Original Research

Automated Analysis of Contrast Enhancement in Breast MRI Lesions Using Mean Shift Clustering for ROI Selection

Mark J. Stoutjesdijk, MD, MSc,[1,2]* Jeroen Veltman, MD,[1] Henkjan Huisman, PhD,[1] Nico Karssemeijer, PhD,[1] Jelle O. Barentsz, MD, PhD,[1] Johan G. Blickman, MD, PhD,[1] and Carla Boetes, MD, PhD[1]

Purpose: To evaluate a new method for automated determination of a region of interest (ROI) for the analysis of contrast enhancement in breast MRI.

Materials and Methods: Mean shift multidimensional clustering (MS-MDC) was employed to divide 92 lesions into several spatially contiguous clusters each, based on multiple enhancement parameters. The ROIs were defined as the clusters with the highest probability of malignancy. The performance of enhancement analysis within these ROIs was estimated using the area under the receiver operator characteristic curve (AUC), and compared against a radiologist's final assessment and a classifier using histogram analysis (HA). For HA, the first, second, and third quartiles were evaluated.

Results: MS-MDC resulted in AUC = 0.88 with a 95% confidence interval (CI) of 0.81–0.95. The AUC for the radiologist's assessment was 0.93 (95%CI = 0.87–0.97). Best HA performance was found using the first quartile, with AUC = 0.79 (95%CI = 0.69–0.88). There was no significant difference between MS-MDC and the radiologist ($P = 0.40$). The improvement of MS-MDC over HA was significant ($P = 0.018$).

Conclusion: Mean shift clustering followed by automated selection of the most suspicious cluster resulted in accurate ROIs in breast MRI lesions.

Key Words: breast; MRI; CAD; mean shift; clustering
J. Magn. Reson. Imaging 2007;26:606–614.

DYNAMIC CONTRAST ENHANCED MRI (DCEMRI) of the breast is a sensitive technique to detect breast lesions, but its specificity is still limited. DCEMRI makes it possible to evaluate the architectural features of a breast lesion in several orientations, and also enables the radiologist to analyze the dynamic contrast enhancement characteristics of the lesion. Both aspects of breast MRI lesions are important to obtain optimal diagnostic accuracy (1). For the evaluation of contrast enhancement kinetics, it is essential to find a representative area within the tumor. A lesion averaged approach has been shown to be inferior to manually selecting a smaller region of interest (ROI) (2,3), although there is no consensus on the optimal location and size of this ROI. A generally accepted way to determine a good ROI is to take a reasonably sized section of the lesion that captures the most malignant aspects of the lesion, usually defined by the fastest and most intense contrast enhancement pattern (4,5). Unfortunately, there is no good definition of either "reasonably sized" or "most malignant contrast dynamics." This leads to substantial variability in the assessment of contrast enhancement dynamics (6–8), an issue that will become even more important when contrast enhancement analysis includes the evaluation of several parameter maps. The most suspicious area in a particular parameter image may not be the most suspicious in another parameter image.

It seems, therefore, that both the current way of evaluating contrast enhancement, as well as anticipated future methods (including parameter map analysis), would benefit from a standardized method of determining the ROI that incorporates all available parameters. As far as we know, there is currently no published method for automated identification of a spatially contiguous ROI. We propose mean shift segmentation as an automated method to divide a breast MRI lesion into several spatially contiguous clusters of similar enhancement characteristics. With a new, iterative classification technique, it is possible to identify the most suspicious cluster, which then represents the lesion's ROI. Such an ROI could be displayed on a work station, perhaps together with its enhancement/time curve, or it could be used in a computer aided diagnosis (CAD) setting in the automated evaluation of contrast enhancement characteristics.

[1]Radboud University Medical Centre, Department of Radiology, Nijmegen, The Netherlands.
[2]Ikazia Hospital, Department of Radiology, Rotterdam, The Netherlands.
*Address reprint requests to: M.S., Radboud University Medical Centre, Department of Radiology, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. E-mail: m.stoutjesdijk@rad.umcn.nl
Received June 4, 2006; Accepted May 3, 2007.
DOI 10.1002/jmri.21026
Published online in Wiley InterScience (www.interscience.wiley.com).

Mean shift clustering is well suited for segmentation tasks where spatial contiguity is required, such as the segmentation of objects in color images (9). Clustering based on mean shift does not require training. Further, the number of resulting clusters is not set a priori, unlike other segmentation techniques such as k-means and fuzzy c-means. Iterative self-organizing data analysis (ISODATA) does allow adjustment of the number of clusters, but there are about six parameters that need to be set.

To find the most suspicious cluster, a classifier is needed. The classifier needs training with examples of the most suspicious clusters, but these are yet unknown. It is, therefore, not straightforward to automatically determine the most suspicious cluster in a lesion. We describe a new iterative training algorithm that addresses this challenge.

The goal of this study was to determine whether CAD based on the enhancement characteristics within the ROIs found by our method has a diagnostic performance similar to an experienced radiologist's assessment, and also to compare it to an alternative automated method to find the most suspicious voxels in a lesion.

MATERIALS AND METHODS

Data Acquisition

Patient Population and MRI Scanning Protocol

The MRI data for this study was obtained retrospectively from 481 consecutive MRI studies, performed from October 2003 to July 2005. From these 481 MRI examinations, the data set for this study was obtained using the following inclusion criteria: first, an enhancing lesion had to be present; second, histological correlation or follow-up by MRI with an interval of at least two years was available; and third, there were no technical problems or patient mobility artifacts during the examination reducing the quality of the data or the number of obtained series. No lesion was included twice, but different lesions in the same patient were allowed. Using these criteria, 92 lesions from 92 studies performed on 87 women were included.

