Model-Based Region-of-Interest Selection in Dynamic Breast MRI

by

Florence Forbes, Nathalie Peyrard, Chris Fraley, Dianne Georgian-Smith, David M. Goldhaber, Adrian E. Raftery

TECHNICAL REPORT No. 472

December 2004

Department of Statistics, Box 354322
University of Washington
Seattle, Washington 98195 USA

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Model-Based Region-of-Interest Selection in Dynamic Breast MRI	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Department of Statistics,Box 354322,Seattle,WA,98195-4322**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Magnetic Resonance Imaging (MRI) is emerging as a powerful tool for the diagnosis of breast abnormalities. Dynamic analysis of the temporal pattern of contrast uptake has been applied in differential diagnosis of benign and malignant lesions to improve specificity. Selecting a region of interest (ROI) is an almost universal step in the process of examining the contrast uptake characteristics of a breast lesion. We propose an ROI selection method that combines modelbased clustering of the pixels with Bayesian morphology, a new statistical image segmentation method. We then investigate tools for subsequent analysis of signal intensity time course data in the selected region. Results on a data base of 19 patients are promising. The method provides informative segmentations and good detection rates are obtained.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**36**	

Model-Based Region-of-Interest Selection in Dynamic Breast MRI *

Florence Forbes[†], Nathalie Peyrard[‡], Chris Fraley[§], Dianne Georgian-Smith[¶], David M. Goldhaber[‖], Adrian E. Raftery[**]

Technical Report No. 472

Department of Statistics
University of Washington
December 2004

Abstract

Magnetic Resonance Imaging (MRI) is emerging as a powerful tool for the diagnosis of breast abnormalities. Dynamic analysis of the temporal pattern of contrast uptake has been applied in differential diagnosis of benign and malignant lesions to improve specificity. Selecting a region of interest (ROI) is an almost universal step in the process of examining the contrast uptake characteristics of a breast lesion. We propose an ROI selection method that combines model-based clustering of the pixels with Bayesian morphology, a new statistical image segmentation method. We then investigate tools for subsequent analysis of signal intensity time course data in the selected region. Results on a data base of 19 patients are promising. The method provides informative segmentations and good detection rates are obtained.

Index terms—Bayesian morphology, magnetic resonance imaging, Markov random fields, model-based clustering, region of interest, time-signal intensity curves.

1 Introduction

Magnetic Resonance Imaging (MRI) is emerging as a powerful tool for the diagnosis of breast abnormalities. Its unique ability to provide morphological and functional information can be used to assist in the differential diagnosis of lesions that other methods find questionable [1].

*This research was supported by NIH grant 8 R01 EB002137-02, by ONR grants N00014-01-10745, N00014-96-1-0192 and N00014-96-1-0330, and by Toshiba America MRI, Inc. The authors are grateful to Leon Kaufman, David Haynor, Doug Ortendahl, and Brad Wyman for useful comments and discussions, and to Andrew Jianhua Li for providing the software for computing the five intensity parameters from the image data.

[†]Florence Forbes, équipe MISTIS, Inria Rhône-Alpes, Zirst, 655 av. de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France. (Florence.Forbes@inrialpes.fr)

[‡]Nathalie Peyrard, Biometrics and Artificial Intelligence Department of INRA, Domaine Saint Paul, site Agroparc, 84914 Avignon cedex 9, France. (nathalie.peyrard@avignon.inra.fr)

[§]Chris Fraley, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322. (fraley@stat.washington.edu)

[¶]Dianne Georgian-Smith, M. D., Harvard Medical School, Department of Radiology, Massachusetts General Hospital, Wang ACC 219Q, 15 Parkman St, Boston, MA 02114. (dgeorgiansmith@partners.org)

[‖]David M. Goldhaber, International MRI Accreditation Resources, 400 Grandview Drive, South San Francisco, CA 94080. (d_g@pacbell.net)

[**]Adrian E. Raftery, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322. (raftery@stat.washington.edu)

MRI has also proved useful in the evaluation of the extent of breast cancer as well as in treatment planning. It appears as a complementary diagnostic modality in breast imaging.

Because of the high reactivity of breast carcinomas after Gadolinium injection, MRI has the potential to allow differentiation between malignant and benign tissues. However, some benign lesions also enhance, as a result reducing the specificity of MRI. Dynamic analysis of the temporal pattern of contrast uptake has been applied to improve specificity. For reviews of the knowledge base on detection and differential diagnosis of breast tumors, see [2] and [1]. The diagnostic criteria that are in use for differential diagnosis can be divided into those related to lesion enhancement kinetics and those related to lesion morphology. The results in [3] suggest that signal intensity time course data are useful for differentiating benign from malignant enhancing lesions. The authors conclude that the overall shape of the time-signal intensity curve is an important criterion, while a single attribute of the curve, such as the enhancement rate, may not be enough.

The evaluation of morphologic features and the extraction of architectural information is usually also based on post-contrast images of enhancing areas so that the above distinction is somewhat arbitrary. Integrating multiple diagnostic criteria (both qualitative and quantitative) is therefore recommended. Selecting a region of interest (ROI) is an important first step in the process of examining the contrast uptake characteristics of a breast lesion from either a morphologic or kinetic point of view. However, no standard method for ROI selection and analysis of dynamic breast MR data has yet been established.

A limitation of many ROI analysis procedures is that only subjectively selected regions are examined. Less subjective selection approaches have also been proposed. The work in [4] proposes a semi-automatic method that consists in selecting as ROI the 3×3 block of pixels with the highest mean enhancement in a manually pre-selected larger suspect region. Another method using a single variable quantifying the enhancement is proposed in [5]. ROI selection is performed by classifying pixels into suspect and nonsuspect pixels, using the Graduated Non-Convexity algorithm.

As regards tissue classification, there has been considerable research in brain MRI. Many methods are based on modeling the image intensity with a Gaussian mixture model via the Expectation-Maximization algorithm [6]. Extensions and variations allow the integration of

spatial information into the classification process, using Markov random fields [7, 8]. However there are differences in the analysis of breast MRI and brain MRI, and less research has been devoted to the former. In breast MRI analysis, there is a greater centrality of image segmentation (e.g. finding the ROI). Also, breast tissues are much more heterogeneous than brain tissues. Normal breasts can consist almost entirely of fatty tissues or include extremely dense fibroglandular tissues. This results in additional challenges for the analysis of breast MRI. Also, in dynamic breast MRI, the information to be modeled at each pixel is not a single intensity measure, as is usual in brain MRI, but a whole signal intensity time curve.

In this paper, we propose an unsupervised ROI selection method based on statistical techniques. We describe a multivariate classification method that enables us to take account of more than one measurement. It results in color classification images in which parts of the breast with similar signal intensity time courses are assigned to a class represented by a color. This gives morphological information that can be used to select an ROI by focusing on the pixels with the strongest enhancement. We have also developed some tools for analyzing the enhancement kinetics for pixels in the selected region.

In the following section, we describe our data set for this study. In Section 3 we present the multivariate classification method. We describe the model-based clustering method used, along with complementary procedures to include spatial information. In Section 4 we discuss how to use the resulting classifications for ROI selection and enhancement kinetics analysis, and we also propose techniques for improving differential diagnosis based on the shapes of the curves in the selected region. The procedure is then illustrated and the results for our data set reported in Section 5.

2 Data

We considered sequences of images for 19 patients representing different cases (malignant and benign lesions). Several two-dimensional slices were available for each patient. For each slice, 25 sequential magnetic resonance (MR) breast images were acquired (one image approximately every 10 seconds). See Figure 1 for examples of such images. Each image records the signal intensity at a given time after injection. Instead of working directly with these MR images, we summarized them in terms of five derived variables considered to be of significance for cancer

diagnosis. These five variables are calculated from a curve fitting procedure developed at Toshiba America MRI, Inc. [9]. A fit analysis is carried out at each pixel location for the signal intensity curve. Figure 1 (e) shows a signal intensity curve at a given pixel, after subtraction of the reference signal. The fitting model is assumed to be made of three successive sections: a zero signal, a second order polynomial curve and a flat line.

We used the following five derived variables in our study:

- **Time to Peak:** the time at which the signal peaks.
- **Difference at peak:** absolute increase of intensity between the beginning of the signal and the time at which the signal peaks.
- **Enhancement slope:** in units of intensity/time.
- **Maximum step:** maximum change between two adjacent dynamic samples.
- **Washout slope:** in units of intensity/time.

In addition to the images, diagnostic information is available. Among the 19 patients, 12 have tumors diagnosed as carcinomas and 5 are diagnosed as not having cancer. For two others, the diagnosis is ambiguous. In addition, for two of the patients, follow-up data sets are available but with no associated diagnosis. See Section 5 for more details.

Our starting point is thus five images for each case, one showing the values of each derived variable at each pixel location, rather than the original 25 images. Although this preprocessing reduces the amount of data to be analyzed, the characterization of breast lesions based on these MR images remains a difficult task. In Section 3, we present the multivariate statistical methods for clustering and spatial segmentation we propose to synthesize the available information into a single classification image.

