The Impact of Motion Correction on Lesion Characterization in DCE Breast MR Images

Martin Bergtholdt, Sven Kabus, Rafael Wiemker, Thomas Buelow

Philips Research Hamburg, Germany

ABSTRACT

In the context of dynamic contrast enhanced breast MR imaging we analyzed the effect of motion compensating registration on the characterization of lesions. Two registration techniques were applied: 1) rigid registration and 2) elastic registration based on the Navier-Lamé equation. Interpreting voxels that exhibit a decline in image intensity after contrast injection (compared to the non-contrasted native image) as motion outliers, it can be shown that the rate of motion outliers can be largely reduced by both rigid and elastic registration. The performance of lesion features, including maximal signal enhancement ratio and variance of the signal enhancement ratio, was measured by area under the ROC curve as well as Cohen's κ and showed significant improvement for elastic registration, whereas features derived from rigidly registered images did not in general exhibit a significant improvement over the level of unregistered data.

Keywords: Breast, Registration, Feature Extraction, Characterization and Staging

1. INTRODUCTION

In dynamic contrast enhanced Breast MRI (DCE BMRI) a T1-weighted 3D image volume is acquired before injection of a Gadolinium-based contrast agent which is then followed by a temporal sequence of 5-6 additional T1-weighted images. The degree and characteristics of the local change of image intensity over time has proven to contain relevant information for the characterization of breast lesions.[1,2] In the literature a number of ways has been proposed to quantify this kinetic behavior and use it in computer aided diagnosis (CADx) systems as quantitative feature for lesion classification.

Even though the patient's breast is fixated to some degree in the MR coils during image acquisition, some amount of patient movement during the course of the MR study cannot always be avoided. This movement becomes apparent as artifacts in derived images obtained by subtracting a native, non-contrasted MR image from a contrasted image (see Fig. 2, top row). Such motion artifacts also lead to inconsistent time-intensity curves if a fixed voxel grid is employed for sampling, since the fixed positions might not represent the same anatomic positions at different points in time due to the patient motion. This can consequently lead to corrupted time-intensity curves. Recently it was indicated that optical flow based motion compensation leads to a modest improvement in PCA based lesion classification.[3]

The purpose of this study is to investigate whether and to which extent motion correction by image registration can improve the performance of quantitative time-intensity features for lesion classification.

2. METHODS

This section describes the registration method employed for motion compensation, the procedure for obtaining the segmentation set $\mathcal{L}$ of voxels for each lesion, as well as the kinetic lesion features used for evaluation.

2.1 Motion compensation by image registration

In the following, in order to compensate for patient motion, each contrast-enhanced 3D image is transformed in order to align with the pre-contrast image of the respective dynamic imaging study. For this study a rigid and an elastic registration algorithm are compared.

Let the pre-contrast reference image be denoted by R and any of the contrast-enhanced template images by T. The template images T are first intensity normalized by a linear transformation such that their 95%-quantile is equal to that of the reference image R.

For the rigid registration six parameters are optimized by using a Gauss-Newton approach in conjunction with Armijo's step-size rule. The similarity of R and T is measured using the sum of squared differences (SSD).

Elastic registration is implemented in a multi-resolution framework.[4] Successive down-sampling defines an image pyramid for both R and T. The registration starts on the coarsest level. After convergence the resulting deformation vector field (DVF) is up-sampled to the next level. This step is repeated until the finest level (the original resolution) is reached. On each level of the pyramid a DVF $u : \mathbb{R}^3 \to \mathbb{R}^3$ is sought such that the displaced template image $T_u(x) := T(\mathrm{id} + u(x))$ minimizes the sum of a certain similarity measure $\mathcal{D}$ and a regularization term $\mathcal{S}$ which assures that the registration problem is well-posed. For $\mathcal{D}$ the SSD is chosen as for rigid registration, while for $\mathcal{S}$ an elastic regularizer based on the Navier-Lamé equation is employed:[5]

$$\mathcal{D}[u] := \frac{1}{2}\int_\Omega [R(x) - T_u(x)]^2 \mathrm{d}x \ ,$$

$$\mathcal{S}[u] := \int_\Omega \left(\frac{\mu}{4} \sum_{i,j=1}^{3} (\partial_{x_j} u_i(x) + \partial_{x_i} u_j(x))^2 + \frac{\lambda}{2} (\nabla \cdot u(x))^2 \right) \mathrm{d}x \ .$$