Each study was performed using the same protocol, in the second week of the menstrual cycle, to minimize contrast enhancement of glandular tissue (10). The protocol consisted of one precontrast T1-weighted fast low-angle shot three-dimensional (FLASH-3D) series, with echo time (TE) = 4 msec, repetition time (TR) = 7.8 msec, 120 slices, and flip angle $\alpha = 20°$. The field of view (FOV) = 340 mm, voxel size = 1.3 × 1.3 × 1.3 mm, and there was no interslice gap. To avoid artifacts from heartbeat, scanning was done in the coronal plane. After this initial series, 0.2 mmol/kg body weight of gadopentetate dimeglumine contrast agent (Dotarem, Guerbet; The Netherlands) was injected using a power injector (Powerinjector®; MedRad, Pittsburgh, PA, USA) followed by a saline flush. At 90 seconds after contrast injection, four consecutive postcontrast measurements were obtained using the same protocol. No T2-weighted series were acquired.

Feature Definitions

In our experiments, we used two commonly used enhancement parameters as features, namely, relative enhancement and washout. The presence of strong initial enhancement and fast washout are considered suggestive for malignancy (4). Both parameters were obtained from the dynamic three-dimensional (3D) volumes, acquired as described in the previous section. We define V as the collection of all voxels from all lesions in our data set. Relative enhancement $RE(v,t)$ in voxel $v \in V$ at time point t is defined as follows:

$$RE(v,t) \equiv \frac{SI(v,t) - SI(v,t_0)}{SI(v,t_0)} \times 100\%, \qquad (1)$$

where $SI(v,t)$ denotes the signal intensity of voxel v at time point t, and $SI(v,t_0)$ the signal intensity at $t = t_0$, before administration of contrast. We used $RE(v,t_1)$ as feature value for relative enhancement, with t_1 the relative enhancement at the first postcontrast time point $t = t_1$. The feature was then normalized as follows:

$$RE_n(v,t_1) = \frac{RE(v,t_1) - \mu_{RE(t_1)}}{\sigma_{RE(t_1)}}, \qquad (2)$$

where $\mu_{RE(t1)}$ and $\sigma_{RE(t1)}$ are the mean and standard deviation of $RE(v,t_1)$ over all $v \in V$, respectively. Furthermore, washout in a given voxel v is defined by the slope of the straight line that best fits all postcontrast relative enhancement values. This slope is determined using linear regression (11), i.e., by minimizing the chi-squared merit function $\chi^2(a,b)$ of the fit to the line $RE(v,t) = a + bt$ through N postcontrast time points $t_j = t_1 \ldots t_N$:

$$\chi^2(a,b) = \sum_{j=1}^{N} \left(\frac{RE(t_j) - a - bt_j}{\sigma_j} \right)^2, \qquad (3)$$

where σ_j is defined as the standard deviation of $RE(t_j)$. The washout of voxel v is then defined as $Wash(v) = -b$. The normalized feature $Wash_n(v)$ is then obtained similar to the normalized relative enhancement.

Lesion Delineation

An operator first identified a voxel within the tumor. The lesion was then delineated using a standard connected threshold operation on the relative enhancement feature (using $t = t_1$). This resulted in a contiguous region, including the identified voxel, with a relative enhancement above a certain value. An empirically determined default threshold of 120% enhancement was employed. If a lesion showed only faint enhancement, this level could be too low for a complete delineation of the tumor, and if a lesion was adjacent to another strongly enhancing structure (e.g., a blood vessel), the use of the standard level could result in inclusion of this structure. Therefore, the operator could adjust the threshold level if the delineation results were not satisfactory.

Mean Shift Clustering

Mean shift clustering is a nonparametric, multivariate method used to find areas of similar enhancement behavior (clusters) within a tumor. A summary of the mean shift algorithm as it applies to our method is given in Appendix A; a more generalized and formal discussion can be found in Fukunaga and Hostetler (12), Cheng (13), and Comaniciu et al (9).

The end result of the clustering step is that each voxel of the lesion has been assigned to one cluster. Each cluster consists of spatially contiguous voxels with similar enhancement parameters. The degree of parameter value (range) variation found within a cluster depends on the choice of the spatial and range bandwidth parameters h_s and h_r. When the spatial bandwidth increases, only larger regions with similar parameters will be assigned to a separate cluster. When the range bandwidth increases, only regions with parameter values that are substantially different from their surroundings will be assigned to a separate cluster. Figure 1 shows the effect of changing the bandwidths on the resulting number of clusters.

To minimize the effects of noise and possible patient motion, the minimum ROI volume was arbitrarily set at 10 mm^3, and clusters with a total volume of less than 10 mm^3 were discarded if at least one larger cluster was available.

In this work, the mean shift procedure was implemented using code from the edge detection and image segmentation (EDISON) package, kindly provided by the Robust Image Understanding Laboratory at Rutgers University (http://www.caip.rutgers.edu/riul/research/code.html; Last accessed June 23, 2007.).

