation of a Prototype

In order to further stabilize the parameters we propose to calculate a prototype for the components and use the prototype during the histogram approximation phase. A simple static prototype can be described with an average value for all parameters of the mixture and a variance around this value. It can be shown that this approach is equivalent to specify a prior probability distribution of the parameters of Gaussian components which are themselves Gaussian.

To use the prototype during the estimation of the mixture components, the EM algorithm must be modified. We will call the modified algorithm P-EM (prototype-EM).

- E-step: Update the λ_{ki} based on the last set of parameters $(p^{(m-1)}, \mu^{(m-1)}, \sigma^{(m-1)})$.
- M-step: The optimization of the parameters of the Gaussian with the examples weighted by the λ_{ki} .
- P-step: Update the parameters of the Gaussians to take into account the prototype.

$$p_i^{(m+1)} = \alpha_{i_p} p_{i_c} + (1 - \alpha_{i_p}) p_i^{(m)} \quad (7)$$

$$\mu_i^{(m+1)} = \alpha_{i_\mu} \mu_{i_c} + (1 - \alpha_{i_\mu}) \mu_i^{(m)} \quad (8)$$

$$\sigma_i^{(m+1)} = \alpha_{i_\sigma} \sigma_{i_c} + (1 - \alpha_{i_\sigma}) \sigma_i^{(m)} \quad (9)$$

$$(10)$$

For the static prototype, the P-step consists in taking a weighted average between the prototype center values $(p_{i_c}, \mu_{i_c}, \sigma_{i_c})$ and the current parameters $(p_i^{(m)}, \mu_i^{(m)}, \sigma_i^{(m)})$ calculated in the M-step. The weights $(\alpha_{i_p}, \alpha_{i_\mu}, \alpha_{i_\sigma})$ determine the importance of the prototype versus the importance of the histogram.

2.7 Prototype with Parameter Interactions

The P-EM algorithm is able to avoid instabilities in the parametrization of the histogram like exchanged components. On the other hand, there is no direct link between the components during the approximation. In case of a histogram which contains a pattern that has not the form of a single component (like a triangular area when using Gaussian components) the pattern can be approximated by a number of components. To find these patterns in a histogram, a link between the parameters of the components used to describe the pattern is necessary, otherwise each component will be updated independently of the others during iteration.

An efficient way to incorporate parameter interactions into the model is to change the P-step in the P-EM algorithm with an update which includes interactions. One example of interaction is a linear relationship between parameters. We can define a matrix \mathbf{A} that describes a linear relationship between the components of the parameter vector Φ .

$$\mathbf{A} \cdot \Phi = 0 \quad (11)$$

In the general case the parameters of the model do not show a perfectly linear relation so that the equation 11 becomes:

$$\mathbf{A} \cdot \Phi = \epsilon \quad (12)$$

The error of the linear model ϵ can be used to define the internal energy of the mixture components:

$$E_{int} = [\mathbf{A}\Phi]^T \mathbf{R}_i^{-1} [\mathbf{A}\Phi] \quad (13)$$

where \mathbf{R}_i is the diagonal matrix of the standard deviation of the linear model. With this notation the algorithm P-EM minimizes an energy function which is the sum of the external energy E_{ext} which describes the correspondence between the mixture model and the histogram, and the internal energy E_{int} which describes the distance of the adapted mixture model to the prototype model. E_{ext} is minimized by the E- and M- steps of the algorithm P-EM and E_{int} is minimized in the P-step.

2.8 Prototype Model Learning from a Database

The parameters for the adaptive membership functions are the center of the prototype Φ_0 , the distance matrix \mathbf{A} and its standard deviation matrices \mathbf{R}_i

(standard deviation of the linear prototype) and \mathbf{R}_e (standard deviation of the static prototype). To learn these parameters from a learning data base, we first apply the hierarchical-EM algorithm on all histograms. This will give a first estimation of the components that are encountered, but they are not necessarily ordered. The next step is to find histogram components that are present in most of the images, in other words to apply a clustering algorithm to the components estimated by the H-EM. The clustering of the components establishes the order which is needed for a successful estimation of the interaction between components

After the clustering, a matrix $\bar{\Phi}$ containing all the ordered parameters Φ_i for all images can be constructed and the matrix \mathbf{A} can be calculated by a least square approach minimizing $\mathbf{A} \cdot \bar{\Phi}$. Using \mathbf{A} , \mathbf{R}_e and \mathbf{R}_i the P-step can now be formulated as:

$$\tilde{\Phi} = \left(\mathbf{R}_e^{-1} + \mathbf{A}^T \mathbf{R}_i^{-1} \mathbf{A} \right) \Phi \quad (14)$$

2.9 Conclusion

For applications such as detection of fibrous lesions on mammograms, fuzzy rule based reasoning can be applied, but the amount of uncertainty introduced by different sources of variability seriously reduces the classification performance. The use of adaptive membership functions can help to reduce the uncertainty by extracting a model for at least some part of the variability and thereby increase the classification performance.

3 Comparison Measures and Prototypes for Image Processing

Comparison of objects is a widespread operation necessary in many frameworks. It is based on the idea that if two objects are sufficiently similar then a transfer of knowledge is possible from one to the other. This comparison is frequently achieved through a measure intended to determine to which extent the descriptions have common points or differ from each other.

Many measures of comparison have been proposed and studied. We have proposed a general form for measures of comparison, enabling us to classify the main existing ones with respect to their properties. Thus, the framework we have obtained provides a means to choose a measure of comparison relevant and appropriate for a given problem. We use the formalization and the framework introduced in [4] to deal with measures of comparison. We show the interest of this framework in two kinds of problems: classification and construction of a fuzzy prototype.