The elastic regularizer depends on two material parameters, the so-called Lamé constants λ and μ which can be translated into Poisson's ratio and Young's modulus. Since the expected deformations are small, $\lambda = 0$ and $\mu = 1$ are chosen corresponding to a larger Young's modulus and resulting in a relatively stiff deformation. This choice of material properties is an alternative to enforcing local volume preservation and ensures robustness against noise and, moreover, minimizes disruptive impact of local intensity changes due to contrast agent inflow.

By combining the similarity measure and the regularization term, the registration problem is formulated as minimizing the joint functional

$$\mathcal{D}[u] + \mathcal{S}[u] \stackrel{u}{\to} \min \ .$$

Based on the calculus of variations, a system of non-linear partial differential equations has to be solved:

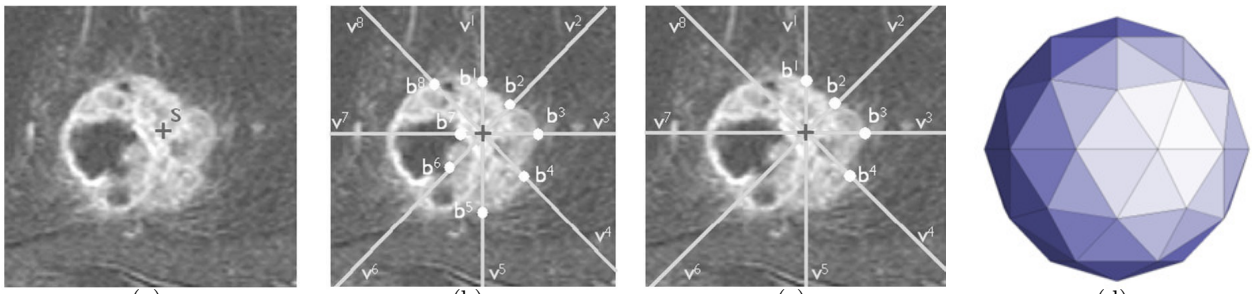
$$\mu \Delta u + (\mu + \lambda) \nabla \cdot \nabla u = \nabla T_u (R - T_u) \ . \tag{1}$$

For the discretization of Eq. (1) finite differences in conjunction with Neumann boundary conditions have been chosen. The resulting system of linear equations consists of a sparse, symmetric and highly structured matrix arising from the regularizer, and of a so-called force vector corresponding to the similarity measure. The system of equations is then iteratively solved by a conjugate gradient scheme. The iteration is stopped if the update in u is below 0.05 mm for all positions indicating convergence.

2.2 Segmentation of mass lesions

The tumors were segmented using a previously presented automatic computerized segmentation algorithm for mass lesions in BMRI images.[6,7] The segmentation algorithm takes a subtraction image (*post-contrast image* − *pre-contrast image*) as input image which is usually also used for visual assessment by radiologists. The main components of the method are 1) lesion selection by the user, 2) automatic intensity threshold estimation, 3) connected component analysis, and 4) a post-processing step consisting of a hole-filling procedure and a leakage removal step leading to the final segmentation result.

The intensity threshold estimated in step 2) is used again to restrict the set of voxels from the final segmentation result to those voxels used for kinetic analysis and feature computation, i.e. the intersection of the voxel sets 2) and 4). Therefore, a brief description of step 2) is included in the following. For the other components of the segmentation algorithm see previous publications Ref. 6, 7.



(a) (b) (c) (d)

Figure 1. Ray casting approach to threshold estimation. (a) A seed point s is located inside the tumor to be segmented. (b) From the seed point a set of rays is cast (shown here in 2D). On each ray a tumor boundary point is estimated. (c) The set of boundary points is pruned to the most likely ones. An intensity threshold is estimated by averaging the intensity values at the remaining boundary point positions. (d) One rays is cast from the seed position for each vertex of a subdivided icosahedron in order to cover the tumor surface as regularly as possible.