Cluster Selection

The next step is to find the cluster that is most representative of the tumor and to discard the others as nondescriptive. In this section, we present a method to identify the cluster that has the highest degree of suspicion. This is done using a trained classifier. Training in this case is complicated, because each lesion is likely to contain multiple clusters, from which the cluster with the highest probability of malignancy should be selected. Yet the classifier still needs training, so there is no certain way to find these most suspicious clusters. This issue was addressed by determining an initial classifier using all clusters from all lesions. This initial classifier was expected to perform rather poorly, but it would allow for reclassification of the data set, producing a slightly improved classifier. From there, an iterative process of adjustments to the training data set was executed until a repetitive sequence was reached. The final classifier was determined from the intersection of clusters that were used during the repetitive sequence. In other words, the training set only contained lesions that showed a stable cluster of highest suspicion. The algorithm was thus:

1. Train a classifier C_0 with all clusters from all lesions, using jackknife leave-1-out cross-validation.

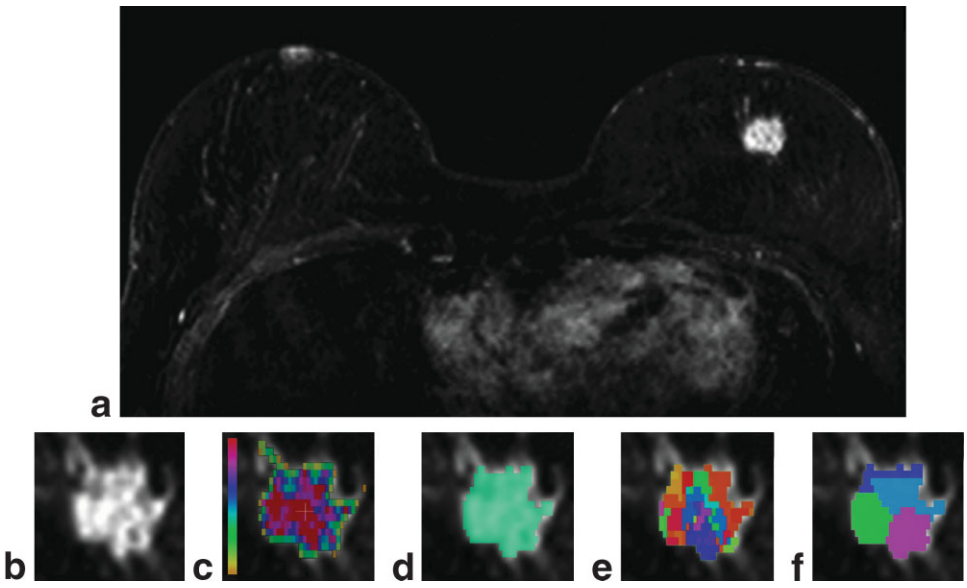

Figure 1. a: Large enhancing mass in the left breast. This image was obtained by subtracting the precontrast series from the first postcontrast series. **b:** Detail of the same lesion, at a slightly different level. **c:** Color overlay representing the relative contrast enhancement. The color scale left of the lesion indicates the level of enhancement, from very low (yellow) to very high (red). Voxels without color did not show enough enhancement. **d:** Delineation of the lesion. Mean shift clustering will be performed on the colored voxels. The threshold for delineation does not necessarily correspond to the lowest enhancement level shown on the color overlay in (c), as can be seen in the top-left region. **e:** Clustering result with small bandwidth, $h_r = h_s = 1.0$. Each cluster is identified by a different color, unrelated to the scale in (c). Some uncolored voxels at the top of the lesion belong to a cluster smaller than 10 mm^3 which was discarded. Note that the layout of the clusters does not at all locations correspond with the color map in (c), because clustering was done based on both washout and relative enhancement. **f:** Clustering result with larger bandwidth, $h_r = h_s = 2.0$. The clusters are now considerably larger and all delineated voxels are assigned to a cluster.

2. Using C_0, determine for each lesion the cluster with the highest posterior probability of malignancy and use the resulting collection to train a new classifier C_1.
3. Iteratively determine C_{j+1} from the clusters with the highest posterior probability of malignancy as determined using C_j. Stop when $C_{j+1} \in C_{1\ldots j}$ (repetitive sequence of R classifiers; $C_{j+1} = C_{j+1-R}$).
4. Determine the final classifier C_F from the intersection of the clusters that were used for the repetitive sequence.

The final classifier performance was estimated using the area under the receiver operator characteristic (ROC) curve (AUC). The classification was done using quadratic discriminant analysis. Note that after the initial first step, the determination of C_F is unsupervised and that there is no optimization of the AUC. Figure 2 shows an example of the variation of the AUC during the execution of the algorithm.

The results of the multidimensional clustering (MDC) procedure were expected to depend somewhat on the selection of the bandwidth parameters h_s and h_r. This current work does not offer automated bandwidth determination and bandwidths of $h_s = 1.0$ mm and $h_r = 1.0$ (unitless with normalized features, see "Feature Definitions") were used as an initial estimate. To evaluate whether these values were reasonable, the experiments were repeated with all combinations of h_s and h_r values from {0.6, 0.8, 1.0, 1.5, 2.0}.

Performance Comparison

To assess accuracy of the ROIs determined using the mean shift clustering algorithm, we performed automated classification of the 92 lesions in our data set using these ROIs. We compared the results to those from an alternative automated classification method, as well as to the final assessment of an experienced radiologist. The performance analysis of both automated methods was performed using a leave-1-out approach.