3.1 A Framework to Compare Fuzzy Subsets

We propose the construction of a general framework enabling to include the majority of all the measures, indices and relations of comparison. For this purpose, different classes of measures are defined according to their properties: comparison, similitude, dissimilarity, satisfiability, inclusion, resemblance... This framework, described in details in [4], is a tool for a classification of the existing measures as well as for a creation of new quantities.

For any set Ω of elements, let $F(\Omega)$ denote the set of fuzzy subsets of Ω , f_A the membership function of any description A in $F(\Omega)$. We use the classical definition of intersection: $f_{A \cap B} = \min(f_A, f_B)$ to describe the elements belonging to A and B . The definition used for the difference between two fuzzy subsets is the following:

An operation on $F(\Omega)$ is called a **difference** and denoted by $-$, if it satisfies for every A and B in $F(\Omega)$:

- if $A \subseteq B$, then $A - B = \emptyset$.
- $B - A$ is monotonous with regard to B : $B \subseteq B'$ entails $B - A \subseteq B' - A$

We suppose that we have a way of evaluating the weight of the elements of the universe characterized by a fuzzy set through a **fuzzy set measure** M . This fuzzy set measure M is a mapping $F(\Omega) \rightarrow \mathbb{R}^+$ such that, for every A and B in $F(\Omega)$:

- $M(\emptyset) = 0$
- if $B \subseteq A$, then $M(B) \leq M(A)$.

A **measure of comparison** between A and B in the most general sense, takes into account common features, $A \cap B$, and distinctive features, $B - A$ and $A - B$. Therefore a M -measure of comparison on Ω is a mapping $S : F(\Omega) \times F(\Omega) \rightarrow [0, 1]$ such that $S(A, B) = F_S(M(A \cap B), M(B - A), M(A - B))$, for a given mapping $F_S : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, 1]$ and a fuzzy set measure M on $F(\Omega)$.

One example of a M -measure of comparison is Zadeh's inclusion: $S(A, B) = \frac{1}{|\Omega|} \sum_x \min(1, 1 - f_A(x) + f_B(x)) = 1 - \frac{1}{|\Omega|} M(A -_1 B)$ with $f_{A -_1 B} = \max(0, f_A - f_B)$ and M the sigma-count [7].

A M -measure of comparison can either evaluate the likeliness of two descriptions (it is called a M -measure of similitude), or their differences (it is then called a M -measure of dissimilarity). We consider the following properties for F_S :

- symmetry: $F_S(u, v, w) = F_S(u, w, v)$
- minimality: $F_S(u, 0, 0) = 0$
- reflexivity: $F_S(u, 0, 0) = 1$
- containment: $F_S(u, 0, w) = 1$ whatever u and w may be.

- exclusiveness: $F_S(0, v, w) = 0$ whatever $(v, w) \neq (0, 0)$ may be.

Obviously, symmetry of F_S implies that S satisfies $S(A, B) = S(B, A)$. Minimality of F_S entails that $S(A, A) = 0$ while reflexivity corresponds to $S(A, A) = 1$. The containment property yields $S(A, B) = 1$ if $B \subseteq A$. The exclusiveness property implies that $S(A, B) = 0$ as soon as $A \cap B = \emptyset$.

A **M -measure of similitude** S on Ω is a M -measure of comparison S such that $F_S(u, v, w)$ is non decreasing in u , non increasing in v and w .

An example for a f -measure of similitude is the relation given by Tversky [28] and generalized to fuzzy sets [26], [4]:

$$S(A, B) = f(A \cap B) / (f(A \cap B) + \alpha f(A - B) + \beta f(B - A))$$

with $\alpha, \beta \geq 0$ and f a fuzzy set measure.

The **measure of dissimilarity** is not defined as the dual of a measure of similitude, but it has specific properties. A M -measure of dissimilarity S on Ω is a M -measure of comparison satisfying the minimality property and such that $F_S(u, v, w)$ is independent of u and non decreasing in v and w .

A definite symmetrical M -measure of dissimilarity (i.e. $S(A, B) = S(B, A)$, $\forall A$ and B) satisfying the triangular inequality (i.e. $S(A, B) \leq S(A, C) + S(C, B)$, $\forall A, B$ and C) is a distance.

In order to classify the existing measures more subtly, we distinguish three types of M -measures of similitude: satisfiability, inclusion and resemblance.

A **measure of resemblance** is used for a comparison between the descriptions of two objects, of the same level of generality, to decide if they have many common characteristics. A M -measure of resemblance on Ω is a M -measure of similitude S which satisfies the reflexivity and the symmetry properties.

In the case where T is the minimum, we obtain extensions of measures of similarity. Examples of M -measures of resemblance are the following:

- $S(A, B) = \exp(-\beta |d_r(A, B)|)$ [19] where $\beta > 0$ and $d_r(A, B) = (\sum |f_A - f_B|^r)^{1/r}$, for $r \geq 1$. This quantity is a product-transitive indistinguishability relation.
- $S(A, B) = M(A \cap B) / M(A \cup B)$ [10] for M such that: $M(A \cup B) = M(A \cap B) + M(A - B) + M(B - A)$.
- $S(A, B) = 1 - \frac{1}{|\Omega|} \sum_x |f_A(x) - f_B(x)| = 1 - \frac{1}{|\Omega|} (M(A - B) + M(B - A))$ [10] with the sigma-count $M(A) = \sum_x f_A(x)$ as a fuzzy set measure.

A **measure of satisfiability** corresponds to a situation in which we consider a reference object or class and we decide if a new object is compatible with it or satisfies the reference. A M -measure of satisfiability S on Ω is a M -measure of similitude S satisfying the containment and the exclusiveness properties and such that $F_S(u, v, w)$ is independent of w .

Analogy relations [5]: $S(A, B) = \inf_x \min(1 - f_B(x) + f_A(x), 1)$, and fuzzy similitude [6]: $S(A, B) = 1 - \sup_{f_A(x)=0} f_B(x)$ are particular M -measures of

satisfiability. Measures of satisfiability have been proven [4] to be compatible with the contrast model introduced by Tversky, satisfying major properties such as matching, monotonicity, independence, solvability [28].