2.2.1 Estimation of the intensity threshold

In the proposed approach the image intensities are sampled along 1D rays emanating from a user defined seed point position $\mathbf{s}$ inside the tumor (Fig. 1 (a)). The 3D spatial distribution of rays is arranged according to a subdivided icosahedron (Fig. 1 (d)). Each vertex ($\mathbf{v}^i \in \mathbb{R}^3, i \in \{1, \ldots, 42\}$) of the subdivided icosahedron corresponds to one ray. Along each of the rays, the image data are sampled to yield a 1D signal r^i per ray.

$$r^i(x) := (I_1 - I_0)\,(\mathbf{s} + x\mathbf{v}^i) \tag{2}$$

Each ray signal r^i is searched for the smallest value x^i exhibiting locally steepest decline $(d^2r^i/dx^2)(x^i) = 0$ and $(d^3r^i/dx^3)(x^i) > 0$) *and* a signal intensity below a given factor α of the maximal intensity along r^i: $r^i(x^i) < \alpha \max_{x<x^i}(r^i(x))$. For all computations $\alpha = 2/3$ was used. The 3D positions $\mathbf{b}^i = \mathbf{s} + x^i\mathbf{v}^i$ are interpreted as tumor boundary candidates (see Fig. 1 (b)).

In order to remove potential outliers, the smallest and largest 10 values x^i are discarded, respectively (Fig. 1 (c)). Assuming that the x^i are sorted such that $x^i \leq x^j$ for $i < j$ the segmentation threshold is defined as

$$t := \frac{1}{22}\sum_{i=11}^{32} r^i(x^i). \tag{3}$$

2.3 Kinetic features for lesion characterization

The characteristics of time-intensity curves in DCE BMRI has long been known as a relevant clinical feature in the prediction of the dignity of breast lesions. The time intensity curves are usually categorized as "persistent enhancement" (Type 1), "plateau" (Type 2), and "washout" (Type 3), with Type 3 being the curve type most indicative of a malignant lesion.[2]

Recently, it was shown that not only the curve type of the most suspicious curve, but also the degree of heterogeneity of the curve types within a lesion is a significant predictor of the status of a lesion.[8,9] The curve type is characterized by the signal enhancement ratio[10]

$$\mathrm{SER}_1(x) = \frac{I_1(x) - I_0(x)}{I_{\mathrm{last}}(x) - I_0(x)}$$

This heuristic feature is invariant under linear intensity transformations and has been shown to be closely correlated to more physiologically motivated pharmacokinetic parameters.[11] In our implementation the intensity values I_1 were chosen from the dynamic contrast enhanced series at acquisition time closest to 2 minutes after

the pre-contrast image I_0 in order to cope with different time intervals for different DCE protocols on the diverse systems.

In this study the effect of motion compensation on the features from Ref. 9 are analyzed within the lesion: the maximal SER value, the extent of the range of SER values, mean and variance of the SER histogram, entropy of the histogram, flatness of the histogram (maximum bin-size), and for comparison also the volume of the lesion.

3. RESULTS

3.1 Data preparation

131 DCE breast MR studies were acquired on Philips MR Intera Achieva 1.5 T and GE Signa 1.5 T scanners. One mass lesion per case was analyzed. Lesions smaller than 20 voxels were discarded as well as lesions without an available biopsy proven diagnosis, resulting in 90 remaining lesions, 33 benign, 57 malignant. Each feature was evaluated on the set $\mathcal{L}$ of voxels which represent a lesion according to the automatic lesion segmentation algorithm from Sec. 2.2.

In a previous study[9] only voxels with initial relative intensity increase (due to contrast injection) of 10% or more were considered. This inclusion criterion was changed such that in the present study only tumor voxels $\mathbf{x}$ that show an absolute intensity uptake above the threshold used in the segmentation procedure $I_1(\mathbf{x}) - I_0(\mathbf{x}) > t$ with t as defined in (3) are included. The reason for this adaptation is that due to the post processing step of the segmentation procedure the segmented tumor volume might contain portions of interior voxels that show very low image intensity, e.g. due to necrotic areas of the tumor. At these low intensity levels even a slight noise level can yield a spurious high relative intensity increase, even though no interpretable kinetic curve is present. By adapting the selection criterion as described above non-enhancing voxels are excluded from the SER analysis.