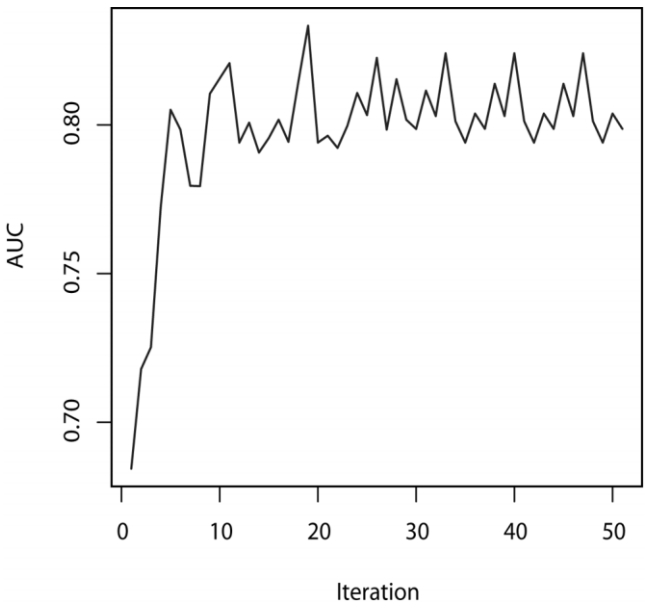

Figure 2. Example of the AUC vs. the number of iterations in the determination of the classifier. Note the repetitive sequence after iteration 24. The exact behavior of AUC vs. the number of iterations varies with the values of the bandwidths h_s and h_r.

With mean shift multidimensional clustering (MS-MDC)-based automated classification, each lesion was divided into several clusters using the mean shift procedure. A standard classifier using quadratic discriminant analysis was then employed to determine a probability of malignancy based on the average value of the normalized *Wash* and *RE* enhancement parameters within each cluster. Training of the classifier was done as described above. The ROI of each lesion was defined as the cluster with the highest probability of malignancy.

The alternative automated method consisted of classification based on quartiles calculated for each feature (*Wash* and *RE*) at all voxels within the same delineated lesions (histogram analysis) (14,15). This was implemented by first determining two histograms (*Wash* and *RE*) from all voxels, within a given delineated lesion, for both *Wash* and *RE*. The first quartile was then obtained from each histogram; i.e., 25% of the voxels are below this level. These values were determined for the entire data set and the same type of classifier as used for MS-MDC was employed to obtain an AUC. This process was repeated for the second and third quartiles, yielding a total of three AUC values. The best performing quartile was then used to compare against the MS-MDC performance.

The radiologist's final assessment of the lesion was done using the breast imaging reporting and data system (BIRADS) classification for MRI (5), which grades suspicious lesions from 1 to 6, with increasing levels of suspicion. A BIRADS score of 1 means that the MRI study is normal, so this classification, by definition, was not used in our study. Similarly, a BIRADS score of 6 indicates histology proven malignancy. No clinical information was given to the radiologist, so this classification was also not used. The remaining scores vary from "almost certainly benign" (BIRADS 2) to "highly suspicious for malignancy" (BIRADS 5).

The comparison of the various results was done by using the AUC as an estimate of diagnostic accuracy. The 95% confidence intervals of the various AUC scores were determined with bootstrapping (16,17) with 5000 samples, and paired testing was used to evaluate differences. A statistically significant difference was defined as having $P < 0.05$.

RESULTS

Breast MRI Data

The data set consisted of 92 lesions (from 92 studies performed on 87 women) with either histology or follow-up correlation. The distribution of the histological findings is given in Table 1. Note that there are several benign cases that typically exhibit suspicious contrast enhancement behavior (e.g., adenosis), as well as some malignant lesions that are known to sometimes show

Table 1
Distribution of the Lesions in the Data Set

Benign	*N*	Malignant	*N*
Fibroadenoma	11	Invasive ductal carcinoma	36
Fibrosis/scar tissue	7	Invasive lobular carcinoma	5
Adenosis	3	Ductal carcinoma in situ	13
Ductal papilloma	2	Adenocarcinoma NOS	5
Chronic Inflammation	1		
Radial scar	1		
Follow-up	8		
Total	33	Total	59

NOS = not otherwise specified.

relatively benign curves, such as invasive lobular carcinoma. The size of the lesions, as determined by an expert reader on MRI, ranged from 0.5 to 9.0 cm (mean = 2.6 cm).

The lesions were identified by one radiologist with extensive experience in breast MRI. They were then automatically delineated based on their relative enhancement with a threshold of 120%. The operator adjusted this threshold 33 times out of 92 (36%). After the adjustments, the semiautomated delineation obtained an acceptable result in all cases, as confirmed by the breast MRI expert.

CAD Performance

The clustering algorithm resulted in an AUC = 0.88 with a 95% confidence interval (95%CI) = 0.81–0.95. Histogram analysis showed the best performance when using the first quartile, with an AUC = 0.79. Table 2 shows the AUCs and 95%CIs for all quartiles. The radiologist's final assessment, as shown in Table 3, resulted in AUC = 0.93 with a 95%CI = 0.87–0.97. The improvement of mean shift clustering over histogram analysis was statistically significant ($P = 0.018$ for the first quartile); the difference between the radiologist's performance and that of mean shift clustering was not statistically significant ($P = 0.40$). The ROC curves are shown in Fig. 3. The scatter plots in Fig. 4 show the feature values within the ROIs and the values as determined from the first quartile.

The CAD performance was found to be acceptable at range and spatial bandwidths $h_r = h_s = 1.0$. The mean cluster volume using these bandwidths was 33 mm^3 (range = 13–8539 mm^3). Performance tended to diminish to less than 0.70 at higher values (e.g., either h_r or h_s at 2.0). Very small bandwidths, resulting in smaller clusters, caused somewhat lower performance as well. Note that extremely low bandwidths could lead to oversegmentation, and, thus, to failure of the clustering algorithm due to the minimum cluster volume requirement (see Mean Shift Clustering). The effect of bandwidth variation on the clustering result is illustrated in Fig. 1.