3 Producing Classification Images

We propose to use statistical segmentation methods to produce a color image for each case, in which each color represents a group or class of pixels with similar time-signal intensity curves (summarized by the five derived variables). An important issue is the determination of the effective number K of components present in the data, *i.e.* the number of colors to use in the classification image. The main components in the breast are blood vessels, air, fat, a possible

tumor and other tissues of less interest. Because of the large number of pixels involved, those corresponding to air are eliminated prior to further analysis, leaving three or four components, depending on whether or not there is a tumor. We therefore considered segmentations into three or four classes.

We also investigated the possibility that allowing more than four classes may provide better statistical performance in terms of identifying the main features of interest in the image. In this connection, we assessed the possibility of using a statistical method to determine the number of classes based on the data. We did this using the Bayesian Information Criterion (BIC) [10]. The BIC is computed given the data and a model and allows comparison of models with differing parameterizations and/or differing number of classes. It is the value of the maximized model loglikelihood with a penalty for the number of parameters in the model. It can be viewed as providing an approximation to the Bayes factor, which is the standard Bayesian approach to model selection [11]. BIC can be compared to other selection criteria. One of them is the Akaike Information Criterion (AIC) of [12] which differs from BIC in the penalty term but has been shown to overestimate the number of parameters in practice. The MDL criterion proposed in [13] is based on stochastic complexity and is similar to BIC, and methods using cross validation ([14]) seem promising but their tractability in our context is not straightforward due to the dependence structure in the data. Many other approaches can be found in the literature on model selection (see for instance the list of references in [11]). BIC has become quite popular due to its simplicity and its good results.

However, BIC tended to select values of K between 10 and 15, which accurately reflects the inhomogeneity of some kinds of tissue, but turned out not to help identifying tumors. In what follows, we have reported results for $K = 3$, 4 and 10. Overall, we found that using $K = 4$, as suggested by the underlying biology, performed best.

Model-based statistical methods for clustering multivariate observations are flexible and have been widely applied [15, 16, 17, 18, 19]. However for complex data such as those associated with tissue segmentation in medical imaging, these methods can produce slightly noisy results that do not correspond directly to a meaningful classification, because they do not take spatial dependence into account. For this reason, we propose refining the clustering results of Section 3.1 with a spatial statistical technique called Bayesian morphology [20]. Small isolated regions are

removed by automatically reassigning pixels located in them, reducing the spatial fragmentation of the classification.

3.1 Model-Based Cluster Analysis

We propose to use *marginal mixture EM segmentation* as a first step in our analysis. The idea is to model the marginal distribution of (possibly multivariate) pixel intensities as a finite Gaussian mixture model, and use the EM (Expectation-Maximization) algorithm [21, 22] to estimate the model parameters.

We used the MCLUST software for model-based clustering [17, 19]. It combines model-based Gaussian agglomerative hierarchical clustering methods [15, 23]) with the EM algorithm for Gaussian mixture models [24]. The EM algorithm is an iterative method widely used in parameter estimation for incomplete data. For clustering applications via mixture models, the missing values are the cluster membership probabilities of each pixel. To be effective, the EM algorithm generally requires a good initial estimate. Classifications produced by model-based Gaussian agglomerative hierarchical clustering, which are often good but rarely optimal, are a good way to initialize the EM algorithm for classification ([25],[16]).

In what follows, observations (corresponding to image pixels) are denoted by y_i and are assumed to be five-dimensional vectors, corresponding to the five derived variables. The y_i are assumed to arise from a K-component Gaussian mixture model, so that if y_i belongs to class $k \in \{1, \ldots, K\}$, its distribution is multivariate normal (Gaussian) with mean vector μ_k and covariance matrix Σ_k. The likelihood for our data is then

$$\mathcal{L}(\theta_1, \ldots, \theta_K; p_1, \ldots, p_K \mid y) = \prod_{i=1}^{n} \sum_{k=1}^{K} p_k f_k(y_i \mid \theta_k), \tag{1}$$

where $\theta_k = (\mu_k, \Sigma_k)$ and f_k is the multivariate normal density of the kth component in the mixture, parameterized by its mean μ_k and covariance matrix Σ_k:

$$f_k(y_i \mid \mu_k, \Sigma_k) \equiv \frac{\exp\left\{-\frac{1}{2}(y_i - \mu_k)^T \Sigma_k^{-1} (y_i - \mu_k)\right\}}{\sqrt{\det(2\pi\Sigma_k)}}. \tag{2}$$

Here p_k is the probability that an observation belongs to the kth component ($p_k \geq 0; \sum_{k=1}^{K} p_k = 1$).

Data generated by mixtures of multivariate normal densities are characterized by groups or clusters centered at the means μ_k. The corresponding surfaces of constant density are ellipsoidal.

Geometric features (shape, volume, orientation) of the clusters are determined by the covariances Σ_k, which may also be parameterized to impose cross-cluster constraints. There are a number of possible parameterizations of Σ_k [15], [24], many of which have been implemented in MCLUST. Common instances include $\Sigma_k = \lambda I$, where all clusters are spherical and of the same size; constant $\Sigma_k = \Sigma$, where all clusters have the same geometry but need not be spherical [26]; and unrestricted Σ_k, where each cluster may have a different geometry [27]. Banfield and Raftery [15] proposed a general framework for geometric cross-cluster constraints in multivariate normal mixtures by parameterizing covariance matrices through eigenvalue decomposition in the form $\Sigma_k = \lambda_k D_k A_k D_k^T$, where D_k is the orthogonal matrix of eigenvectors, A_k is a diagonal matrix whose elements are proportional to the eigenvalues, and λ_k is an associated constant of proportionality. Their idea was to treat λ_k, A_k and D_k as independent sets of parameters, and either constrain them to be the same for each cluster or allow them to vary among clusters.

We obtained a first segmentation via MCLUST with the constant-shape model $\Sigma_k = \lambda_k D_k A D_k^T$. The algorithm provides an estimate of the conditional probability that each pixel belongs to each of the classes, given the observations. These probabilities are obtained using the EM algorithm. The segmentation derived from these conditional probabilities is the one which assigns each pixel to the class with the greatest probability.

3.2 Spatial classification

In this section, we discuss refinements of our model-based classification to incorporate spatial information. In Section 3.2.1, we describe image smoothing via morphological filters, which treat each pixel in the initial classification only in the context of neighboring pixels. In Section 3.2.2, we discuss methods that make use of the original data in addition to the initial model-based classification.

3.2.1 Morphological filters

Morphological filters are procedures that successively apply a morphological rank operator to each pixel and its neighbors until there are no further changes [28, 29, 30]. Each pixel is considered in conjunction with its 8 surrounding neighbors, and the rank operator depends on a parameter r (the rank) as follows. Let c be the color which is taken most often by the neighbors of pixel i and let z_i be the current color of pixel i. We denote by $nb(k)$ the number of neighbors

of pixel i which have color k. If $nb(c)$ is greater than or equal to $nb(z_i) + r$, the color of pixel i is changed to c; otherwise the color of pixel i remains equal to z_i.

For example, when $r = 1$, this is equivalent to apply the majority rule, according to which a pixel assumes the color which is taken most often by its neighbors. This corresponds to a *median filter* except that the pixels are updated successively rather than simultaneously.

Morphological filters tend to produce very smooth classifications, and more smoothing occurs as r decreases. We refer to morphological filters as *blind restoration*, since they do not make use of the original data. They have the disadvantage that useful information may be lost in the process, and it is not difficult to find instances in which they perform relatively poorly [20].

Nevertheless, morphological filters may be useful in our application to help isolate possible tumors from the initial MCLUST classification. Blind restoration can smooth the data considerably, which can be useful because it eliminates clutter and extraneous minor features in the image. However, in images where the tumor is less clearly identified, this level of smoothing runs the risk of eliminating the tumor altogether, and should probably be used only in conjunction with less smoothed images as well. Such images can be obtained using the procedure described in the next section. It has features similar to morphological filters but is statistically based.

3.2.2 Bayesian Morphology

Bayesian image segmentation is based on probability models. The unknown classification, $z = \{z_i,\ i \in S\}$, where S is a set of pixels, is interpreted as a particular realization of a random vector Z. The observed data set y is interpreted as a realization of a random vector Y. The vector Y depends on Z through a known conditional probability density function $\mathcal{L}(y|z)$ which incorporates the observed data formation model and the noise model.

The unknown classification Z is assumed to be a realization of a random field with distribution $p(z)$. Then the estimated classification $\hat{z}$ is based on the posterior density of z, namely $p(z|y) \propto \mathcal{L}(y|z)p(z)$. A standard restoration criterion consists of maximizing this density, leading to the Maximum A Posteriori (MAP) estimate of z .