Furthermore, outlier voxels due to motion artifacts are excluded from the analysis. A voxel is defined as *motion outlier* whenever $I_{\text{last}}(\mathbf{x}) - I_0(\mathbf{x}) \leq 0$ (indicating an intensity decrease after contrast agent injection which is not physiologically plausible and thus likely due to motion, assuming a negligible noise-level).

The registration algorithms are applied to the region of interest (ROI) around each lesion. Given the center position c and maximum radius r of a lesion (as indicated by the user when launching the lesion segmentation algorithm), an ROI is initialized as a cube with center c and $\max\{2.4r, 80\text{mm}\}$ edge length to ensure that the ROI contains the entire lesion and sufficient additional image content. In order to reduce the influence of MR cardiac artifacts for breast lesions in dorsal position, the ROI is shrinked in anterior-posterior direction by 20mm. All images are cropped according to the ROI prior to application of the registration.

It should be noted that in contrast to the outlier sets, the segmentation set $\mathcal{L}$ was defined on the unregistered native data. The same set was used for all three methods, unregistered, rigid registration and elastic registration.

3.2 Evaluation

Visual results for an example lesion are shown in Fig. 2 indicating significant reduction of motion artifacts by both registration algorithms.

Outlier rates. The rate of voxels labeled as motion outliers per lesion is analyzed for all 90 lesions. The results are shown in Fig 3. Without motion compensation 13% of all lesions contain more than 2% motion outliers per lesion whereas only 2% of the lesions contain this amount of motion outliers after rigid and elastic registration. Clearly motion compensation reduces the number of outlier voxels with respect to this criterion.

To contrast these multi-vendor results, an additional selection of cases for a single vendor/system consisted of all instances from the Philips Intera Achieva system. Considering only these cases, the data heterogeneity is significantly reduced. The subsequent results are presented for both, all data and only the Philips Intera Achieva data. Table 1 summarizes the data selection and results with respect to the motion outliers. In the following results all lesions with more than 2% outlier voxels in *any* of the three cases (unregistered, rigid, elastic) are omitted, resulting in 77 lesions (13 outlier) and 33 lesions (5 outliers) for "all" and "Achieva only" data, respectively.

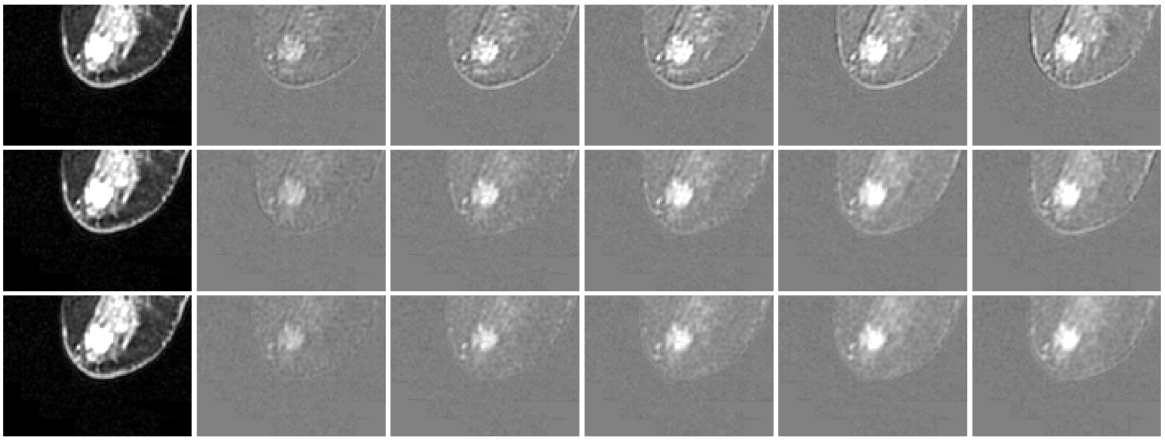


Figure 2. From left to right: A slice from the pre-contrast MR volume, subtraction images (post-contrast - pre-contrast) at the 5 time-points of the dynamic study. Top row: Unregistered. Middle row: Rigid motion compensation. Bottom row: Elastic motion compensation. Especially the dark lines in the original, unregistered case are a strong indication of motion artifacts.