Table 2
Results of Classification Using Features Obtained From Histogram Analysis

Quartile	AUC	95%CI
First	0.79	0.69–0.88
Second	0.74	0.62–0.83
Third	0.70	0.57–0.80

Table 3
Final BIRADS Assessments of the Radiologist

BIRADS	Benign	Malignant
2	21	4
3	4	2
4	6	36
5	2	17

DISCUSSION

In this work, we present an accurate method to automatically determine a contiguous region of interest for the evaluation of breast MRI contrast enhancement characteristics. This is accomplished in two steps. First, mean shift segmentation is used to divide a lesion into several clusters. The number of clusters is not known a priori; this depends on the size and heterogeneity of the lesion. Second, the cluster with the most suspicious enhancement characteristics is identified. This will represent the ROI. It is determined using a classifier that was trained using a new iterative method.

Classification results using the enhancement characteristics from the automatically determined ROI were significantly better than those based on histogram analysis. After the classifier was trained, new lesions could be evaluated in only a few seconds. This is much faster than a manual search for the most suspicious enhancement, which may take up to several minutes. Also, this technique finds clusters in three-dimensional

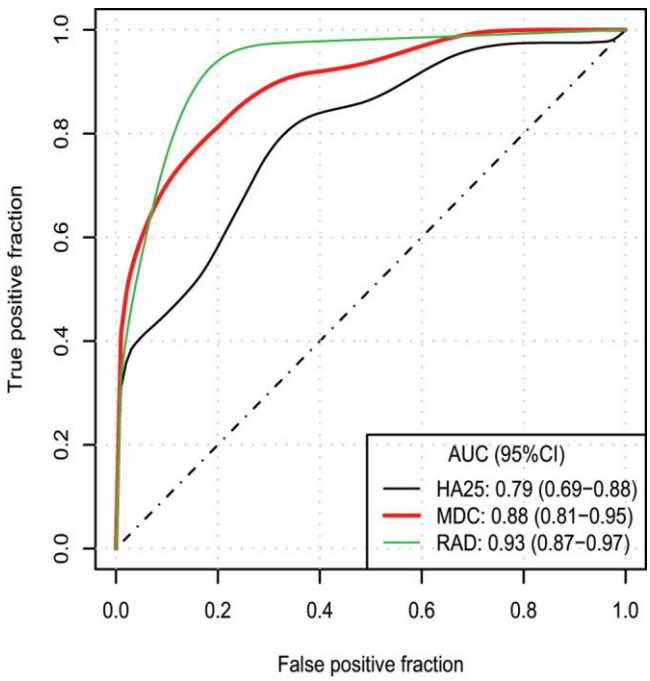

Figure 3. ROC curves for the MS-MDC algorithm, histogram analysis for the best performing quartile (first quartile, HA25), and an experienced radiologist's BIRADS assessment (RAD).