A **measure of inclusion** also concerns a situation with a reference object and measures if the points common to A and B are important with regard to A . A M -measure of inclusion S on Ω is a M -measure of similitude satisfying the reflexivity and the exclusiveness properties such that $F_S(u, v, w)$ is independent of v .

As an example of M -measure of inclusion, we can find the degree introduced by Sanchez [25]: $S(A, B) = |A \cap B|/|A| = M(A \cap B)/(M(A \cap B) + M(A - B))$ with the sigma-count as M .

A classification of all the measures of comparison discussed above is shown in figure 10. We point out that there exist measures of comparison which

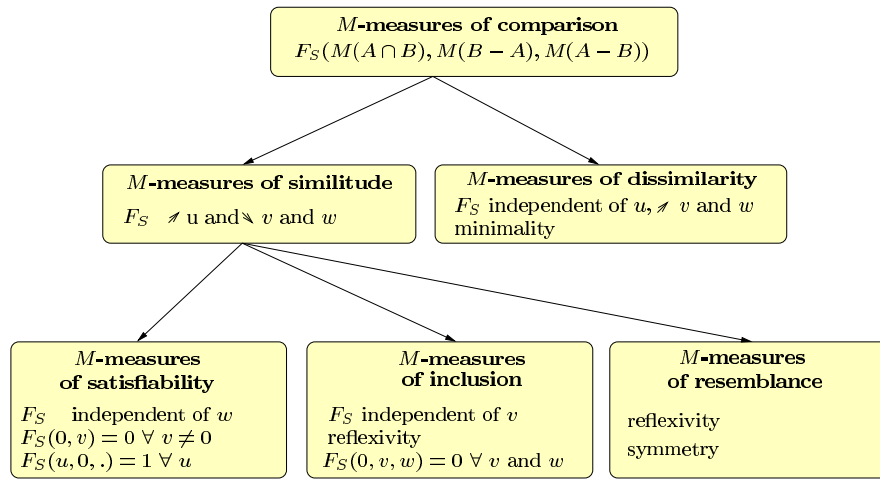


Fig. 10. Classification of M -measures of comparison

are neither similitudes nor dissimilarities. The same holds for measures of similitude: they can be different from satisfiability, inclusion or resemblance. In the rest of this section, we show the use of this framework for two kinds of problems in mammographic image processing: classification and construction of fuzzy prototypes.

3.2 Measures of Comparison for Classification Process

A similarity based-classification method has to solve the problem of the choice of a measure of similarity or, more generally, a family of measures of comparison. A new object we do not know the class is often classified thanks to

a comparison with a rule. Indeed, a prototype can be considered as a rule describing a class. For example, the prototype of the class “round” might be: **around 15 pixels** for surface, **between 5 and 6** for compactity. In other words, **if surface = around 15 pixels, and compactity = between 5 and 6** then **class = round**. A decision tree is also equivalent to a set of IF...THEN rules. Each path in the tree is associated with a rule, where premises are composed by the tests of the encountered nodes, and the conclusion of the rule is composed by the class associated with the leaf of the path.

The classification process is based on the question: *does the new object satisfy the rule?* This question entails the use of a measure which is *maximal when the object is included in the prototype* (exclusiveness property) and can be considered as a particular case of the prototype, and *minimal when no common features are shared by the two objects* (minimality property). In the proposed framework, these properties are shared by *measures of satisfiability*.

3.3 Construction of a Fuzzy Prototype

The prototype is intrinsically interesting because of its power of description. This power can be used for a classification process for example the semiological classification of microcalcifications in mammography.

According to E. Rosch [23], all objects do not represent in a same manner the category they belong to. They are spread along a scale of typicality. According to Rosch and Mervis [24] :

[..] categories tend to become defined in terms of prototypes or prototypical instances that contain the attributes most representative of items inside and least representative of items outside the category.(p.30)

Therefore the notion of prototype is linked to the notion of typicality. Zadeh [30] has emphasized this aspect: the typicality is a matter of degree and it implies that the concept of prototype is a fuzzy concept. A prototype, as L. A. Zadeh said, is not a unique object or a group of objects. It is more a fuzzy schema enabling us to generate a set of objects because of the synthesized information it contains.

In the approach we describe here, we retain Rosch’s idea that an object is all the more typical as it resembles the members of its class and it is distinct with regard to the members of others classes and Zadeh’s idea that a prototype is not a particular example but a fuzzy schema. The difference with the standard methods of determining prototype lies on the fact that values of attributes rather than objects themselves are compared pair by pair.

We need to determine the typicality of *each value appearing* in a learning database in order to construct a fuzzy prototype.

This situation of comparison is based on the assumption that values of attributes are considered to have the same level of generality, and no value can be taken as a reference. This situation needs a *symmetrical* measure

because all values have the same level of generality and a *reflexive* measure because there does not exist any pair of distinctive objects which perfectly look like each other. In the proposed framework, these properties are shared by *measures of resemblance*.

Let X be a set of objects. We suppose that there exists a partition given on X composed by crisp classes C_j . For an object O of the class C_i , the typicality of the value v of an attribute A is computed as follows:

- Step 1. Compute the resemblance $r(v, v_j)$ between v and the value v_j of the attribute A for any example of the same class C_i . The global resemblance $R(v)$ relative to the set of values of A present in examples, is obtained by aggregating the degrees $r(v, v_j)$.
- Step 2. Compute the dissimilarity $d(v, v_j)$ between v and the value v_j of the attribute A for any example of class C_k different from C_i . The total dissimilarity $D(v)$ relative to the set of values of A present in examples, is obtained by aggregating the degrees $d(v, v_j)$.
- Step 3. The aggregation of this two values (for example with a t-norm), $R(v)$ and $D(v)$, gives the typicality $T(v)$ of v , according to the attribute A , for the class C_i .