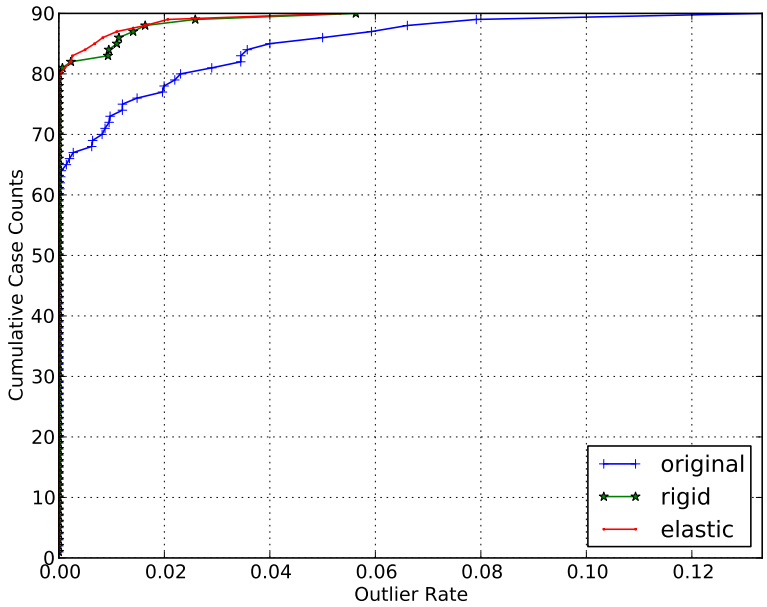


Figure 3. Cumulative rate of motion outlier voxels per lesion. The graph can be understood as, e.g., "78(87%) images have less than 2% outlier voxels for the unregistered case. This number increases to 88(98%) for both registered cases. Each marker represents a single data instance."

Table 1. Data selection. The first row shows the results of all data instances, the second using only data from the Philips Intera Achieva system. The columns are: total number of instances, mean outlier voxels for unregistered/rigid/elastic in ‰, invalid instances with more than 2% motion outlier voxels (union and unregistered/rigid/elastic individually), number of valid instances after removing of union of invalid, of which benign, and malignant lesions.

	Total	Mean Outlier Rate (‰)	Invalid	Valid	Benign	Malignant
All	90	8.24/1.74/1.41	13 (12/2/2)	77	30	47
Achieva	38	6.20/1.00/0.58	5 (4/1/1)	33	12	21

Lesion characterization. The performance of the lesion features from Sec. 2.3 is summarized in Fig. 4 and Table 2. Fig. 4 presents a performance analysis based on the receiver operating characteristic (ROC),[12] which plots the true positive rate (TPR) vs. the false positive rate (FPR). Each operating point on the graph corresponds to the classification based on a single threshold, individually for each respective feature type.

In order to quantify the feature performance in Table 2 the area under the ROC curve (AUC) is computed, as well as the highest accuracy (ACC) (true positive rate + true negative rate), and the highest κ-statistics[13] for particular thresholds. The κ-statistics is more suited than accuracy in cases where positive and negative class have different prior occurrences, since it indicates the agreement beyond that expected by pure chance. It is computed as

$$\kappa = \frac{\text{accuracy} - \text{chance}}{1 - \text{chance}},$$

where chance is the expected agreement when randomly selecting classes according to their marginal occurrence probability and can be computed as

$$\text{chance} = (\text{TPR} + \text{FNR})(\text{TPR} + \text{FPR}) + (\text{TNR} + \text{FPR})(\text{TNR} + \text{FNR}),$$

with TNR, the true negative rate and FNR, the false negative rate.