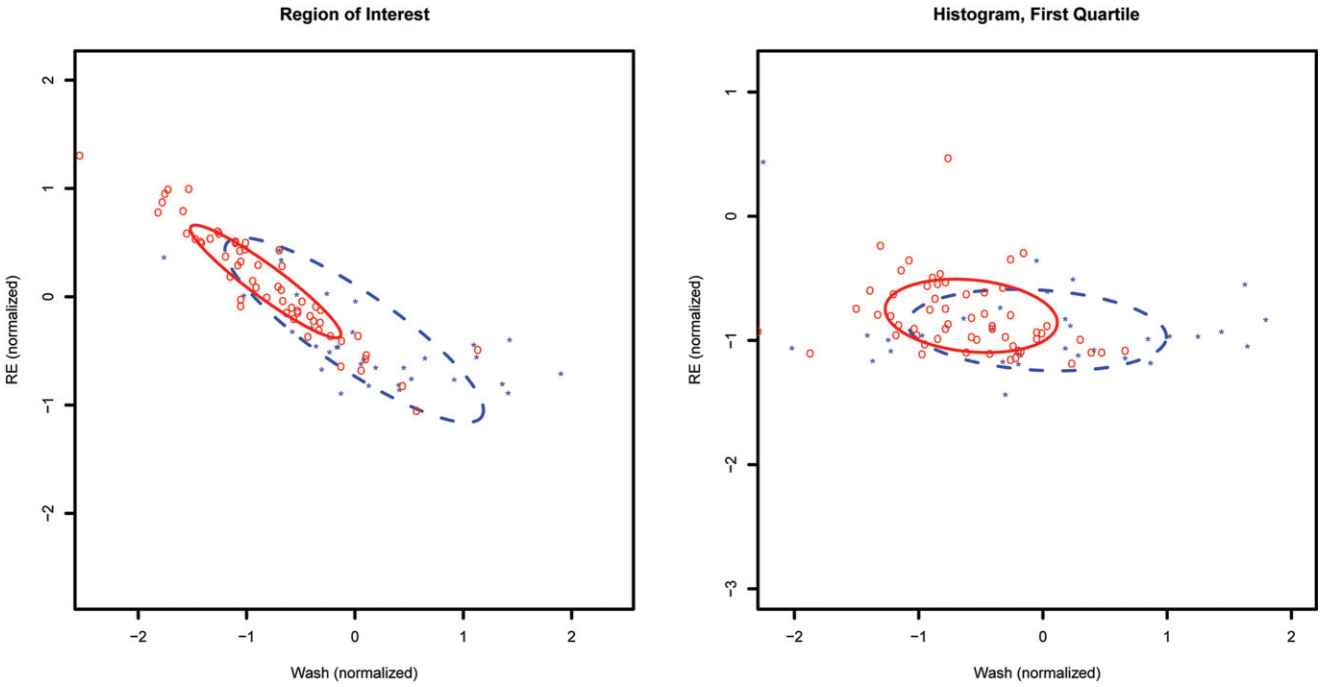

Figure 4. Scatter plots indicate the representative values of the *Wash* and *RE* features for each lesion as determined by the ROI method and the histogram method (first quartile). Malignant lesions are shown as red circles and benign lesions as blue asterisks. The ellipsoids summarize the groups of malignant and benign lesions by fitting a bivariate normal distribution and displaying the outline at a radius of 1SD.

space. This is an advantage compared to the method offered by typical high-end workstations, namely a free-hand region drawn in one slice only. Automated classification of the lesions in our data set resulted in a diagnostic performance that was comparable to that of an experienced radiologist. It should be noted that to reach a final assessment, the radiologist used the morphological appearance of the lesions in addition to the contrast enhancement characteristics, whereas mean shift clustering and histogram analysis were based on enhancement kinetics only.

Histogram analysis, as implemented in our study, resulted in a significantly lower diagnostic performance than mean shift clustering. The first, second, and third quartiles of each feature were used for classification, and similar to Furman-Haran et al (15), we found the first quartile to yield the best performance. Mean shift clustering has some advantages compared to using quartiles. For example, the quartile approach discards potentially valuable information such as the overall distribution of feature values. Also, histogram analysis in general does not offer spatial contiguity; the most suspicious voxels are typically scattered throughout the lesion. This property is independent of the features extracted from the histograms. Mean shift MDC inherently produces clusters with contiguous voxels.

There is limited published data on the use of computer-aided diagnosis for the analysis of breast MRI lesions in general, as well as its clinical application. Gilhuijs et al (18) developed a method to evaluate several morphological features of breast lesions. This approach resulted in encouraging results in particular using analysis of lesion sharpness and shape, using a variance of margin gradient and a radial gradient histogram feature, respectively. An improved version of this system was tested in a screening situation (19), which resulted in an negative predictive value (NPV) of 99% while still maintaining an acceptable positive predictive value (PPV) of 50%. Deurloo et al (20) later showed that this system could potentially increase performance in the evaluation of clinically and mammographically occult lesions.

Penn et al (21) reported on a CAD system based on fractal discrimination of lesion borders. Their technique is based on earlier work where an expert description of lesion morphology is combined with the fractal dimension feature and then used for lesion discrimination using a back-propagation neural network or logistic regression (22,23). A possible limitation of this approach lies in the dependence on expert description of the lesions, because several features generally accepted in radiology literature show considerable interobserver variability (6).

In order to obtain spatially contiguous ROIs, fuzzy c-means clustering incorporating spatial information (24) could also be used, but this method inherently requires a preset number of clusters. Our method has an adaptive cluster size and, thus, results in more clusters (and a smaller ROI) when lesions are heterogeneous rather than when lesions are relatively homogeneous. This property better reflects the method of ROI determination used by most radiologists.

The delineation of the lesions was obtained using a simple algorithm that needed user intervention 36% of the time. This work was aimed at the determination of an ROI in a defined lesion, and although it should be

possible to automatically determine the exact extent of lesions with less user interventions, improvement of delineation accuracy is not straightforward and will be the subject of future research.

This study has some limitations. First, the lesion delineation step (based on relative enhancement) required frequent operator adjustments. This may somewhat lower the reproducibility of the delineation, even though the final clustering result is probably not very sensitive to minor variations in lesion delineation. Only when the most malignant section of the lesion is erroneously not included, a different cluster would be identified as ROI. Such a situation is, however, very unlikely as the most malignant part of a lesion usually enhances strongly and would, therefore, be included in the delineation. Operator adjustments during lesion delineation also have a potential influence on the feature values, since these were normalized using all voxels within all delineated lesions. We expect this effect to be minimal, thanks to the very large total number of voxels.

Second, the current implementation of mean shift clustering does not include automated bandwidth selection. This is an important issue since the bandwidth defines the number of clusters as well as the size of the ROI. The currently used fixed bandwidth values h_s and h_r were determined empirically and produced encouraging results. Better results may be obtained by setting different h_s and h_r based on, for example, an analysis of the stability of the segmentation results. In this situation, h_s and h_r are defined by the center of the largest range of bandwidths that leads to the same number of clusters. Alternatively, an adaptive bandwidth selection method (25) could be used, one which is aimed at optimizing the bandwidth parameters for each individual case. Theoretically, minimization of the asymptotic mean square error (AMISE) as described in Cacoullos (26) could be attempted, but in a multivariate situation such as in the current work, the resulting formula cannot be directly solved (27) and is therefore of limited use.

Third, the performance of radiologists was estimated by the final assessment of only one, albeit experienced, reader. Further research, aimed at the practical use of automated determination of the ROI, will include a multiple reader-multiple case analysis. It was also not possible to properly evaluate the performance of the CAD methods and the radiologist for the lesion subtypes (Table 1), due to the relatively small number of lesions. It would be interesting to know if automated ROI determination is more beneficial to the radiologist for certain subtypes than for others, and further research into this subject seems warranted.