Degrees of typicality participate in the construction of a fuzzy prototype of a given class. For an attribute A , the degree of typicality of each value of A is computed for each class. Then, the fuzzy prototype of any given class is characterized by the most typical value(s) of each attribute. This means that a fuzzy prototype is a virtual object described by means of the same attributes as the objects in the learning database. The values taken by the fuzzy prototype are the most typical.

One of the intermediate steps of the semiological classification is the classification of the microcalcifications according to their form: *round* or *not round*. Figure 11 shows the prototypes obtained for the classes *round* and *not round* by the method described above, applied to a learning database of annotated mammograms. More specifically, these prototypes have been obtained by using :

- $M(A) = \int_{\Omega} f_A(x)dx$ for the fuzzy set measure,
- $S(A, B) = \frac{2}{\pi} \arctan\left(\frac{2M(A \cap B) \sup(A \cap B)}{M(A \cup B)}\right)$ for the measure of resemblance,
- $S(A, B) = \frac{1}{\Omega} (\sum_{x/f_A > f_B} f_{A-B}(x) + \sum_{x/f_B > f_A} f_{B-A}(x))$ for the measure of dissimilarity,
- the mediane for the aggregation of resemblance and for the aggregation dissimilarity in the steps 1 and 2 of the method.
- the probabilist t-conorm (i.e. $\perp(x, y) = x + y - xy$) for the aggregation operator in the third step of the method.

The prototypes can be provided to the radiologist as a help to formalize his reasoning and to better understand the structure of microcalcifications and its link to the semiological classes. For example the first row in figure 11

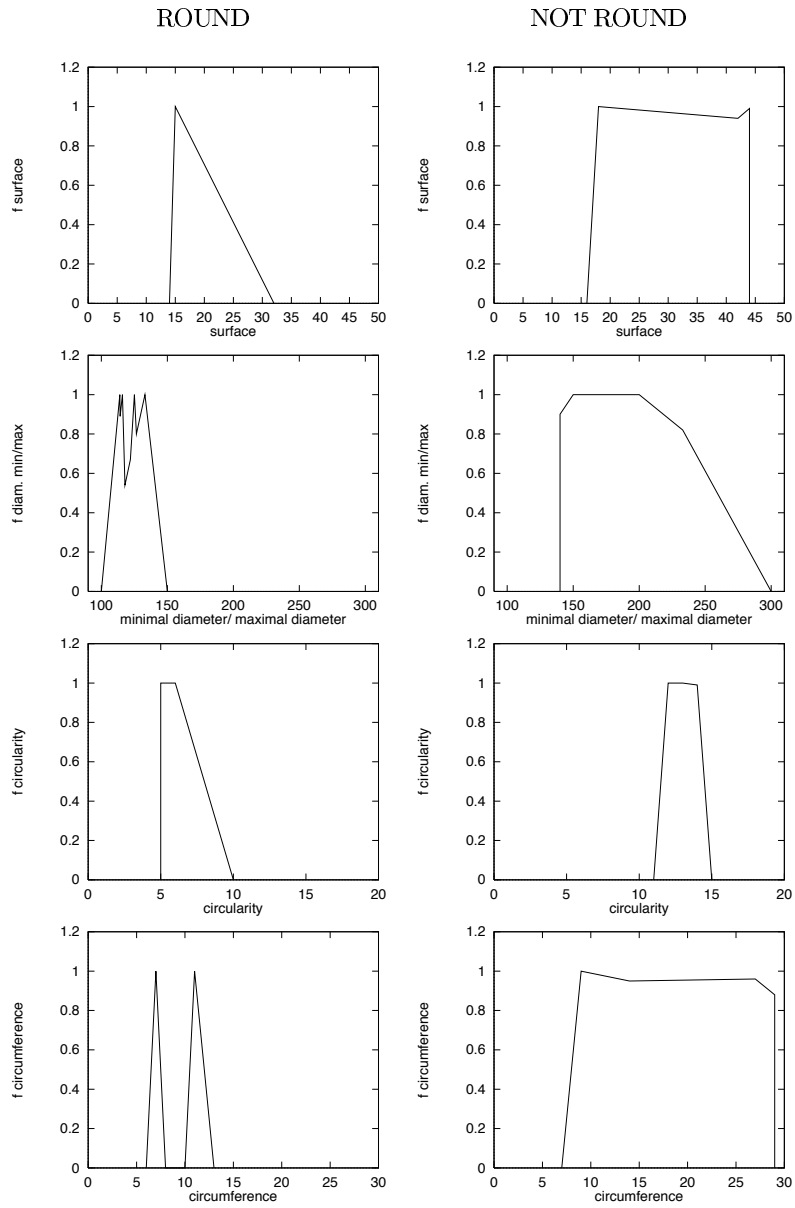


Fig. 11. Prototypes of the classes “round” and “not round”

shows that the microcalcifications in the class *round* have a surface of about 15 pixels and at the same time that the surface of microcalcifications in the class *not round* is less specific (between 16 and 44 pixel).

3.4 Conclusion

The proposed formal framework of comparison measures can help to choose the right fuzzy measures for different tasks which require comparisons between fuzzy objects. The two applications in mammography to which we have applied it (the classification of microcalcifications, and the matching of microcalcifications) have confirmed its efficiency. Furthermore, degrees of typicality based on measures of comparison can be used for the construction of fuzzy prototypes which are effective for a classification problem [22].

Part2: Applications

In this part we will discuss applications of the fuzzy methods described above. We will first describe a system for the semiological classification of microcalcification clusters. In the second section we will show some results of the application of fuzzy comparisons to the matching of microcalcifications seen on different views of a stereotactic exam. The matching is used for the reconstruction of a 3D representation of a microcalcification cluster.