One of the most popular modeling assumptions is to consider the image z as being a realization of a Markov random field. This means that $p(z)$ satisfies $p(z_i|z_{S\setminus\{i\}}) = p(z_i|z_{N(i)})$, for all pixels i in S, i.e. the conditional distribution of $p(z)$ depends only on the values of

pixels in a subset $N(i)$ of $S\backslash\{i\}$, called the *neighborhood* of pixel i. Another usual assumption is that, given $Z = z$, the Y_i are conditionally independent and have the same conditional density function $f(y_i|z_i)$ that depends only on z_i. Thus $\mathcal{L}(y|z)$ can be written as the product $\mathcal{L}(y|z) = \prod_{i\in S} f(y_i|z_i)$.

Finding the MAP estimate under these assumptions can require heavy computation. A less computationally demanding method that provides a fast approximation to the MAP is the Iterated Conditional Modes (ICM) algorithm [31]. The ICM algorithm is iterative. Given a current estimate $\hat{z}$ of the image, a new one is computed by visiting each pixel in turn. When at pixel i, the current value there is replaced by the value that maximizes the conditional density

$$p(z_i|\hat{z}_{S\backslash\{i\}}, y), \tag{3}$$

given all other current pixel values $\hat{z}_{S\backslash\{i\}}$ and the fixed observation y. This choice is motivated by the fact that $p(z|y) = p(z_i|z_{S\backslash\{i\}}, y)\ p(z_{S\backslash\{i\}}|y)$. When pixels are updated sequentially, choosing values that maximize the conditional probability $p(z_i|z_{S\backslash\{i\}}, y)$ increases the posterior distribution and ensures the convergence to a local maximum of $p(z|y)$.

Under the previous modeling assumptions, maximizing the conditional density (3) is equivalent to maximizing $f(y_i|z_i)\ p(z_i|\hat{z}_{N(i)})$, since only the dependence on z_i is relevant for the maximization.

A widely used model for image segmentation or classification in a number K of classes is the Potts model defined below. The standard nearest-neighbor Potts model depends on a parameter β and is defined by $p(z) = Z(\beta)^{-1} \exp(\beta v(z))$, where $v(z) = \sum_{i\sim j} \delta(z_i, z_j)$ is the number of pairs of neighboring pixels having the same color in z. In the above sum, $i \sim j$ denotes the statement that the pixels i and j are neighbors, and $\delta(z_i, z_j)$ refers to the Kronecker delta function, equal to 1 if z_i and z_j are the same, and to 0 otherwise. The quantity $Z(\beta)$ is the normalizing constant, or partition function, $Z(\beta) = \sum_z \exp(\beta v(z))$. This function is usually difficult to compute because of the intractably large number of terms in the summation. The conditional distributions of $p(z)$ have the simple form $p(z_i|z_{S\backslash\{i\}}) \propto \exp(\beta u_i(z_i))$, where $u_i(z_i) = \sum_{j\in N(i)} \delta(z_i, z_j)$ is the number of neighbors of pixel i having color z_i. The model depends on a parameter β which is taken to be positive, reflecting the assumption that neighboring pixels tend to be of the same color. The true classification is then assumed to be degraded by channel transmission noise.

For all i,

$$f(y_i|z_i) = \begin{cases} 1-p & \text{if } y_i = z_i \\ \frac{p}{C-1} & \text{otherwise.} \end{cases}$$

A good transmission is assumed to be the most probable, which means that $C(1-p) > 1$.

In [20], we showed that the ICM algorithm could be formulated using morphological terminology and proposed *Bayesian morphology*, a procedure that combines the speed of mathematical morphology with the principled statistical basis of ICM. In Bayesian morphology, a succession of morphological rank operators is applied, and at each iteration, the rank is estimated from the data and a current classification. One advantage of such a mathematical morphology formulation is to reveal the existence of insensitivity conditions for the parameters. This means that the final segmentation is not sensitive to the precise values of the model parameters. This is a key observation that we use to reduce the complexity of the estimation step in traditional unsupervised ICM and to save a lot of computation time. When performed on discrete images (or if an initial segmentation has been carried out), the resulting algorithm is equivalent to ICM in the sense that the final segmentation or classification is the same. In this case, it differs from ICM essentially in the way the parameter estimation step is carried out. According to the insensitivity conditions, point estimates need not be computed.

By estimating the rank of the morphological operator at each iteration instead of using a predetermined or arbitrary chosen rank, these methods make more use of the available information than blind restoration, and as a result tend to produce classifications with more detail. In comparison to blind morphology, less noise is likely to be eliminated, while ambiguous features worthy of further consideration are more likely to be retained. These classifications provide a first tool for guiding diagnosis. Often the lesion is easily identified and the radiologist can be asked to select regions that are suspicious or otherwise of interest to be further examined. In the next section, we show that additional information is present in the data that can be used for a more automatic detection and localisation of lesions of interest.

4 Region Of Interest Selection

We now propose a way to automatically select an ROI using the color classification images produced in the first part of our study (Section 3). Our second step consists of studying the

values of each of the five derived variables within each class, on the basis of which we propose an ROI selection. The analysis can then be carried out by looking at the shape of the ROI (morphologic feature analysis) or at the curves (kinetics analysis) for pixels in a selected cluster or region in order to identify the nature of the lesion.

4.1 Deciding which of the segments is the ROI

In breast MRI, lesions are usually identified because they enhance after intravenous injection. Although the pathophysiologic basis of this feature is not yet been fully known, some basic facts are known that should help in designing tools for lesion detection and differential diagnosis [1, 2]. In our study, we focus on rapidly and strongly enhancing lesions. We look at enhancement rates because malignant lesions tend to enhance more quickly than benign ones [1]. Strong enhancements are characterized by a large *difference at peak*, *i.e.* the absolute increase in intensity between the beginning of the signal and the time at which the signal peaks. For a given classification image, the mean value of *difference at peak* is computed for each class in the image. We then select the class with the largest mean value as containing the ROI, and we identify the ROI as the biggest connected component in the class, as for instance in Figure 8(e). Note that another criterion for rapid enhancement would be a small time to peak. This would correspond to a large value for the variable referred to as *time to peak* in our data, which is a linear transformation of the real time to peak. However, this criterion would sometimes select the heart area which enhances faster than the lesion of interest. Using the largest difference at peak instead, the class containing the suspicious region is always selected as desired. Illustrations are given in Section 5.2.

The difference at peak can also be used to determine a meaningful color assignment. In most classification methods, images are produced using colors (or equivalently class labels) artificially assigned to the different regions (see Table 2). In our study, we propose to automatically display our results using the highest difference at peak criterion so that suspicious regions can be marked with a pre-determined label and always displayed with a specific color (*e.g.* red).

4.2 Enhancement kinetics analysis

In diagnosis, an important point is to produce a *good* estimated pattern of uptake which should be representative of the lesion under study. A first idea is to use the ROI and compute the mean

of all signals in the region. This should get rid of some noise but may be biased if the region is too big. We could also select one or a few pixels in the ROI with the highest probability. We also investigated other ways to compute such *mean* curves using weights. The idea is to give more weight to pixels in the ROI which are typical of the lesion and less weight to pixels for which we are more uncertain. The question is then to find reasonably good weights as automatically as possible.

Let S_{roi} denote the set of pixels in the ROI and $w = \{w_i, i \in S_{roi}\}$ a vector of weights associated to each pixels. A mean curve can be computed by multiplying each signal in the ROI by $w_i / \sum_{i \in S_{roi}} w_i$ before summing all the signals. If w_i is equal to either 0 or 1, using weights is equivalent to selecting some of the pixels. The question then is which pixels to keep ($w_i = 1$) and which to discard ($w_i = 0$). As an illustration, we first kept the 33% pixels in the ROIs with the highest difference at peak values. We then used the conditional probability estimates provided by MCLUST (see Section 3.1) and kept the pixels for which the probability of belonging to the lesion class was very close to one. More generally, the estimated conditional probabilities estimates can be used as weights. See Figures 9 and 10 for an illustration of the different curves obtained this way.

4.3 Time-intensity signal analysis in the ROI

Assuming that we have assigned a representative curve to the lesion under study, our third step is then to focus on the shape of this curve. We take into account information from other sources [3, 1, 2]. They distinguish three patterns of signal intensity curves on the basis of three characteristics, the *enhancement rate*, the presence of a *plateau* and that of a *washout slope*. Type I shows a monophasic enhancement that persists until the late post-contrast period (linear time course). This type is indicative of a benign lesion. Type II is a biphasic enhancement where signal intensity reaches the maximum approximately 2-3 minutes after injection and stays at this level (plateau curve). This type has been observed in both benign and malignant lesions. Type III is characterized by a washout enhancement. As in type II, peak enhancement is already reached in the early post-contrast period but then it is followed by an intensity loss. A type III pattern strongly supports the diagnosis of a malignant breast tumor.

5 Results

To evaluate our procedure, we first focus on its ability to produce informative classification images (Section 5.1). We show detailed results for three cases. We analyze the various choices of number of segments and segmentation methods that can be made. ROI analysis results (Section 5.2) are then given for all 19 patients (see Table 3).