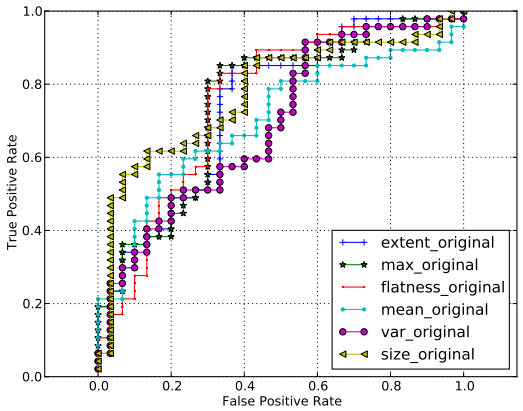
Almost all features benefit from the elastic registration in their ability to discriminate between benign and malignant. The benefit seems larger for features with high discriminative power such as “max”, and “extent”, compared to features with apparently lower discriminative power (“mean”, “variance”). For the latter, a clear trend is not attestable, with one method winning over the others by small margins at an overall low performance level and depending on the measure used; leaving a feeling of fitting the small differences in the data due to noise.

These observations are even more pronounced when looking only at the homogeneous set of Philips Intera Achieva cases: again with the exception of the poor “mean” feature, all measures improve significantly for the elastic method compared to the original data as well as the rigid method. E.g. for the “max”, “extent”, “flatness”, “entropy” features, the mean κ measure increases from 0.61 for unregistered data to 0.72 for the elastically registered data. Interestingly performance seems to degrade using the simple rigid registration, with only 0.49 for the mean κ value for these features. Also, as expected, the performance of all measures and methods, generally increases with this less heterogeneous data set.

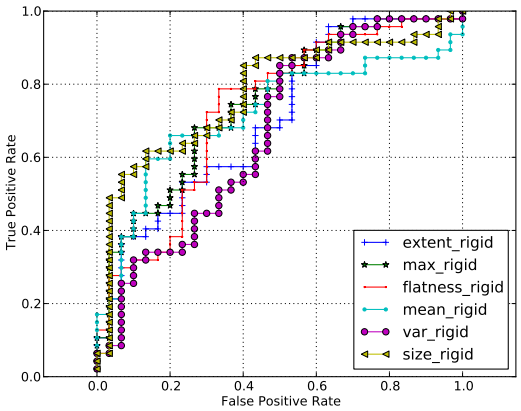
As a first step towards multi-feature integration, logistic regression with a maximum-likelihood estimator[14] was performed on the set of “max”, “mean”, “variance”, “flatness”, and “size”. These features were chosen since they represent complementary aspects of the data (“extent” is similar to “max”, “entropy” is similar to “flatness”). Using more features can improve the classification results significantly, e.g. up to 95% sensitivity (TPR) at a level of 100% specificity (1-FPR), see Table 2, last row of Philips data and Fig. 4 (f). It has to be noted, however, that the optimization of the regression parameters was performed on the same data for which these results are presented. To overcome problems of overfitting, a much more rigorous analysis is necessary with more data instances to allow for separate training- and test-data.

4. CONCLUSION

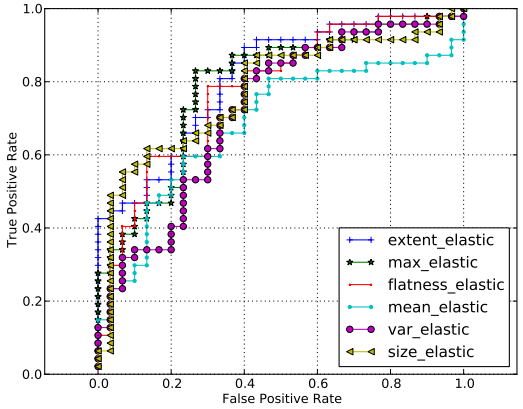
In this study, the impact of motion compensation on the characterization of mass lesions was investigated. It was shown that the amount of voxels perceived as motion outliers by our definition in Sec. 3.1 can be consider-