4 Computer-Aided Diagnosis for mammography

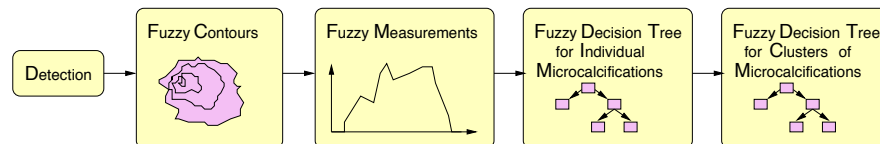


Fig. 12. Overview of the CAD system

Systems for computer-aided diagnosis for breast cancer are in most cases multi-level decision processes. The knowledge of uncertainty in each part of the processing is important for a decision process, if the uncertainty can be transmitted from step to step. In the computer-aided detection and characterization system we describe here (fig. 12), this principle has been realized using a fuzzy representation. The steps of the algorithm are:

- detection of the microcalcifications

- fuzzy segmentation of the microcalcifications
- fuzzy measurements on each calcification
- fuzzy decision tree for the classification of individual microcalcifications
- fuzzy decision tree for the classification of microcalcification clusters.

4.1 Automated Construction a Fuzzy Decision Tree for the Classification of Lesions from a Learning Database

The image database is transformed into an attribute database by using the fuzzy measurements described in sections 1. Each case is characterized by 73 attributes. We have to classify them to give a diagnosis. Radiologists provide us some indications about the way they examine the mammographic images. But they only study the films, and have only qualitative information. On the contrary, we have digital images, with measurable qualitative information. Images can also be filtered, and they provide a kind of data ignored by the radiologists (gradient, ...). Among all these attributes, a few are relevant and should be identified by a learning system.

We choose the decision tree as a learning and classification system for the following reasons:

- It can learn from a small database (our database, which is noisy, has 118 examples for 73 attributes),
- It is able to handle a large number of attributes, choosing only the most relevant among them,
- The rules provided by the radiologists can be easily translated into nodes in the tree,
- It is possible to track the path followed by an example through the tree and to compare this path to the radiologist reasoning.

4.2 Decision Tree

Learning step: The tree looks for a partition of the database according to an attribute, which is well suited to the ideal partition realized by the class. The values of attributes we use are numerical, and the partition corresponds to a threshold. The error between the two partitions is evaluated with an entropy (as in the method ID3 proposed by Quinlan [20]), or other measures (contrast, ...). We use the following contrast measure:

$$C = \frac{\int \min(f_{A_1}(x), f_{A_2}(x))dx}{\int \max(f_{A_1}(x), f_{A_2}(x))dx} \quad (15)$$

If f_{A_1} is the membership function for the attribute *compactness* for all malignant calcifications and f_{A_2} the membership function for the same attribute for all benign calcifications, C gives the contrast for the attribute *compactness*. The attribute which minimizes this contrast is chosen to constitute a test

associated with the node. Each element of the partition follows the branches of the tree, and the operation is repeated, until a predefined stopping criterion is reached (e.g. the homogeneity of the classes of the elements in the leaf is sufficiently high). An example for a decision tree, constructed in this way from a training set of different shaped microcalcifications, is shown in figure 13

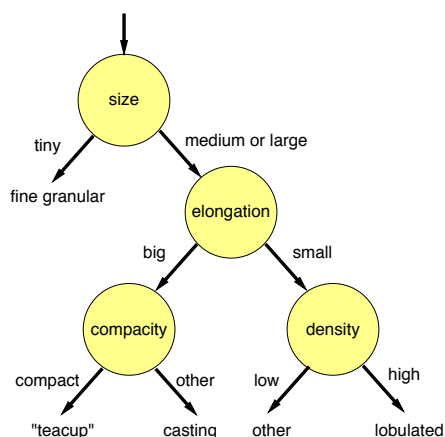


Fig. 13. Decision tree

Classification step: A new example is introduced at the root of the tree. At each node a fuzzy measure of satisfiability is calculated to estimate to which degree the example satisfies the test on the attribute. According to the outcome of this measure, the example is propagated to one or several branches with its corresponding degree of satisfiability. At the end, the example falls into one or several leaves with different degrees which are aggregated to determine the final class of the example.

In our implementation, we have used the following satisfiability:

$$S(A, B) = \frac{\int f_{A \cap B}(x) dx}{\int f_B(x) dx} \quad (16)$$

and the Zadeh t-norm (i.e. the minimum) for the aggregation operator.

4.3 Classification Results

We compared the performance of our approach with several different approaches for the classification of microcalcifications into small, round and elongated classes which are the usual categories used by the doctors to determine the risk of cancer (table 1). It is not easy to compare an optimized

fuzzy logic approach with crisp methods that are not necessarily optimized. Nevertheless, it is interesting to have some comparative figures. We tested our database with the MLC (Inducer) software. The techniques provided by this software handle only crisp data, therefore the database has been defuzzified with the max operator for those methods.

Method	KNN	Naive Bayes	ID3	Fuzzy tree
Success for “small” class	91.8 %	93.7%	93.1%	96.4 %
Success for “round” class	79.6 %	74.1%	84.2%	84.4 %
Success for “elongated” class	73.5 %	72.8%	80.1%	83.1 %

Table 1. Comparison of fuzzy tree performance with other classification methods

These results represent the characterization of individual microcalcifications. We added a second decision tree to take into account the global cluster attributes. For difficult cases, especially those with many superimpositions, we observed that even if a high level of uncertainty remains in the classification of some of the individual microcalcifications, the system can make use of the more certain measurements of the majority of microcalcifications in the cluster to give a valid result. We think this behavior is similar to that of radiologists.