5.1 Breast MRI segmentation

We report detailed results for three data sets. We have a set of twenty-five 512×256 images corresponding to one slice for patient 05, containing a spherical lesion diagnosed as a carcinoma. Another set consists of twenty-five 176×352 images for slice 6 of patient 08. The MR data was obtained less than a week after surgery and the radiologists concluded that there was no residual carcinoma, *i.e.* the margins of the surgery site were not suspicious. For patient 28, three slices are available, slices 10, 11 and 12, each one consisting of twenty-five 192×192 images. A spherical carcinoma is known to be present in slices 10 and 11. After a tumor biopsy, the tumor size was estimated to be less than 16mm, which means that there should be no malignant tissues on slice 12 (a slice is 8mm).

Figure 2 shows the segmentations for 3, 4 and 10 clusters for slice 09 for Patient 05. These numbers do not include the background as a class so that the number of colors in the final segmented images is equal to the number of clusters plus one.

In all three images, one can easily recognize the heart and tumor locations. The cluster corresponding to the heart and vessels is shown in blue while the tumor is in red. The remaining colors indicate other tissues. We will refer to these three clusters as *heart, tumor* and *misc.* The latter group is composed of more than one cluster in the $K = 4$ and $K = 10$ cases.

In the segmentation obtained for $K = 3$ (Figure 2(b)), many pixels in the skin area are classified as tumor, an indication that more classes are needed for useful image segmentation and tumor identification. This conclusion is further supported by the results for $K = 4$, in which the number of red pixels in the skin area is much smaller (see Figure 2 (c)), but the tumor remains solidly red.

If we compare the size (number of observations) of each cluster for $K = 3$ and $K = 4$ (Table 1), we can see that the second cluster of type *misc* and the cluster *tumor* are merged into a

Table 1: Cluster volumes for $K = 3$ and $K = 4$ (Patient 05, slice 09).

	tumor	*heart*	$misc_1$	$misc_2$
K=3	1112	1597	707	
K=4	975	1312	751	378

single one when K decreases from 4 to 3. The behavior of the five derived variables (see Figure 3) in these two clusters is similar for the time to peak and the maximum step. The tumor shows a greater range of values for difference at peak and enhancement slope and a smaller range of values for the washout.

For both $K = 3$ and $K = 4$, the main difference between the *heart* cluster and the *tumor* cluster lies in the difference at peak, enhancement slope and washout variables. In this case the tumor shows a greater range of values for the three parameters. The additional cluster produced for $K = 4$, referred to as *misc1* in the figures, seems to be mainly composed of outliers. The enhancement slope and the washout are equal to zero for most of the pixels in this cluster. Another difference between the segmented images for $K = 3$ and $K = 4$ is the classification of the area to the left of the tumor, which is classified as tumor for $K = 3$ and non-tumor for $K = 4$. For a higher number of clusters (Figure 2(d)), the tumor area is represented by more than one cluster. For instance, when $K = 10$, four colors can be distinguished in this area. Note that, as before, the method detects the presence of a cluster of pixels whose enhancement slope and washout variables are equal to zero.

In the case $K = 10$, it appears also that the vessels above the heart are classified as *tumor* instead of *heart*, which does not occur when $K = 3$ or $K = 4$. Also of note is that when $K = 4$ and $K = 10$, the tumor is surrounded by a thin border, composed of pixels from several clusters.

Similar analysis has been made for the other data sets. Figures 4 and 5 show the segmentations for Patients 08 and 28. They illustrate the ability of model-based clustering to produce simple segmented images that reproduce the important features contained in the full set of 25 sequential images. Note that the tumor area is not always painted red as one may wish (Figures 4(c) and 5(b)), because in the clustering method, the color assignment is arbitrary.

In what follows, we detect the suspicious regions and automatically assign them to a predetermined color (red) using the difference at peak parameter. The spatial techniques described

in Section 3.2 are then applied to further refine this initial nonspatial analysis. Going further for patient 05, using the initial MCLUST classifications (Figure 2) to apply blind restoration with $r = 1$, we obtained, for $K = 3, 4, 10$, the segmentations shown in the first column of Figure 6. For $r = 3$, the results are shown in the second column of Figure 6. Blind restoration smooths the data considerably, and, as expected, the results are even smoother with $r = 1$ than with $r = 3$. For this image it clearly eliminates extraneous minor features and retains the tumor, and so in this case the result is satisfactory. However, in general it may be helpful to use less smooth images as well.

Using Bayesian morphology, we obtained the classifications shown in Figure 7. In addition to an initial classification, the method requires an initial value for the parameter p (see Section 3.2.2) between 0 and 1 that controls the first operator applied. When p is close to 1, the resulting rank operator is a *median filter* ($r = 1$). In Figure 7, p was set initially to 0.99. For this example, the classifications with p set initially to 0.45 were similar for $K = 3$ and $K = 4$. For $K = 10$, the analysis with $p = 0.99$ was clearly better because it eliminates more noise, without removing any features of actual or potential interest.

Similar investigations were carried out for all the patients in our data set. Here are our main conclusions:

1. Model-based clustering techniques provide informative initial segmentations.
2. Partitions into 4 colors/segments were adequate to reveal the tumor of interest. Three segments were too few because the resulting partition was not sufficient to distinguish the tumor from other tissue classes. Ten segments were too many, because the resulting partition divided the tumor pixels among several classes.
3. Bayesian morphology is useful in refining these initial classifications by:
   (a) giving a simultaneous (color) picture of all the (grey-level) bands;
   (b) eliminating noise and distracting features; and
   (c) enhancing features of potential interest.
4. Moderate blind restoration (second column of Figure 6) provides much more smoothing and a clearer picture, but at the possible cost of eliminating unclear features of potential interest. Strong blind restoration (first column of Figure 6) smooths the image even further,

Table 2: Patient 05, slice 09. The classifications in the first column of Figure 6 into $K = 3, 4, 10$ classes are used to compute mean values for variable *difference at peak* in each non-background component. The largest value corresponds to the lesion of interest while taking the largest mean *time to peak* would select the heart area.

Patient 05 slice 09			
K=3	mean difference at peak	mean time to peak	interpretation
class 2	10363	**-2061**	heart
class 3	**17179**	-4578	lesion and skin
K=4			
class 2	13006	- 5468	skin
class 3	10519	**-1834**	heart
class 4	**27068**	-2819	lesion
K=10			
class 8	**29840**	-2222	lesion (main)
class 9	25512	-3063	lesion (border)
class 10	9736	**-1582**	heart

so that there is even more potential loss of useful information.

Based on these results, we recommend providing radiologists with two different color synthetic images, one to which statistical smoothing has been applied (*e.g.* Figure 7(b)), and another based on a more drastic heuristic smoothing method (*e.g.* Figure 6, first column, $K = 4$). Note that there is a solid statistical basis for Bayesian morphology, but less so for the more drastic smoothing performed by blind restoration.

5.2 ROI analysis

Considering this first classification step, the second step is to decide which of the segments is the ROI. We based our choice on the values of the difference at peak parameter we considered. As an illustration, Table 2 shows the values for the mean difference at peak and mean time to peak for the classifications shown in the first column of Figure 6. The suspicious regions, in red, are the ones selected when using a maximum mean difference at peak criterion.

The subsequent enhancement kinetics analysis can then be made on the basis of the signals observed in the selected ROI. The goal is to compute a curve representative of the lesion under study. As described in Section 4.2, we tried various ways to compute a good estimated pattern of uptake. A first natural choice is to compute the mean of all signals in the ROI. Figures 8(g)-(i) show mean curves computed using classifications in Figures 8(d)-(f). These curves are obtained

by averaging all the signals in the ROIs, where the ROI is defined as the largest connected component in the red region of Figures 8(d)-(f). Another possibility is to use only selected pixels, for instance those for which we have a good confidence of their belonging to the ROI. As an illustration, we considered pixels with the highest difference at peak values. Figure 9 shows the selected pixels and the corresponding mean curves in three cases. We also selected pixels in the ROI according to their membership probability estimates as provided by MCLUST (see Section 3.1). We kept the pixels for which the probability of belonging to the lesion class was very close to 1 (within 10^{-7}). Results are shown in Figure 10 (upper curves). More generally we used the membership probability estimates as weights to compute a weighted mean curve. This gives the results in Figure 10 (middle curves). The curves are very similar to the mean curves in Figure 8 (g)-(i) because in the lesions the conditional probabilities estimates are close to one. Selecting only a few pixels in the ROIs, those with the highest difference at peak values or with the highest conditional probabilities, provides mean curves where features (slope enhancement, washout, etc.) are more clearly marked.

The resulting curves are usually easily assigned to a curve type, I, II or III, where the types were described in Section 4.3. For patients 05, 08 and 28 the assignments are respectively II, I and III. This is consistent with the known diagnostics which correspond respectively to a carcinoma, a benign lesion and a carcinoma, and confirms that our procedures are of interest for the differentiation between malignant and benign lesions.