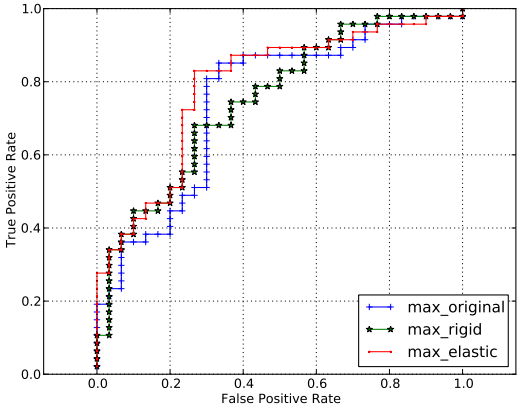
(a) original data



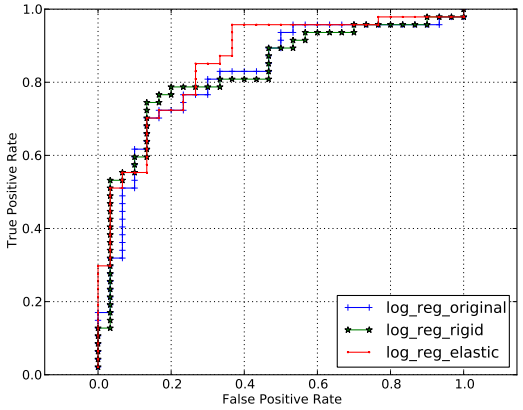
(b) rigid registration



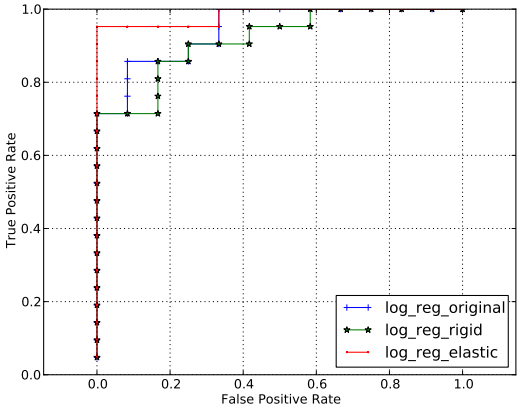
(c) elastic registration



(d) "max" feature



(e) logistic regression (all)



(f) logistic regression (Intera Achieva)

Figure 4. The Receiver Operating Characteristic for the features "extent", "max", "flatness", "mean", "variance", and "size" ((a) through (c)). ROC curves for each individual methods and the "max" feature (d). Results using logistic regression are presented for all data instances (e) and using only instances from the Philips Achieva system (f). For discussion, see text.

Table 2. Quantitative evaluation of feature performance depending on the registration type. The upper table shows the performance on all 77 valid cases, the lower using only Philips Intera Achieva cases. For each method (unregistered, rigid registration, elastic registration), three performance measures are presented: area under ROC curve (AUC), accuracy (ACC), and κ-statistics. Each row represents a particular feature type as described in Sec. 2.3. The last row corresponds to a logistic regression using the features "max", "mean", "variance", "flatness", and "size". The best performing method is set in bold font individually for each feature and performance measure. Additionally the overall best (single) feature per performance measure is set in italic. For comparison, lesion volume ("size") yields (AUC, ACC, κ): (0.7759, 0.7532, 0.4651) for all instances and (0.8492, 0.8485, 0.6541) for the Intera Achieva instances. For discussion, see text.

All 77 valid cases									
	No registration			Rigid registration			Elastic registration		
Feature	AUC	ACC	κ	AUC	ACC	κ	AUC	ACC	κ
Max	0.7479	0.7792	0.5273	0.7482	0.7013	0.3974	**0.7894**	***0.7922***	***0.5631***
Mean	0.7014	0.6623	0.3525	**0.7184**	**0.7143**	**0.4334**	0.6830	0.7013	0.3525
Variance	0.6908	**0.7273**	0.3778	0.6681	0.7143	0.3689	**0.7220**	**0.7273**	**0.4160**
Extent	0.7433	0.7662	0.4964	0.7050	0.7273	0.3611	***0.8099***	**0.7792**	**0.5154**
Flatness	0.7525	**0.7662**	**0.5025**	0.7287	0.7403	0.4539	**0.7784**	0.7532	0.4843
Entropy	0.7050	**0.7662**	0.4702	0.7213	**0.7662**	0.4836	**0.7532**	**0.7662**	**0.5025**
Log. Reg.	0.8234	0.7662	0.5365	0.8277	0.7922	0.5832	**0.8582**	**0.8312**	**0.6247**
33 Philips Intera Achieva cases									
Max	0.7897	0.8182	0.6071	0.7183	0.6970	0.3678	**0.8492**	**0.8485**	**0.6784**
Mean	**0.6944**	0.6667	0.3795	0.6468	**0.6970**	**0.3889**	0.5992	0.6667	0.2924
Variance	0.6905	0.7576	0.4359	0.6548	0.7879	0.4762	**0.8095**	**0.8182**	**0.5769**
Extent	0.8135	0.8485	0.6667	0.7381	0.7879	0.4967	**0.9286**	***0.9091***	***0.7925***
Flatness	0.8571	0.8485	0.6541	0.7619	0.8182	0.5926	***0.9444***	**0.8788**	**0.7556**
Entropy	0.7937	0.7879	0.4967	0.7103	0.7879	0.4967	**0.8929**	**0.8485**	**0.6541**
Log. Reg.	0.9444	0.8788	0.7471	0.9167	0.8485	0.6784	**0.9841**	**0.9697**	**0.9357**