5 3D Reconstruction of Microcalcification Clusters

When breast lesions containing ambiguous microcalcifications are detected, a biopsy using stereotactic guidance can be used to confirm the diagnostic. The stereotactic exam consists in taking two views of the breast from two different directions. If the object to be punctured - typically a microcalcification - can be detected on both views and the geometry of the acquisition is known, the 3D position of the object can be calculated. A positioning device can then be used to introduce a needle at the exact position of the object, to obtain a tissue sample. Figure 14 shows images obtained for the two angles and the *scout view*, which is a centered view that is taken to verify the correct positioning. Generally, the radiologist has to identify the object to be analyzed on both images from the two images only.

The application we show in this chapter is the 3D reconstruction of the microcalcification cluster from these two views.

5.1 Microcalcifications Matching

The 3D reconstruction of a cluster microcalcification cluster from two or more views requires that the microcalcifications can be matched between the views.

An overview of the matching process we propose is shown in figure 15. The first step is the detection of all microcalcifications on all views. We

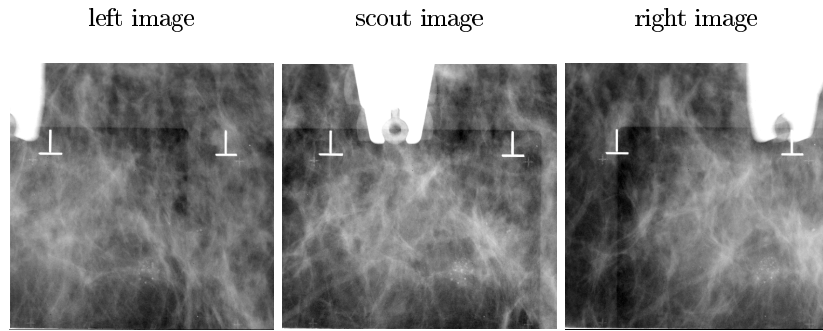


Fig. 14. Stereotactic exam

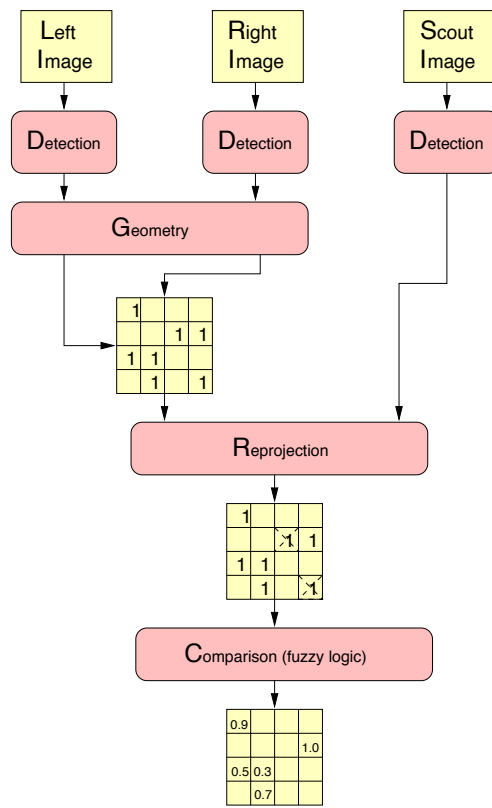


Fig. 15. Matching process for microcalcifications

construct a correspondence matrix which has one row for every candidate in the left image and one column for every candidate in the right image. This correspondence matrix is initialized to 1.0 for each element to indicate that all matches between any pair of candidates is possible. The geometrical constraints from the known acquisition geometry (epipolar line) are used to eliminate all candidates combinations which are not possible by setting to zero the corresponding matrix component. If a scout view was taken (a third view with a centered tube angle) it can be used to eliminate further candidate combination which do not project at the right place on the third view.

5.2 Using Measures of Similarity and Dissimilarity for Comparing Microcalcifications

The remaining candidates are compared by using fuzzy attributes calculated on the fuzzy contours of the calcifications. All candidates from one view are compared to all remaining candidates in the second view by calculating similarity measures based on the fuzzy attribute. We use the same approach as for the construction of a fuzzy prototype of a class of objects (see 3.3). For a given microcalcification in one view we need to find the *most similar* microcalcification in the other view and *at the same time, the most dissimilar* microcalcification to the other microcalcifications it matches. For the same reasons given for the construction of a fuzzy prototype, the similarity is determined thanks to the computation of a resemblance measure.

Comparisons of microcalcifications are made attribute by attribute. A step of aggregation is needed to give a single index of similarity or dissimilarity.

We denote the value of attribute j of the microcalcification l for the view i by $m_{l,j}^i$. The method is decomposed as follows:

For a given microcalcification on the first view (described by $m_{l,1\dots J}^1$ or in short m_l^1),

- Step 1. For each non-zero cell in the correspondent column (i.e. a matched microcalcification m_k^2 in the second view), compute the resemblance $r(m_{l,j}^1, m_{k,j}^2)$ for all j . The global resemblance $R(m_l^1, m_k^2)$ is obtained by aggregating the degrees $r(m_{l,j}^1, m_{k,j}^2)$.
- Step 2. For each non-zero cell in the row of m_k^2 (i.e. a matched microcalcification m_h^1 in the first view), compute the dissimilarity $d(m_{k,j}^2, m_{h,j}^1)$ for all j . The dissimilarity $D(m_k^2, m_h^1)$ is obtained by aggregating the degrees $d(m_{k,j}^2, m_{h,j}^1)$. The global dissimilarity $D(m_k^2)$ of m_k^2 relative to all matched m_h^1 is given by aggregating the degrees $D(m_k^2, m_h^1)$.

The result is a matrix obtained from the correspondence matrix where, in each non-zero cell, two indices are computed: the resemblance index and the dissimilarity index.