The same analysis was carried out for all 19 patients in our data sets. The results are summarized in Table 3. We computed some rates following the "minimum risk" strategy, *i.e.* considering doubtful lesions as malignant. We used the following parameters: **a**: number of patients diagnosed as having cancer for which our method conclusion is "cancer" (**true positive**), **b**: number of patients diagnosed as not having cancer for which our method conclusion is "cancer" (**false positive**), **c**: number of patients diagnosed as having cancer for which our method conclusion is " no cancer" (**false negative**), **d**: number of patients diagnosed as not having cancer for which our method conclusion is "no cancer" (**true negative**). The **sensitivity** was calculated as the number of true positive results divided by the number of patients having cancer (as given by the diagnostic information), $\frac{a}{a+c} = 93\%$. The **specificity** was calculated as the number of true negative results divided by the number of patients without cancer,

Table 3: Results (number of patients) of the ROI analysis for 19 patients.

Diagnosis Curve type	No cancer	Ambiguous	Cancer
I (benign)	5	1	0
II (doubtful)	0	1	4
III (malignant)	0	0	8

$\frac{d}{b+d} = 100\%$. We also computed the probability that there is actually cancer when the analysis indicates cancer (positive predictive value): $\frac{a}{a+b} = 1$ and that with a conclusion indicating "no cancer" there is effectivily no cancer (negative predictive value): $\frac{d}{d+c} = 0.83$.

These good results illustrate the gain in using more than a single enhancement measure and in combining two complementary type of analysis. The classification images provide a good analysis of the different regions in the breast. Among these regions one of them is usually clearly emerging as a potential tumor. The following signal intensity time course data analysis enables us to further identify the lesion. Note that as regards the final conclusion, a detailed analysis of the classification images is not always necessary. In most cases the images make the lesion appear very clearly and our ROI selection method selects the right region automatically.

Typically, in our experiments we observed two situations requiring more care. One data set was that of a patient with no tumor. Not surprisingly, the initial MCLUST classifications produced very fragmented segmentations, after which the spatial procedures smoothed out all features considered as noise and failed to identify homogeneous regions as candidates for a subsequent kinetics analysis. This could have been interpreted as a sign in favor of a "no tumor" conclusion. We were able to confirm this conclusion by performing the analysis on the fragmented MCLUST classification which identified a type I curve (benign).

In another case, after segmentation a region showing significant enhancement was selected. However the location (near the patient skin) and shape of the region was such that the possibility of a malignant tumor could be discarded.

When investigating larger numbers of patients, more complex situations can happen. For example, the breast cancer data could be distorted by motion of the patient. Although this is a serious issue, motion artifact was not present in significant levels in any of the images in our study. Methods for motion correction exist (*e.g.* [5]) and could be used prior to further

processing.

6 Discussion

We have proposed tools for guiding diagnosis of breast abnormalities when MRI data are available. As a diagnosis criterion, we relied on the examination of the contrast uptake characteristics of a breast lesion from both morphologic and kinetic points of view. We first focused on tools to help isolating potential lesions (regions of interest) prior to more specific analysis of the enhancement curves in the ROIs. We have applied model-based clustering followed by spatial smoothing techniques to data derived from a breast MRI with the object of producing one or more classifications useful to experts for breast cancer diagnosis. In particular, the classifications obtained after morphological filtering ($K = 4$ in Figure 6) clearly indicate the tumor in the image that we analyzed. However, this particular image may not represent a typical situation since the tumor is relatively easy to distinguish by eye. While the more conservative segmentations (*e.g.* Figure 7 (b)) lack smoothness to some extent, they may well retain features of potential interest when applied to images in which tumors are less apparent. This trade-off between smoothness and resolution needs to be assessed by further empirical research on other images. The trained human eye can often discern features that are not completely delineated, but it cannot reconstruct features that have been removed. However the ideal for radiologists would presumably be to have very definite, ideally black and white, images showing the tumor and non-tumor areas. We then presented tools for the enhancement kinetics analysis in the selected ROI. We obtained very good results as regards the correspondence between the estimated curve types and the known diagnosis.

This investigation indicates that our proposed statistical methods, which enable us to take into account more than a single enhancement measure, are quite promising for tumor identification. There is a clear gain in combining segmentation with kinetics analysis. Associating the location and shape of a lesion with its pattern of uptake proved to be useful in resolving questionable cases.

References

[1] C.K. Kuhl. MRI of breast tumors. *European Radiology*, 10:46–58, 2000.

[2] T. H. Helbich. Contrast-enhanced magnetic resonance imaging of the breast. *European Journal of Radiology*, 34:208–219, 2000.

[3] C.K. Kuhl, P. Mielcareck, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H. H. Schild. Dynamic Breast MR Imaging: Are Signal Intensity Time Course Data Useful for Differential Diagnosis of Enhancing Lesions? *Radiology*, 211:101–110, 1999.

[4] S. Mussurakis, D. Buckley, and A. Horsman. Dynamic MRI of invasive breast cancer: assessment of three Region-of-Interest analysis methods. *Journal of Computer Assisted Tomography*, 21(3):431–438, 1997.

[5] P. Hayton. *Analysis of contrast enhanced dynamic breast magnetic resonance images.* PhD thesis, University of Oxford, 1998.

[6] W. Wells, W. ad Grimson. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[7] K. Held, E. Kops, B. Krause, W. Wells, R. Kikinis, and H.-W. Müller-Gätner. Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, 1997.

[8] Y. Zhang, M. Brady, and Smith S. Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[9] A. Jianhua Li. Streaming Data Analysis Algorithms and Implementations. Technical report, TAMI RIL report, 1999.

[10] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[11] R. Kass and A. Raftery. Bayes factors. *Journal of the American Statisticial Association*, 90:733–795, 1995.

[12] M. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrox and F. Caski, editors, *Second International Symposium on Information Theory*, page 267, 1973.

[13] J. Rissanen. Stochastic complexity in statistical inquiry. *Word Scientific, Teaneck, New Jersey*, 1989.

[14] P. Zhang. Model selection via multifold cross validation. *The Annals of Statistics*, 21:299–313, 1993.

[15] J. D. Banfield and A. E. Raftery. Model-based Gaussian and Non-Gaussian Clustering. *Biometrics*, 49:803–821, 1993.

[16] C. Fraley and A.E. Raftery. How many clusters? Which clustering methods? - Answers via model-based cluster analysis. *The Computer Journal*, 41:578–588, 1998.

[17] C. Fraley and A.E. Raftery. MCLUST: Software for model-based cluster analysis. *Journal of Classification*, 16:297–306, 1999.

[18] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistican Association*, 97:611–631, 2002.

[19] C. Fraley and A. E. Raftery. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. *Journal of Classification*, 20:263–286, 2003.

[20] F. Forbes and A. E. Raftery. Bayesian morphology: Fast unsupervised Bayesian image analysis. *Journal of the American Statistical Association*, 94:555–568, 1999.

[21] A. P. Dempster, N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, series B*, 39:1–38, 1977.

[22] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, 1997.

[23] C. Fraley. Algorithms for model-based gaussian hierarchical clustering. *SIAM Journal of Scientific Computing*, 20(1):270–281, 1998.

[24] G. Celeux and G. Govaert. Gaussian Parsimonious Clustering Models. *Pattern Recognition*, 28:781–793, 1995.

[25] A. Dasgupta and A. E. Raftery. Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93:294–302, 1998.

[26] H.P. Friedman and J. Rubin. On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62:1159–1178, 1967.

[27] A.J. Scott and M.J. Symons. Clustering methods based on likelihood ratio criteria. *Biometrics*, 27:387–397, 1971.

[28] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.

[29] H. J. A. M. Heijmans. Mathematical morphology: a modern approach in image processing based on algebra and geometry. *SIAM Review*, 37:1–36, 1995.

[30] H. J. A. M. Heijmans. *Morphological Image Operators*. Academic Press: Boston, 1994.

[31] J. E. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, series B*, 48:259–302, 1986.

Figure 1: Patient 05, slice 09: dynamic MR images at (a) 10 seconds, (b) 70 seconds, (c) 150 seconds, (d) 250 seconds, (e) signal intensity curve at a given pixel, twenty-five measures were acquired, one measure every 10 seconds.

Figure 2: MCLUST classifications for Patient 05, slice 09. (a) reference image, (b) three-class segmentation, (c) four-class segmentation, (d) ten-class segmentation.

Figure 3: Histograms for the five parameters in the different classes of Figure 2(c)

Figure 4: MCLUST classifications for Patient 08, slice 06. (a) reference image, (b) three-class segmentation, (c) four-class segmentation, (d) ten-class segmentation.

Figure 5: MCLUST classifications for Patient 28, slice 10. (a) reference image, (b) three-class segmentation, (c) four-class segmentation, (d) ten-class segmentation.