Finally, the results presented in this work are not independent of MRI scanner type, scanning protocol, or even patients. We used two simple features commonly used by radiologists, namely the relative enhancement at the first time point after contrast injection and the amount of contrast washout. The exact value of both features depends on many factors, including scanner field strength, the amount and concentration of the injected contrast medium, the MRI scanning protocol, as well as patient characteristics such as cardiac output. Because of this variation, the current system of clustering and classification is not well suited for generic application. Only the use of pharmacokinetic modeling (see, for example, Tofts (28)) will allow building a clustering procedure and classifier that could be used with data from varying MRI scanning protocols; we therefore intend to use modeled parameters as enhancement features in future work. Further research will focus on these limitations: improvement of the reproducibility of lesion delineation, automated bandwidth selection, and the implementation of pharmacokinetic modeling.

In conclusion, MS-MDC of breast MRI lesions followed by automatic selection of the most suspicious cluster is an accurate method to automatically determine a contiguous region of interest for the analysis of contrast enhancement. There was no statistical difference in diagnostic accuracy with an experienced radiologist. Compared to histogram analysis, this technique performed significantly better.

APPENDIX A

Mean Shift Procedure

Mean shift segmentation is an iterative technique that finds cluster centers at areas of highest density in feature space. Each data point in the feature space is iteratively translated until a stationary point is reached. Such a stationary point is found at locations where the density gradient estimate equals zero and is also known as a mode. The iterative translations take place along the mean shift vector, which for each location represents the direction of the greatest increase in density.

The mean shift vector can be calculated without actually estimating the local density itself. Consider the kernel density estimator $\hat{f}(\mathbf{x})$ for n d-dimensional data points $\mathbf{x}_i \in \mathbf{X}$, $i = 1, \ldots, n$ and a symmetric positive definite bandwidth matrix $\mathbf{H}$ of size $d \times d$:

$$\hat{f}_K(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x}) \cdot (\mathbf{x} - \mathbf{x}_i). \qquad \text{(A1)}$$

Employing a single bandwidth h and assuming Euclidean metric, the bandwidth matrix $\mathbf{H}$ becomes $\mathbf{H} = h^2\mathbf{I}$, and Eq. [A1] can be written as:

$$\hat{f}_K(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{(\mathbf{x} - \mathbf{x}_i)}{h}\right\|^2\right), \qquad \text{(A2)}$$

where $k(\|\mathbf{r}\|^2)$ is the profile of a radially symmetric normalized kernel K so that $K(\mathbf{r}) = k(\|\mathbf{r}\|^2)$. The locations where the kernel density gradient estimate $\hat{\nabla} f(\mathbf{x}) \equiv \nabla \hat{f}(\mathbf{x})$ equals 0 indicate the local maxima, local minima, or plateaus in the density estimate. These points are determined as follows. After defining the derivative profile $g(\|\mathbf{r}\|^2) = -k'(\|\mathbf{r}\|^2)$ and using Eq. [A2], we can write the density gradient estimate as:

$$\hat{\nabla} f_{h,K}(\mathbf{x}) = \frac{2}{nh^{d+2}}\sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x}) g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right) \qquad \text{(A3)}$$

$$= \frac{2}{h^2}\hat{f}_{h,G}(\mathbf{x}) \cdot m_{h,G}(\mathbf{x}). \quad \text{(A4)}$$

The second part of Eq. [A4] is called the mean shift, which is the difference between a point **x** and the mean of all other points from *X*, weighted by the kernel *G* around **x**:

$$m_{h,G}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \quad \text{(A5)}$$

The mean shift vector thus points toward the maximum increase in density. Note that $m(\mathbf{x})$ can be calculated directly using Eq. [A5], without determining the density estimate itself.

The mean shift procedure is the iterative process that consists of successive translations of **x** along the local mean shift vector, which is calculated anew for each iteration. It can be proven that this is guaranteed to converge to a stationary point, given two mild conditions for kernels *G* and *K* (9). Since such a stationary point could represent a local minimum or plateau, its location is perturbed by a small amount, after which the mean shift procedure is again started. An (almost) equal result indicates a local maximum of the density estimate, i.e., a mode.

An important feature of the mean shift procedure is that the step size of the iterations is adaptive, and inversely proportional to the local density in feature space. This is apparent after writing Eq.[A3] as

$$m_{h,G}(\mathbf{x}) = \frac{h^2}{2} \cdot \frac{\hat{\nabla} f_{h,K}(\mathbf{x})}{\hat{f}_{h,G}(\mathbf{x})}, \quad \text{(A6)}$$

which shows that the mean shift vector computed with kernel *G* is proportional to the gradient density estimate using *K*, normalized with the density estimate using *G*. In areas of low density, the magnitude of $m_{h,G}(\mathbf{x})$ will therefore be large, and in areas with high density (i.e., near the modes), the magnitude will be small.

Clustering

In our application, the spatial domain and range domain are combined, resulting in five-dimensional data points $\mathbf{y}_i$. The two domains are assumed to have Euclidean metric, but are allowed separate values for the bandwidth. The spatial domain, with bandwidth h_s, represents the voxel locations, the (normalized) range domain with bandwidth h_r the MRI features (washout and relative enhancement). In this case, the kernel *K* in Eq. [A1] is defined as the product of two radially symmetric kernels:

$$K_{h_s,h_r}(\mathbf{x}) = \frac{1}{h_s^2 h_r^2} k\left(\left\|\frac{\mathbf{x}_s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}_r}{h_r}\right\|^2\right), \quad \text{(A7)}$$

where $\mathbf{x}_s$ is the spatial part and $\mathbf{x}_r$ the range part of **x**. The two parts use a common kernel profile *k*. We used a Gaussian kernel in our experiments.