Several approaches can be imagined to use the correspondence matrix. The most obvious possibility is to aggregate the resemblance index and the dissimilarity index into a single similarity value A by taking the value obtained by:

$$A(m_i^1, m_k^2) = R(m_i^1, m_k^2) \wedge \neg D(m_k^2) \quad (17)$$

or more generally:

$$\top(R(m_i^1, m_k^2), n(D(m_k^2))) \quad (18)$$

where \top is a triangular norm and n is a strict negation.

If the remaining uncertainty about the matching is to be transmitted to the user, the remaining candidates may be visualized in 3D using a color indication for the similarity measure which can be interpreted as a confidence level in the reconstruction position. An example for this approach is shown in the following section.

In most cases the visualization of results after defuzzification is preferable. In other words only the reconstructions with the highest confidence levels are shown. In this case we have to define a defuzzification method for the similarity matrix.

Several defuzzification methods can be imagined:

1. take the maximal similarity value in the matrix and display it, remove the example from both images (take out the row and column of the similarity matrix) and continue with the next maximum
2. find the combination of candidates that maximizes the total sum of similarity values without using a candidate in two combinations
3. maximize the number of possible combinations and the sum of similarity values without using a candidate in two combinations
4. maximize the contrast (similarity between the chosen candidates versus similarity with regard to the other candidates)

From a complexity point of view, the first method is by far the most favorable, but its tendency to emphasize the maximal similarity values, even if this choice prevents many other matches with a slightly lower similarity, is problematic. The choice of a defuzzification method for this type of application with constraints between the fuzzy values needs further investigation.

5.3 3D Visualization

In figure 16 the reconstructed 3D-positions corresponding to the result of the matching process are shown. Each candidate is represented by a small point in 3D in front of one of the images acquired during the stereotactic exam. The image is positioned exactly where the detector was during the exam. In the top row all combinations of candidates are shown from three different positions, the underlying image is the scout view. In the middle row the candidates are shown as if the eye position were at the X-ray tube position

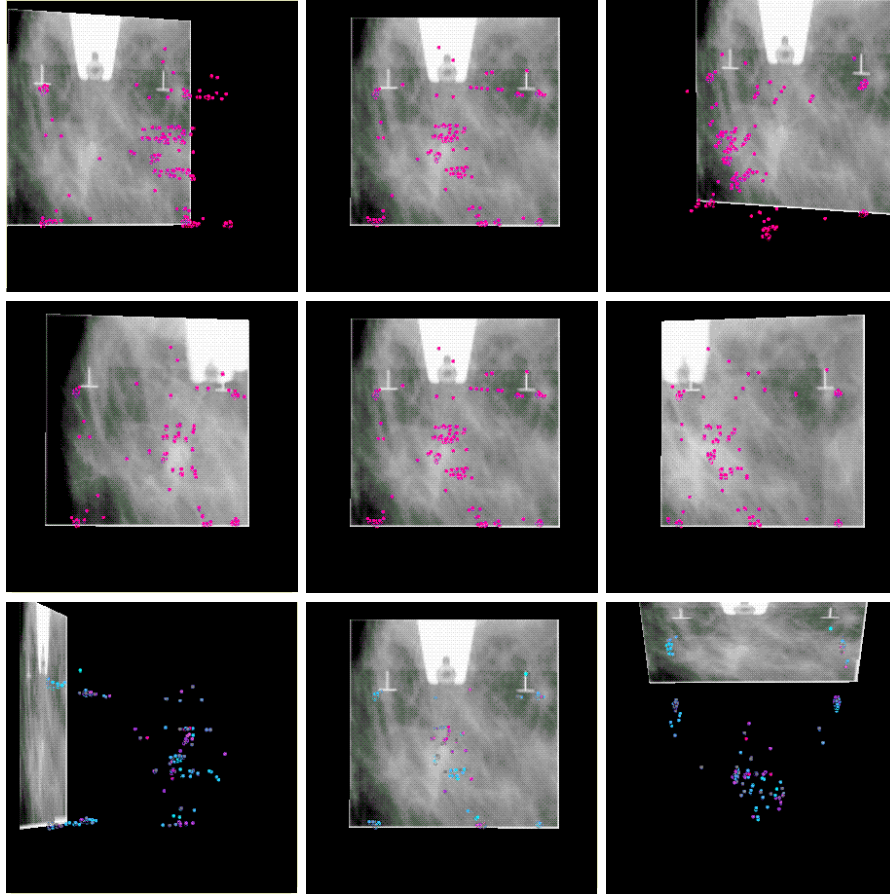


Fig. 16. Visualization of the reconstructed clusters

to show the exact projection onto the acquired images. In the last row, the similarity measure of the remaining candidates after projection on the scout view is shown using different shades of grey. The most similar correspondences are shown in dark grey, the most different in light grey.

Conclusion

Fuzzy techniques can effectively be used in mammographic image processing to represent and treat the uncertainty present in the different steps of the processing. The use of fuzzy techniques during the segmentation allows to capture the ambiguity of the segmentation without taking a decision about the true contour too early. Fuzzy rules (as a part of a fuzzy decision tree or any other framework) can be useful to represent expert knowledge or knowledge

acquired by learning from a database. Adaptive membership functions can be used to represent class memberships even with varying attributes so that fuzzy rules remain valid across large variations.

Using these fuzzy techniques, many applications in mammographic image processing can be realized. In the second part of this paper we presented two of them: the classification of breast lesions and the 3D reconstruction of microcalcifications clusters from a stereotactic exam.