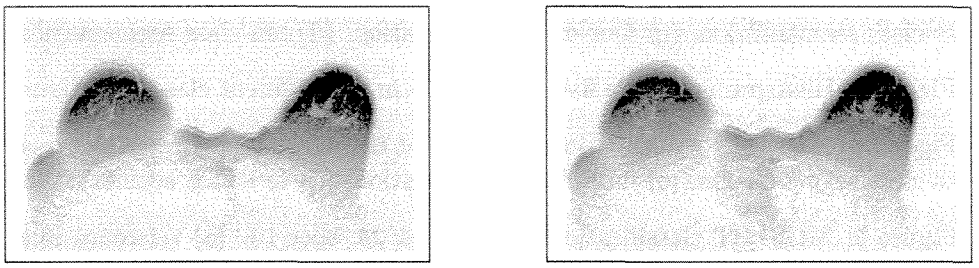
Figure 6: Blind restorations with $r = 1$ (first column) and $r = 3$ (second column) using images in Figure 2 as initial classifications, for K= 3, K=4 and K= 10.

Figure 7: Bayesian morphology using images in Figure 2 as initial classifications. (a) K= 3, (b) K=4, (c) K= 10.

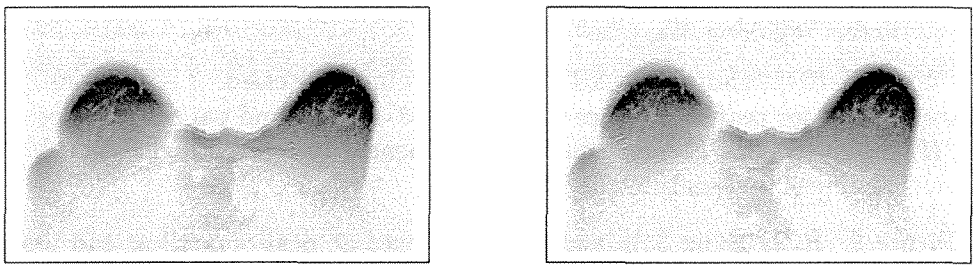
Figure 8: ROIs and associated mean curves in three cases. (a), (b), (c) dynamic image at $t = 1$ for patient 05, slice 09, patient 08, slice 06 and patient 28, slice 10. (d), (e), (f) ROI selections (largest connected component in the red regions). (g), (h), (i) mean time-intensity signals in the ROIs.

Figure 9: ROIs using 33% and 67% difference at peak quantiles, and associated mean curves in three cases. (a), (b), (c) zoomed ROIs from MCLUST classifications with $K = 4$, for patient 05, slice 09, patient 08, slice 06 and patient 28, slice 10. (d), (e), (f) ROI segmentations using difference at peak quantiles: highest 33% values in red, lowest 33% values in green. (g), (h), (i) mean time-intensity signals in each class (upper curve for the red class, lower curve for the green class).

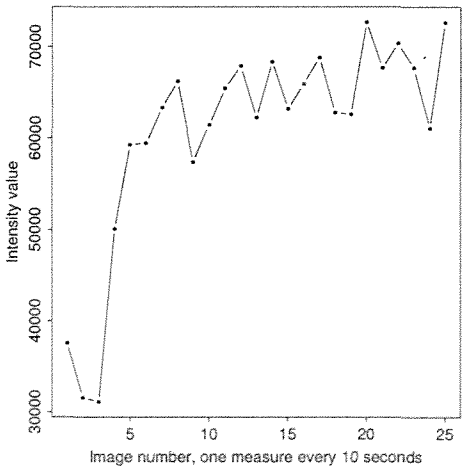
Figure 10: Mean curves using conditional probabilities estimates from MCLUST. (a) patient 05, slice 09, (b) patient 08, slice 06 (c) patient 28, slice 10. Upper curves: mean curves when selecting pixels with conditional probability very close to 1 (within 10^{-7}). Middle curves: weighted mean curves when using conditional probabilities as weights. Lower curves: mean curves using all pixels in the ROIs.



(a) (b)

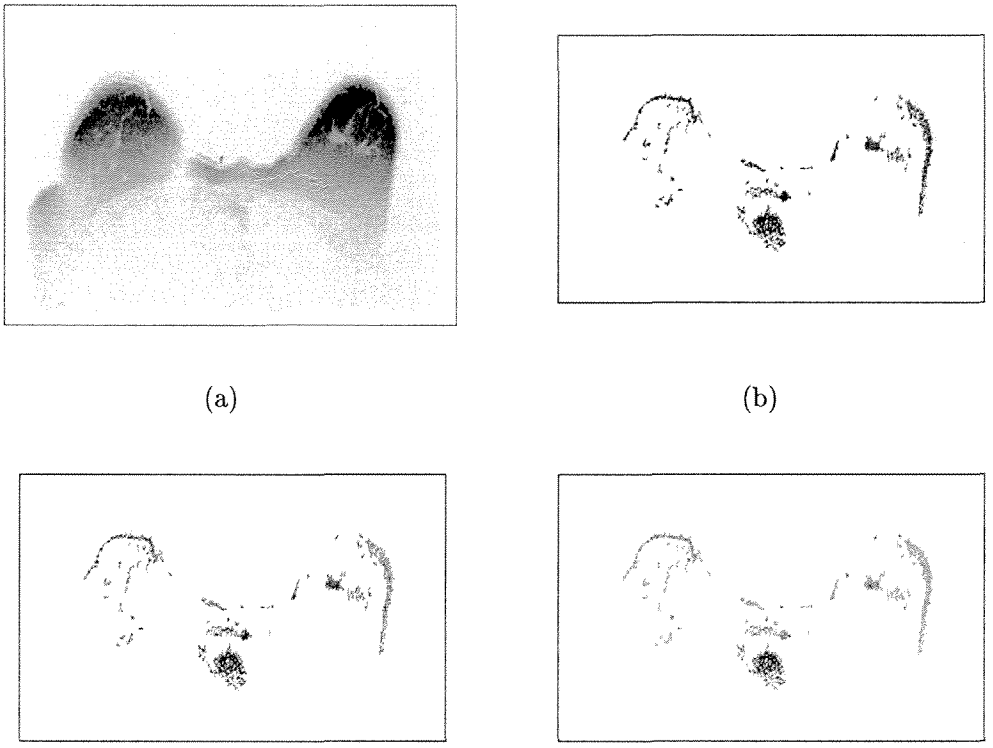


(c) (d)



(e)

Figure 1:

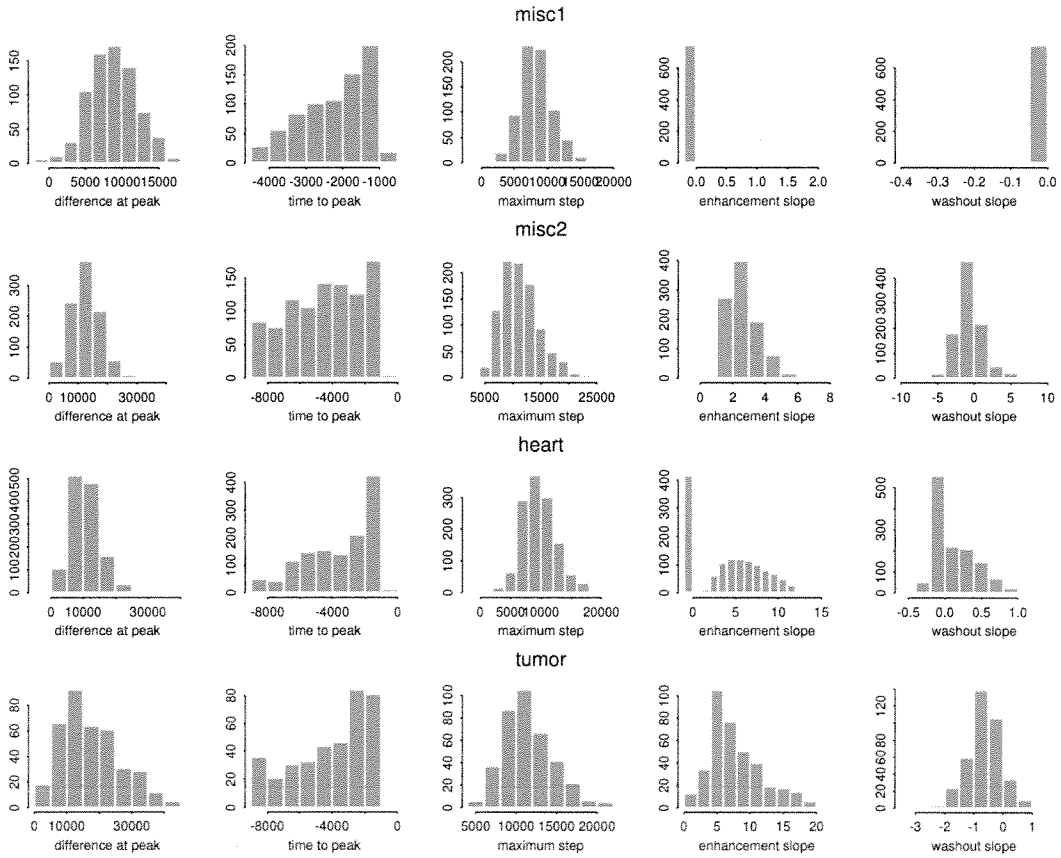


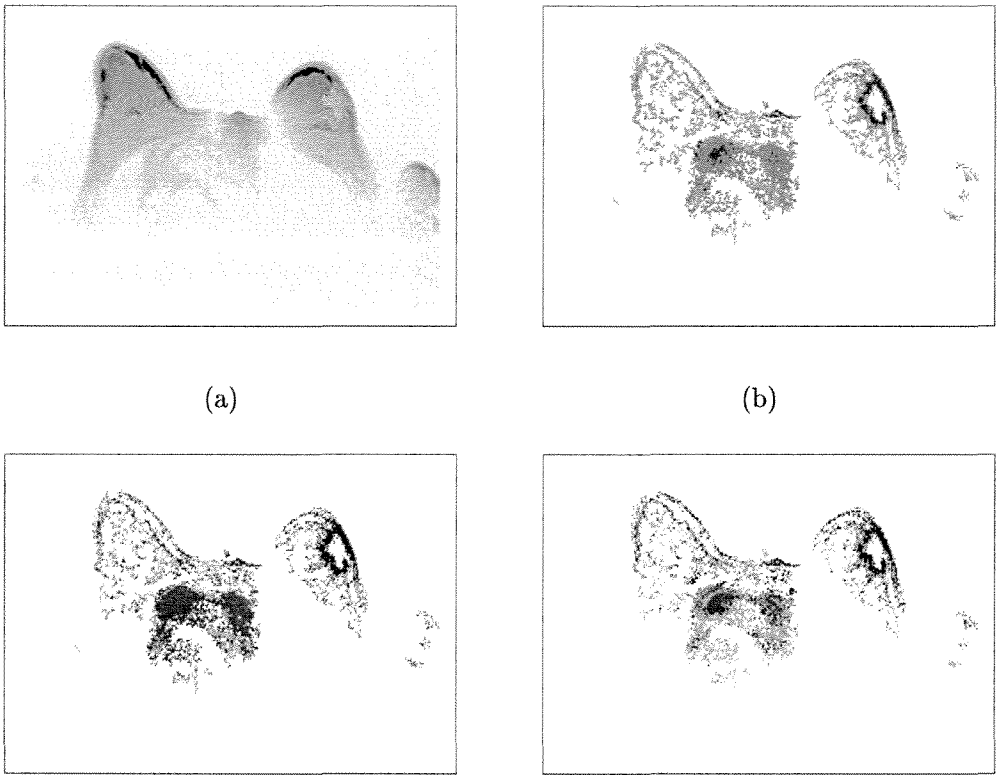
(a) (b)

(c) (d)

Figure 2:

Figure 3:

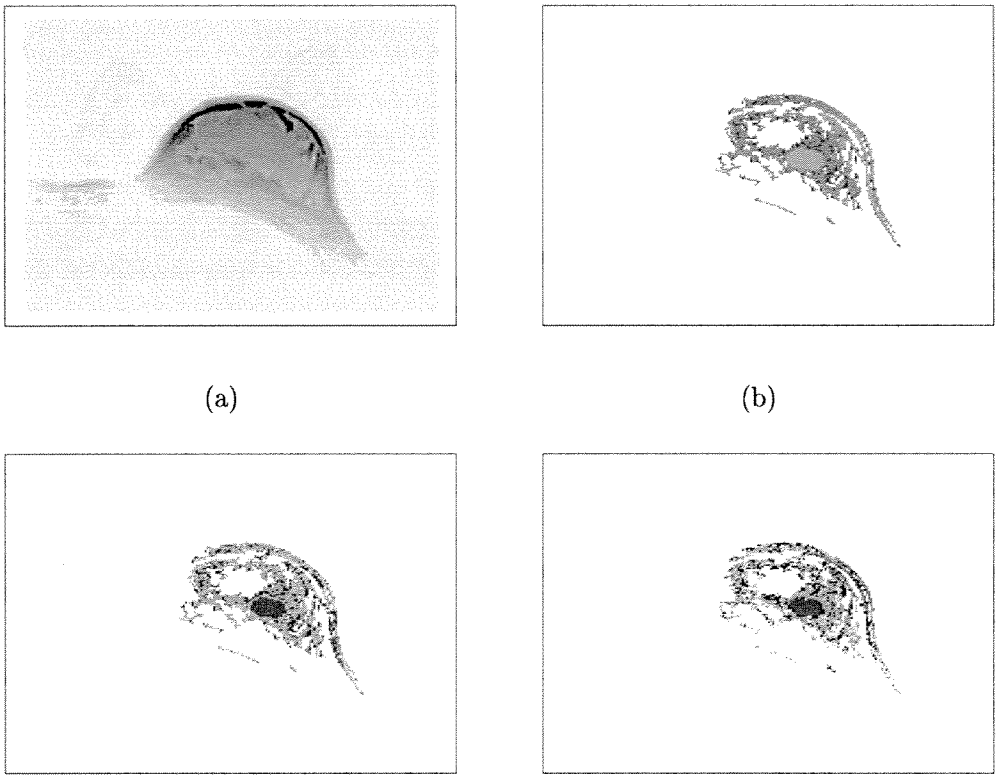

misc1
difference at peak
time to peak
maximum step
enhancement slope
washout slope
misc2
difference at peak
time to peak
maximum step
enhancement slope
washout slope
heart
difference at peak
time to peak
maximum step
enhancement slope
washout slope
tumor
difference at peak
time to peak
maximum step
enhancement slope
washout slope



(a) (b)

(c) (d)

Figure 4:



(a) (b)

(c) (d)

Figure 5:

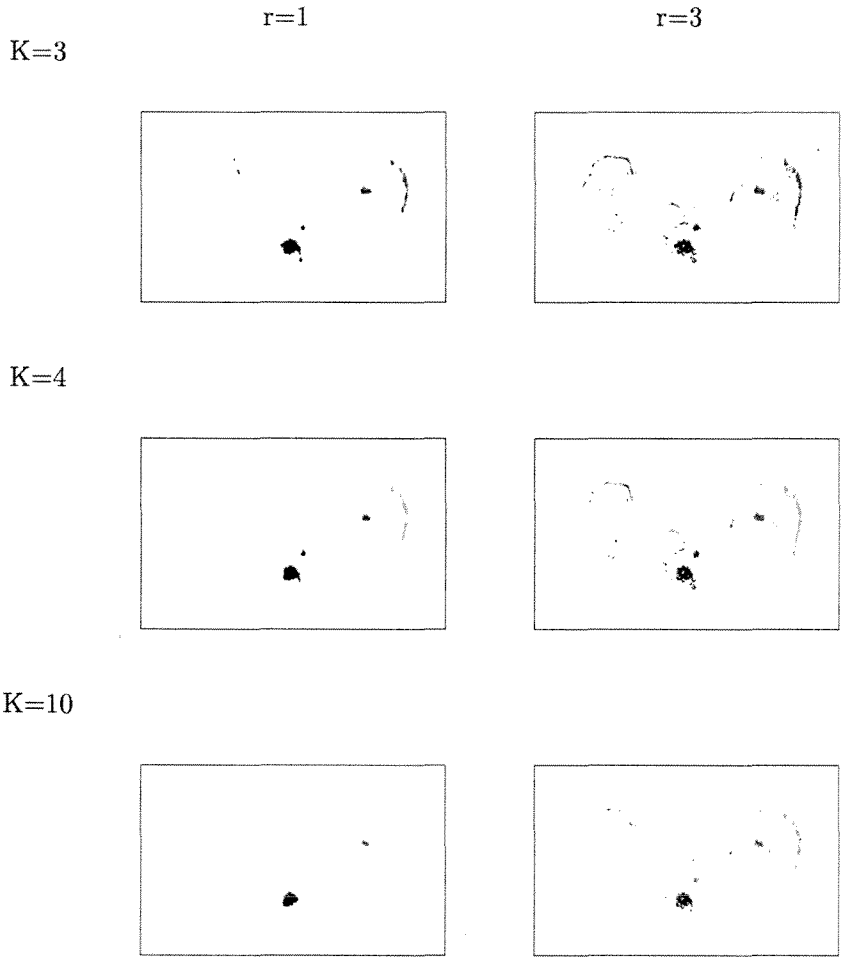
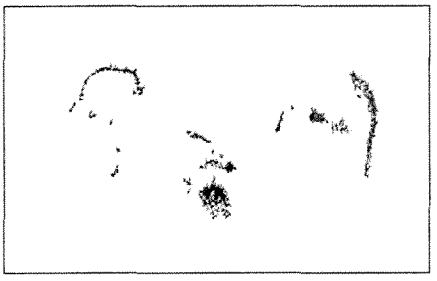
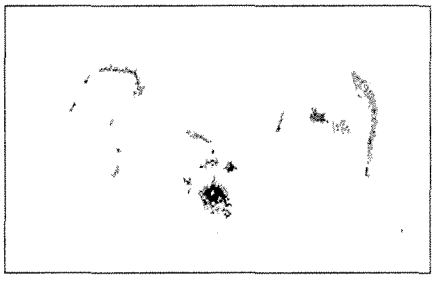