For each data point $\mathbf{y}_i$ (represented by each voxel in the delineated lesion volume) the mean shift procedure is performed, leading to a point of convergence $\mathbf{z}_i$ for each $\mathbf{y}_i$. This results in *m* groups of closely related $\mathbf{z}_i$, and each of these groups represents a cluster. The clusters C_p with $p = 1, \ldots, m$ are thus defined by pooling all $\mathbf{y}_i$ which have $\mathbf{z}_i$ within a distance h_s of each other in the spatial domain, and h_r in the range domain. All $\mathbf{y}_i | \mathbf{z}_i \in C_p$ form the basin of attraction of cluster C_p.

REFERENCES

1. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. Radiology 2006;238:42–53.
2. Mussurakis S, Buckley D, Horsman A. Dynamic MR imaging of invasive breast cancer: correlation with tumour grade and other histological factors. Br J Radiol 1997;70:446–451.
3. Mussurakis S, Buckley DL, Coady AM, Turnbull LW, Horsman A. Observer variability in the interpretation of contrast enhanced MRI of the breast. Br J Radiol 1996;69:1009–1016.
4. Kuhl C, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology 1999;211:101–110.
5. American College of Radiology (ACR). ACR BIRADS—magnetic resonance imaging. In: ACR breast imaging reporting and data system, breast imaging atlas. Reston, VA: American College of Radiology (ACR); 2003.
6. Stoutjesdijk M, Futterer J, Boetes C, van Die L, Jager G, Barentsz J. Variability in the description of morphologic and contrast enhancement characteristics of breast lesions on magnetic resonance imaging. Investigative Radiology 2005;40:355–362.
7. Kim S, Morris E, Liberman L, et al. Observer variability and applicability of BI-RADS terminology for breast MR imaging: invasive carcinomas as focal masses. Am J Roentgenol 2001;177: 551–557.
8. Ikeda D, Hylton N, Kinkel K, et al. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. J Magn Reson Imaging 2001; 13:889–895.
9. Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 2002;24:603–619.
10. Kuhl C, Bieling H, Gieseke J, et al. Healthy premenopausal breast parenchyma in dynamic contrast enhanced MR imaging of the breast: normal contrast medium enhancement and cyclical-phase dependency. Radiology 1997;203:137–144.
11. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes: the art of scientific computing, 2nd edition. Cambridge (UK) and New York: Cambridge University Press, 1992. p 661–666.
12. Fukunaga K, Hostetler LD. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Information Theory 1975;21:32–40.
13. Cheng Y. Mean shift, mode seeking, and clustering. IEEE Trans on Pattern Analysis and Machine Intelligence 1995;17:790–799.
14. Hayes C, Padhani AR, Leach MO. Assessing changes in tumour vascular function using dynamic contrast-enhanced magnetic resonance imaging. NMR Biomed 2002;15:154–163.
15. Furman-Haran E, Schechtman E, Kelcz F, Kirshenbaum K, Degani H. Magnetic resonance imaging reveals functional diversity of the vasculature in benign and malignant breast lesions. Cancer 2005; 104:708–718.
16. Rutter CM. Bootstrap estimation of diagnostic accuracy with patient-clustered data. Acad Radiol 2000;7:413–419.
17. Obuchowski NA, Lieber ML, Powell KA. Data analysis for detection and localization of multiple abnormalities with application to mammography. Acad Radiol 2000;7:516–525.

18. Gilhuijs KG, Giger ML, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. Med Phys 1998;25:1647–1654.
19. Gilhuijs KGA, Deurloo EE, Muller SH, Peterse JL, Kool LJS. Breast MR imaging in women at increased lifetime risk of breast cancer: clinical system for computerized assessment of breast lesions initial results. Radiology 2002;225:907–916.
20. Deurloo EE, Muller SH, Peterse JL, Besnard APE, Gilhuijs KGA. Clinically and mammographically occult breast lesions on MR images: potential effect of computerized assessment on clinical reading. Radiology 2005;234:693–701.
21. Penn A, Kumar N, Thompson S, Schnall M, Wang F, Gatsonis C. Preliminary performance analysis of breast MRI cad system. In: SPIE Medical Imaging 2001: Image Processing. Bellingham, WA: Society of Photo-Optical Instrumentation Engineers, 2001. p 1944–1953.
22. Penn AI, Loew MH. Estimating fractal dimension with fractal interpolation function models. IEEE Trans Med Imaging 1997; 16:930–937.
23. Penn AI, Bolinger L, Schnall MD, Loew MH. Discrimination of MR images of breast masses with fractal-interpolation function models. Acad Radiol 1999;6:156–163.
24. Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ. Fuzzy c-means clustering with spatial information for image segmentation. Comput Med Imaging Graph 2006;30:9–15.
25. Comaniciu, D. An algorithm for data-driven bandwidth selection. IEEE Trans Pattern Anal Mach Intell 2003;25:281–288.
26. Cacoullos T. Estimation of a multivariate density. Ann Inst Stat Math 1966;18:178–189.
27. Wand MP, Jones M. Kernel smoothing. London: Chapman & Hall, 1995. p 98.
28. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging 1997;7:91–101.