References

1. Stephen R. Aylward, Bradley M. Hemminger, Etta D. Pisano *Mixture Modeling for Digital Mammogram Display and Analysis*, in: Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erming (Eds.) *Digital Mammography*, Nijmegen 1998, Kluwer Academic Publishers, 1998, pp. 305 - 312
2. Sylvie Bothorel, *Analyse d'image par arbre de décision floue - Application à la classification sémiologique des amas de microcalcifications en mammographie numérique*, Thèse de l'université Paris 6, 1996
3. Bothorel, S., Bouchon B., S. Muller, *A Fuzzy Logic-based Approach for Semiological Analysis of Microcalcification in Mammographic Images*, *International Journal of Intelligent Systems*, vol. 12, 1997, pp. 819-843
4. B. Bouchon-Meunier, M. Rifqi and S. Bothorel, *Towards general measures of comparison of objects*, *Fuzzy Sets and Systems*, Vol. 84, N. 2, pp.143-153, 1996,
5. B. Bouchon-Meunier and L. Valverde, *Analogy relations and inference*, *IEEE International Conference on Fuzzy Systems*, San Fransisco, pp. 1140-1144, 1993
6. B. Bouchon-Meunier, *Fuzzy similitude and approximate reasoning*, *Advances in Fuzzy Theory and Technology*, Editor P. P. Wang, Bookwrights Press, pp. 161-166, 1993
7. A. De Luca and S. Termini, *A definition of a non-probability entropy in the setting of fuzzy sets theory*, *Information and Control*, Vol. 20, pp.301-312, 1972
8. M. Demster, N.M. Laird, D.B.Jain, " *Maximum likelihood for incomplete data via the EM algorithm*, *J. Royal Statistics Society, Series B*. Vol. 39, 1977, pp 1 - 38
9. Atam P. Dhawan, Yateen Chitre, Christine Kaiser-Bonasso, Myron Moskowitz, *Analysis of mammographic Microcalcifications Using Gray-Level Image Structure Features*, *IEEE Transactions on Medical Imaging*, Vol. 15, No.3, June 1996, pp. 246 - 259
10. D. Dubois and H. Prade, *Fuzzy Sets and Systems, Theory and Applications*, Academic Press, New- York, 1980
11. Takehiro Ema, Kunio Doi, Robert M. Nichikawa, Yulei Jiang, John Papaioannou, *Image feature analysis and computer-aided diagnosis in mammography: Reduction of false-positive clustered microcalcifications using local-gradient analysis*, *Medical Physics*, Vol. 22, No.2, February 1995, pp. 161 - 169
12. Michel Grimaud, *La géodesie numérique en morphologie mathématique. Application a la détection automatique de microcalcifications en mammographie numérique*, Thèse de doctorat à l'Ecole Nationale Supérieure des Mines de Paris,1991
13. R. Gupta, P. E. Undrill, *The use of texture analysis to delineate suspicious masses in mammography*, *Phys. Med. Biol.*, Vol. 40, 1995, pp. 835 - 855

14. Ralph Highnam, Michael Brady, Brasil Shepstone *A representation for mammographic image processing*, Medical Image Analysis, Vol. 1, No.1, 1996, pp. 1 - 18
15. Nico Karssemeijer, Guido M. te Brake *Detection of Stellate Distortions in Mammograms*, IEEE Transactions on Medical Imaging, Vol. 15, No.5, October 1996, pp. 611 - 619
16. Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erning *Digital Mammography*, Nijmegen 1998, Kluwer Academic Publishers, 1998
17. Marton Lanyi, *Diagnosis and Differential Diagnosis of Breast Calcifications*, Springer Verlag 1986
18. Christophe Marsala, *Apprentissage inductif en présence de données imprécises: construction et utilisation d'abres de décision flous*, Thèse de l'informatique, Université Paris 6, 1998
19. S. V. Ovchinnikov, *Representations of transitive fuzzy relations*, Aspects of Vagueness, Editors H. J. Skala and S. Termini and E. Trillas, D. Reidel Publishing Company, pp. 105-118, 1984
20. Quinlan, *Induction of decision trees*, Machine Learning, Vol. 1 No. 1, 1986, pp. 86-106
21. Maria Rifqi, *Mesures de comparaison, typicalité et classification d'objets flous: théorie et pratique*, Thèse de doctorat de l'informatique, Université Paris 6, 1996
22. M. Rifqi, S. Bothorel, B. Bouchon-Meunier and S. Muller, *Similarity and prototype based approach for classification of microcalcifications*, 7th IFSA World Congress, Prague, 1997 pp. 123-128,
23. E. Rosch, *Principles of categorization, Cognition and categorization*, Hillsdale, N. J. : Laurence Erlbaum Associates, 1978
24. E. Rosch and C. B. Mervis, *Family resemblances: Studies in the internal structure of categories*, Cognitive Psychology, Vol. 7, pp. 573-605, 1975
25. E. Sanchez, *Inverses of fuzzy relations. Applications to possibility distributions and medical diagnosis*, Fuzzy Sets and Systems, Vol. 2, pp. 75-86, 1979
26. K. Shiina, *A fuzzy-set-theoretic feature model and its application to asymmetric data analysis*, Japanese psychological research, Vol. 30, pp. 95-104, 1988
27. E. Trillas and L. Valverde, *On implication and indistinguishability in the setting of fuzzy logic*, Management Decision Support Systems Using Fuzzy Sets and Possibility Theory, Editors J. Kacprzyk and R. R. Yager, Verlag TUV, Rheinland, 1984
28. A. Tversky, *Features of similarity*, Psychological Review, Vol. 84, pp. 327-352, 1977
29. L. Valverde, *On the structure of t-indistinguishability operators*, Fuzzy Sets and Systems, Vol. 17, pp. 313-328, 1985
30. L. A. Zadeh, *A note on prototype theory and fuzzy sets*, Cognition, Vol. 12, pp. 291-297, 1982
31. Xinhua Zhuang, Yan Huang, K. Palaniappan, Yunxin Zhao, *Gaussian Mixture Density Modeling, Decomposition, and Applications*, IEEE Transactions on Image Processing Vol. 5, No.9, Sep 1996, pp. 1293-1302