ably reduced by both rigid and elastic registration. The performance of quantitative features was significantly improved by elastic registration, but apparently degraded by the simpler rigid registration. The latter finding seems unexpected since images are visually improved by rigid registration and outlier rate is also reduced. The cause of this phenomenon will be subject to further investigation.

REFERENCES

[1] Heywang, S. H., Hahn, D., Schmidt, H., Krischke, I., Eiermann, W., Bassermann, R., and Lissner, J., "MR imaging of the breast using gadolinium-DTPA," *Journal of computer assisted tomography* **10**(2), 199204 (1986).

[2] Kuhl, C. K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieseke, J., and Schild, H. H., "Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions?," *Radiology* **211**(1), 101—110 (1999).

[3] Steinbrücker, F., Meyer-Base, A., Wismüller, A., and Schlossbauer, T., "Application and evaluation of a motion compensation technique to breast MRI," in [*Proceedings of SPIE*], 73470J–73470J–8 (2009).

[4] Modersitzki, J., [*Numerical Methods for Image Registration*], Oxford University Press (2004).

[5] Broit, C., *Optimal registration of deformed images*, PhD thesis, Pennsylvania (1981).

[6] Buelow, T., Meinel, L. A., Wiemker, R., Kose, U., Shimauchi, A., and Newstead, G., "Segmentation of suspicious lesions in dynamic contrast-enhanced breast MR images," in [*SPIE Medical Imaging: CAD*], **6514**, 27 (2007).

[7] Meinel, L. A., Buelow, T., Huo, D., Shimauchi, A., Kose, U., Buurman, J., and Newstead, G., "Robust segmentation of mass-lesions in contrast-enhanced dynamic breast MR images," *Journal of Magnetic Resonance Imaging* **32**(1), 110–119 (2010).

[8] Shimauchi, A., Jansen, S., Zak, L., Abe, H., and Newstead, G., "Can assessment of kinetic heterogeneity of breast mass lesion enhancement seen on MR imaging predict malignancy?," 6008894 (2008).

[9] Buelow, T., Saalbach, A., Bergtholdt, M., Wiemker, R., Buurman, H., Meinel, L. A., and Newstead, G., "Heterogeneity of kinetic curve parameters as indicator for the malignancy of breast lesions in DCE MRI," in [*SPIE Medical Imaging: CAD*], **7624** (2010).

[10] Esserman, L., Hylton, N., George, T., and Weidner, N., "Contrast-enhanced magnetic resonance imaging to assess tumor histopathology and angiogenesis in breast carcinoma," *The Breast Journal* **5**(1), 13–21 (1999).

[11] Li, K., Henry, R. G., Wilmes, L. J., Gibbs, J., Zhu, X., Lu, Y., and Hylton, N. M., "Kinetic assessment of breast tumors using high spatial resolution signal enhancement ratio (SER) imaging," *Magnetic Resonance in Medicine* **58**(3), 572–581 (2007).

[12] Fawcett, T., "An introduction to ROC analysis," *Pattern Recognition Letters* **27**, 861–874 (June 2006).

[13] Cohen, J., "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement* **20**(1), 3746 (1960).

[14] Czepiel, S. A., "Maximum likelihood estimation of logistic regression models: theory and implementation,"