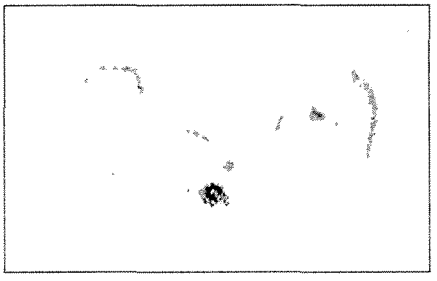
Figure 6:



(a)

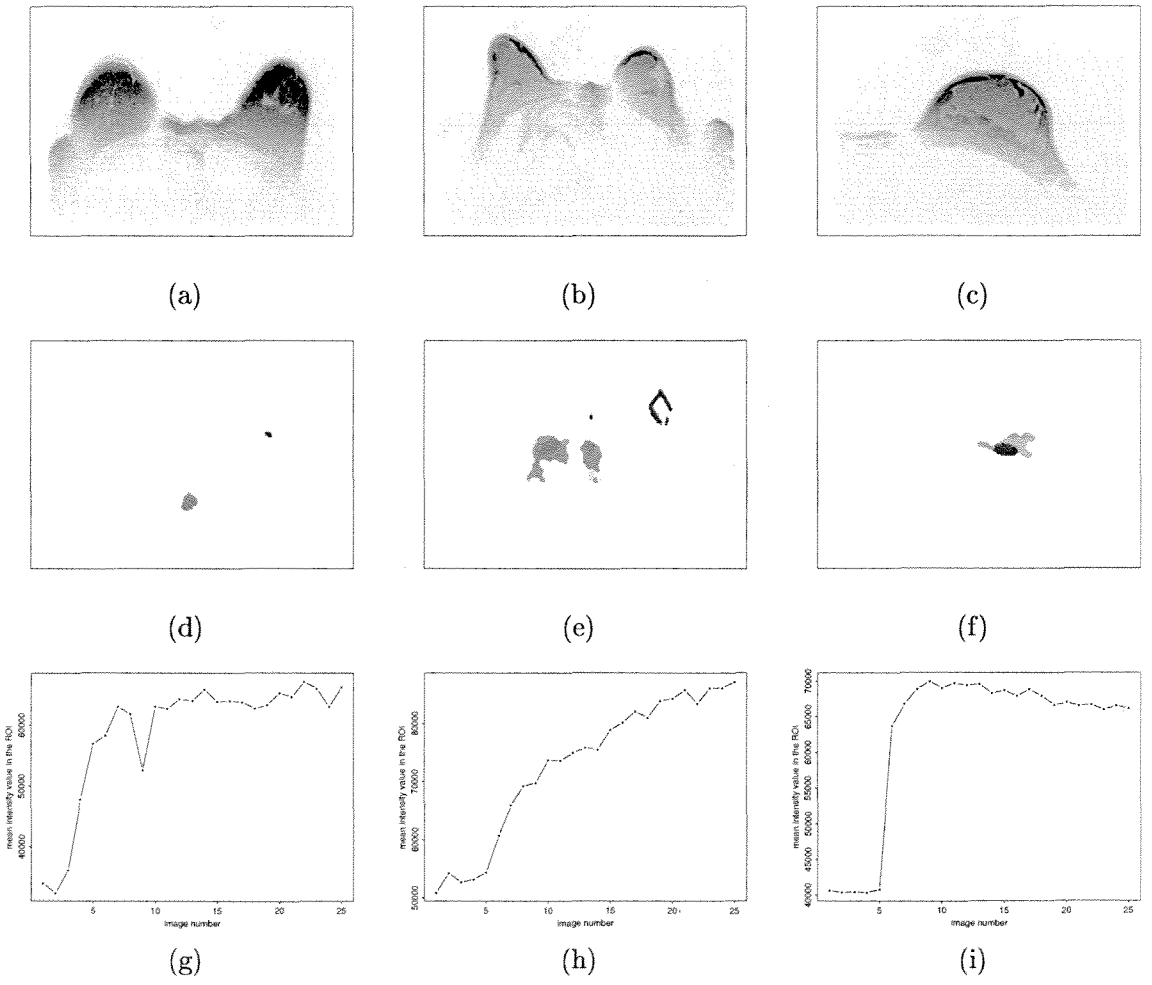


(b)



(c)

Figure 7:



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8:

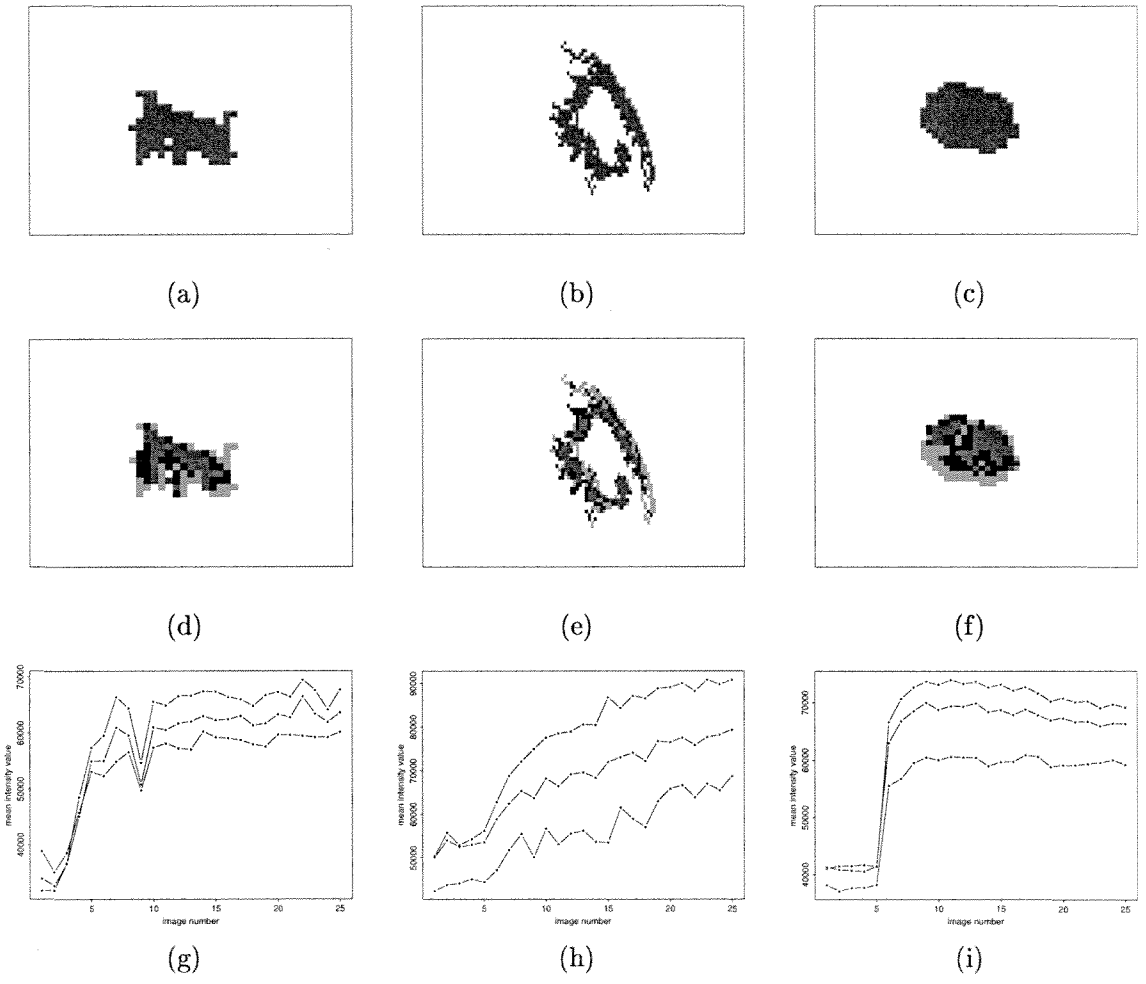
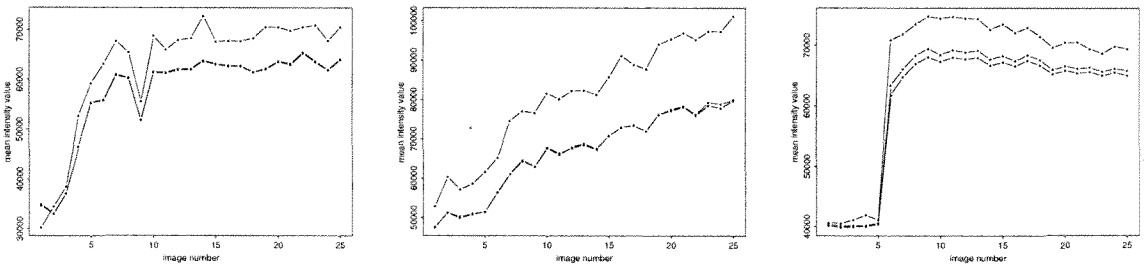


Figure 9:



(a) (b) (c)

Figure 10: