

# Comparative analysis of nonlinear dimensionality reduction techniques for breast MRI segmentation<sup>a)</sup>

Alireza Akhbardeh

*The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*

Michael A. Jacobs<sup>b)</sup>

*The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 and Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*

(Received 31 May 2011; revised 17 January 2012; accepted for publication 17 January 2012; published 3 April 2012)

**Purpose:** Visualization of anatomical structures using radiological imaging methods is an important tool in medicine to differentiate normal from pathological tissue and can generate large amounts of data for a radiologist to read. Integrating these large data sets is difficult and time-consuming. A new approach uses both supervised and unsupervised advanced machine learning techniques to visualize and segment radiological data. This study describes the application of a novel hybrid scheme, based on combining wavelet transform and nonlinear dimensionality reduction (NLDR) methods, to breast magnetic resonance imaging (MRI) data using three well-established NLDR techniques, namely, ISOMAP, local linear embedding (LLE), and diffusion maps (DfM), to perform a comparative performance analysis.

**Methods:** Twenty-five breast lesion subjects were scanned using a 3T scanner. MRI sequences used were T1-weighted, T2-weighted, diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) imaging. The hybrid scheme consisted of two steps: preprocessing and postprocessing of the data. The preprocessing step was applied for  $B_1$  inhomogeneity correction, image registration, and wavelet-based image compression to match and denoise the data. In the postprocessing step, MRI parameters were considered data dimensions and the NLDR-based hybrid approach was applied to integrate the MRI parameters into a single image, termed the embedded image. This was achieved by mapping all pixel intensities from the higher dimension to a lower dimensional (embedded) space. For validation, the authors compared the hybrid NLDR with linear methods of principal component analysis (PCA) and multidimensional scaling (MDS) using synthetic data. For the clinical application, the authors used breast MRI data, comparison was performed using the postcontrast DCE MRI image and evaluating the congruence of the segmented lesions.

**Results:** The NLDR-based hybrid approach was able to define and segment both synthetic and clinical data. In the synthetic data, the authors demonstrated the performance of the NLDR method compared with conventional linear DR methods. The NLDR approach enabled successful segmentation of the structures, whereas, in most cases, PCA and MDS failed. The NLDR approach was able to segment different breast tissue types with a high accuracy and the embedded image of the breast MRI data demonstrated fuzzy boundaries between the different types of breast tissue, i.e., fatty, glandular, and tissue with lesions (>86%).

**Conclusions:** The proposed hybrid NLDR methods were able to segment clinical breast data with a high accuracy and construct an embedded image that visualized the contribution of different radiological parameters. © 2012 American Association of Physicists in Medicine. [DOI: 10.1118/1.3682173]

Key words: manifold learning, dimensionality reduction, machine learning, data fusion, wavelet transform, image segmentation, visualization, multimodal medical images, multiparametric MRI, tumor detection, diffusion-weighted imaging, breast, cancer

## I. INTRODUCTION

Diagnostic radiological imaging techniques are powerful noninvasive tools with which to identify normal and suspicious regions within the body. The use of multiparametric imaging methods, which can incorporate different functional radiological parameters for quantitative diagnosis, has been

steadily increasing.<sup>1-3</sup> Current methods of lesion detection include dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) and/or positron emission tomography/computed tomography (PET/CT) images, which generate large amounts of data. Therefore, image-processing algorithms are required to analyze these images and play a key role in helping radiologists to differentiate normal from

abnormal tissue. In multiparametric functional radiological imaging, each image sequence or type provides different information about a tissue of interest and its adjacent boundaries.<sup>4-7</sup> For example, in multiparametric breast MRI, diffusion-weighted imaging (DWI) and Dynamic Contrast Imaging (DCE) MRI provide information about cellularity and the vascular profile of normal tissue and tissue with lesions.<sup>8-10</sup> Similarly, PET/CT data provide information on the metabolic state of tissue.<sup>11,12</sup> However, combining these data sets can be challenging, and because of the multidimensional structure of the data, methods are needed to extract a meaningful representation of the underlying radiopathological interpretation.

Currently, most computer-aided diagnosis (CAD) can act as a second reader in numerous applications, such as breast imaging,<sup>13-15</sup> Most CAD systems are based on pattern recognition and use Euclidean distances, correlation, or similar methods to compute similarity between structures in the data segmentation procedure.<sup>13,16</sup> It has been shown that methods based on Euclidean distance and other similarity measures cannot fully preserve data structure, which will negatively affect the performance of a CAD system.<sup>17,18</sup> To overcome this limitation, we propose a new hybrid methodology, based on advanced machine learning techniques, and called “dimensionality reduction” (DR), which preserves the prominent data structures and may be useful for quantitatively integrating multimodal and multiparametric radiological images. Dimensionality reduction methods are a class of algorithms that use mathematically defined manifolds for statistical sampling of multidimensional classes to generate a discrimination rule of guaranteed statistical accuracy.<sup>19-23</sup> Moreover, DR can generate a two- or three-dimensional map, which represents the prominent structures of the data and provides an embedded image of meaningful low-dimensional structures hidden in the high-dimensional information.<sup>19</sup> Pixel values in the embedded image are obtained based on distances over the manifold of pixel intensities in the higher dimension, and could have higher accuracy in detecting soft and hard boundaries between tissues, compared to the pixel intensities of a single image modality. To investigate the potential of DR methods for medical image segmentation, we developed and compared the performance of DR methods using ISOMAP, local linear embedding (LLE), and Diffusion Maps (DfM) on high-dimensional synthetic data sets and on a group of patients with breast cancer.

## II. METHODS AND MATERIALS

### II.A. Dimensionality reduction

To detect the underlying structure of high-dimensional data, such as that resulting from multiparametric MRI, we used dimensionality reduction methods. Dimensionality reduction (DR) is the mathematical mapping of high-dimensional data into a meaningful representation of the intrinsic dimensionality (lower dimensional representation) using either linear or nonlinear methods (described below). The intrinsic dimensionality of a data set is the lowest number of images or variables that can represent the true struc-

ture of the data. For example, T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI) are the lowest-dimension set of images that can represent anatomical tissue in the breast. Mathematically, a data set,  $X \subset \mathbb{R}^{D(x_1, x_2, \dots, x_n)=D(\text{Images})}$  where,  $x_1, x_2, \dots, x_n = \text{T1WI, T2WI, DWI, DCE MRI}$  or others and the data set has an intrinsic dimensionality,  $d < D$ , if  $X$  can be defined by  $d$  points or parameters that lie on a manifold.  $\mathbb{R}^D$  refers to  $D$  dimensional space with real numbers. Typically, we use a manifold learning method to determine points or locations within a dataset  $X$  (e.g., MRI, PET, etc.) lying on or near a manifold with intrinsic (lower) dimensionality,  $d$ , that is embedded in the higher dimensional space ( $D$ ). By definition, a manifold is a topological space that is locally Euclidean, i.e., around every point, there exists a neighborhood that is topologically the same as the open unit ball in Euclidean space. Indeed, any object that can be “charted” is a manifold.<sup>24</sup>

Dimensionality reduction methods map dataset  $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^{D(\text{images})}$  into a new dataset,  $Y = \{y_1, y_2, \dots, y_n\} \subset \mathbb{R}^d$ , with dimensionality  $d$ , while retaining the geometry of the data as much as possible. Generally, the geometry of the manifold and the intrinsic dimensionality  $d$  of the dataset  $X$  are not known. In recent years, a large number of methods for DR have been proposed, which belong to two groups: linear and nonlinear and are briefly discussed in this paper.

### II.B. Linear DR

Linear DR methods assume that the data lie on or near a linear subspace of some high-dimensional topological space. Some examples of linear DR methods are: Principal components analysis (PCA),<sup>25</sup> linear discriminant analysis (LDA),<sup>26</sup> and multidimensional scaling (MDS).<sup>27</sup>

#### II.B.1. PCA

Principal component analysis finds a lower dimensional subspace that best preserves the data variance, and where the variance in the data is maximal. In mathematical terms, PCA attempts to find a linear mapping,  $M$ , which maximizes the sum of the diagonal elements (trace) of the following matrix:

$$M^T * \Sigma * M, \quad (1)$$

under the constraint that  $|M| = 1$ , where  $\Sigma$  is the covariance matrix of the  $D$  dimensional data and  $X = [x_1, x_2, \dots, x_N]$  is zeroed mean. “\*” represents the multiplication of two matrices. Reports have shown that the linear map  $M$  could be estimated using  $d$  eigenvectors, i.e., PCA, of the covariance matrix of the data

$$\Sigma * V = \lambda V, \quad (2)$$

$\lambda$  is the eigenvalue corresponding to the eigenvector  $V$ . The data  $X$  now can be mapped to an embedding space by

$$Y = M * X, \quad (3)$$

where the first  $d$  largest PCAs are stored in the columns of matrix  $M$ . If the size of the  $\Sigma$  is high, the computation of the eigenvectors would be very time consuming. To solve this

problem, Partridge *et al.* proposed an approximation method, termed “Simple PCA,” which uses an iterative Hebbian approach to estimate the principal eigenvectors of  $\Sigma$ .<sup>28</sup> PCA has been applied in several applications for pattern recognition and the intrinsic dimension of the data is distributed on a linear manifold; however, if the data are not linear, then an overestimation of the dimensionality could occur or PCA could fail.<sup>29,30</sup>

**II.B.2. Multidimensional scaling (MDS)**

Classical MDS determines the subspace ( $Y \subset \mathbb{R}^d$ ) that best preserves the interpoint distances by minimizing the cost function of error between the pair-wise Euclidean distances in the low-dimensional and high-dimensional representation of the data. That is, given  $X \subset \mathbb{R}^{D(\text{images})}$ , MDS attempts to preserve the distances into lower dimensional space,  $Y \subset \mathbb{R}^d$ , so that inner products are optimally preserved.

The cost function is defined as

$$E(X, Y) = \sum_{ij} (\|x_i - x_j\| - \|y_i - y_j\|)^2, \tag{4}$$

where  $\|x_i - x_j\|$  and  $\|y_i - y_j\|$  are Euclidean distances between data points in the higher and lower dimensional space, respectively. Similar to PCA, the minimization can be performed using the eigen decomposition of a pairwise distance matrix as shown below

$$\begin{aligned} B &= \tau(D) = X^T * X, \\ B * V &= \lambda V, \\ M &= [V_1, V_2, \dots, V_d], \\ Y &= M * X, \end{aligned} \tag{5}$$

where  $B$  is the pair-wise distance matrix,  $V$  is the eigenvectors,  $M$  is the eigenvalues, and  $Y$ , the first  $d$  largest eigenvectors stored in the columns of matrix  $M$ , and the embedded data in the reduced dimension, respectively.

MDS has been applied in several applications in pattern recognition and data visualization.<sup>19–22,31,32</sup> We selected the most frequently used linear DR methods (PCA and MDS) to compare with nonlinear DR techniques in our application.

**II.C. Nonlinear dimension reduction**

Nonlinear techniques for DR do not rely on the linearity assumption for segmentation and, as a result, more complex embedding of high-dimensional data can be identified where linear methods often fail. There are a number of nonlinear techniques, such as Isomap,<sup>21,33</sup> locally linear embedding,<sup>22,34</sup> Kernel PCA,<sup>35</sup> diffusion maps,<sup>36,37</sup> Laplacian Eigenmaps,<sup>32</sup> and other techniques.<sup>38,39</sup>

**II.C.1. Isometric feature mapping (Isomap)**

As mentioned before, DR technique map dataset  $X$  into a new dataset,  $Y$ , with dimensionality  $d$ , while retaining the geometry of the data as much as possible. If the high-dimensional data lies on or near a curved smooth manifold, Euclidean distance does not take into account the distribution

of the neighboring data points and might consider two data points as close, whereas their distance over the manifold is much larger than the typical interpoint distance. Isomap overcomes this problem by preserving pairwise geodesic (or curvilinear) distances between data points using a neighborhood graph.<sup>21</sup> By definition, Geodesic distance (GD) is the distance between two points measured over the manifold and generally estimated using Dijkstra’s shortest-path algorithm.<sup>40</sup>

GDs between the data points are computed by constructing a neighborhood graph,  $G$  (every data point  $x_i$  is connected to its  $k$  nearest neighbors,  $x_{ij}$ ), where, the GDs between all data points form a pair-wise GD matrix. The low-dimensional space  $Y$  is then computed by applying multidimensional scaling (MDS) while retaining the GD pairwise distances between the data points as much as possible. To accomplish this, the error between the pair-wise distances in the low-dimensional and high-dimensional representation of the data are minimized using the equation below

$$\sum (\|x_i - x_{ij}\| - \|y_i - y_{ij}\|)^2. \tag{6}$$

This minimization is performed using various methods, such as the eigen-decomposition of the pair-wise distance matrix, the conjugate gradient method, or a pseudo-Newton method.<sup>41</sup> We used the eigen decomposition for our implementation and Fig. 1 demonstrates the steps used to create the Isomap embedding.

**II.C.2. Locally linear embedding (LLE)**

In contrast to Isomap, LLE preserves the local properties of the data, which allow for successful embedding of non-convex manifolds. LLE assumes that the global manifold can be reconstructed by “local” or small connecting regions (manifolds) that are overlapped. That is, if the neighborhoods are small, the manifolds are approximately linear. LLE performs linearization to reconstruct the local properties of the data by using a weighted summation of the  $k$  nearest neighbors for each data point. This approach of linear mapping of the hyperplane to a space of lower dimensionality preserves the reconstruction weights. Thus, this allows the reconstruction weights,  $W_i$ , to reconstruct data point  $y_i$  from its neighbors in the reduced dimension.

Therefore, to find the reduced ( $d$ ) dimensional data representation  $Y$ , the following cost function is minimized for each point  $x_i$ :

$$\varepsilon(W) = \sum_{i=1}^n \left\| x_i - \sum_{j=1}^k w_{ij} x_{ij} \right\|^2, \tag{7}$$

subject to two constraints,  $\sum_{j=1}^k w_{ij} = 1$  and  $w_{ij} = 0$  when  $x_j \notin \mathbb{R}^{D(\text{images})}$ . Where  $X$  is input data,  $n$  is the number of points, and  $k$  is the neighborhood size. The optimal weights matrix  $W$  ( $n \times k$ ), subject to these constraints, is found by solving a least-squares problem.<sup>22</sup>

Then, we compute the embedding data ( $Y$ ) by calculating the eigenvectors corresponding to the smallest  $d$  nonzero eigenvalues of the matrix

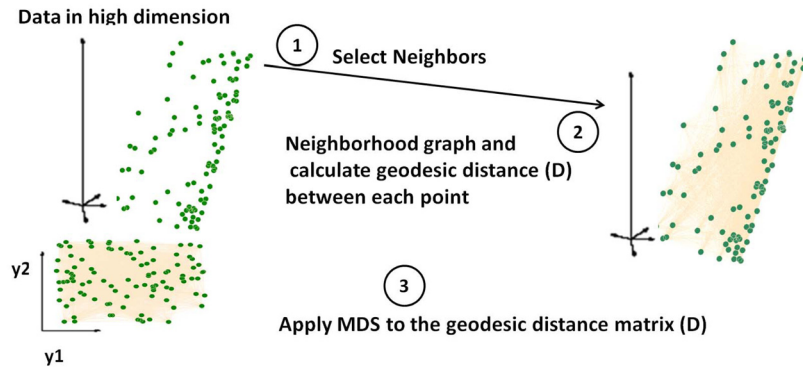


FIG. 1. The steps of Isomap algorithm used for mapping data into lower dimension d: (1) find  $K$  nearest neighbors for each data point  $X_i$ ; (2) calculate pairwise geodesic distance matrix and reconstruct neighborhood graph using Dijkstra search algorithm; (3) apply Multidimensional scaling (MDS) on the reconstructed neighborhood graph (matrix  $D$ ) to compute the low-dimensional embedding vectors.

$$\begin{aligned}
 E &= (I - W)^T * (I - W), \\
 E * V &= \lambda V, \\
 M &= [V_1, V_2, \dots, V_d], \\
 Y &= M * X,
 \end{aligned}
 \tag{8}$$

where  $I$  is the identity matrix and  $W$  is the weights matrix ( $n \times k$ ). Figure 2 shows the steps for LLE.

**II.C.3. Diffusion maps (DfM)**

Diffusion maps find the subspace that best preserves the diffusion interpoint distances based on defining a Markov random walk on a graph of the data called a Laplacian graph.<sup>36,37</sup> These maps use a Gaussian kernel function to estimate the weights ( $K$ ) of the edges in the graph

$$K_{ij} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad 1 \leq i, j \leq L,
 \tag{9}$$

where  $L$  equals the number of multidimensional points and  $\sigma$  is the free parameter, sigma. In the next step, the matrix  $K$  is normalized such that its rows add up to 1

$$p_{ij}^{(t)} = \frac{K_{ij}}{\sum_{n=1}^L K_{in}},
 \tag{10}$$

where  $P$  represents the forward transition probability of  $t$ -time steps of a random walk from one data point to another data point. Finally, the diffusion distance is defined as

$$\begin{aligned}
 D_{ij}^{(2)} &= \sum_{r=1}^L \frac{(p_{ir}^{(t)} - p_{jr}^{(t)})^2}{\Psi(x_r)}, \\
 \Psi(x_m) &= \frac{\sum_{j=1}^L p_{jm}}{\sum_{k=1}^L \sum_{j=1}^L p_{jk}}
 \end{aligned}
 \tag{11}$$

Here, the high density portions of the graph defined by the diffusion distance have more weight, and pairs of data points with a high forward transition probability have a smaller diffusion distance. The diffusion distance is more robust to noise than the geodesic distance because it uses several paths throughout the graph to obtain the embedded image. Based

**LLE Pipeline**

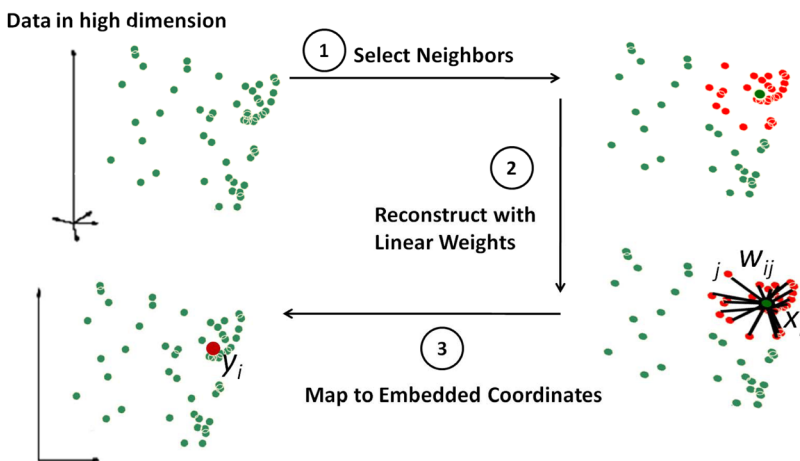


FIG. 2. Steps of Locally Linear Embedding (LLE) algorithm used for mapping data into lower dimension d (1) search for  $K$  nearest neighbors for each data point  $X_i$ ; (2) solve the constrained least-squares problem in Eq. (7) to obtain weights  $W_{ij}$  that best linearly reconstruct data point  $X_i$  from its  $K$  neighbors, (3) Compute the low-dimensional embedding vectors  $Y_i$  best reconstructed by  $W_{ij}$ .



on spectral theory about the random walk, the embedded image of the intrinsic dimensional representation  $Y$  can be obtained using the  $d$  nontrivial eigenvectors of the distance matrix  $D$

$$Y : x \rightarrow \{\lambda_2 V_2, \dots, \lambda_d V_d\}. \tag{12}$$

The DfM graph is a fully connected, eigenvector  $v_1$  of the largest eigenvalue ( $\lambda_1 = 1$ ) that is discarded, and the remaining eigenvectors are normalized by their corresponding eigenvalues.

**II.D. DR-based manifold unfolding**

As mentioned above, the DR methods map high-dimensional data to a lower dimension while preserving data structure. This property of DR can be used to unfold  $N$ -dimensional manifolds into a lower dimension. To illustrate DR capabilities, the advantages and disadvantages of each linear and nonlinear DR method for manifold unfolding, three well-known synthetic datasets—the Swiss-roll, 3D-clusters, and sparse data sets—were used to test each of the methods and determine which one(s) to use on the clinical data.

**II.E. DR-based multidimensional image data integration**

By using DR and its manifold unfolding ability, we can integrate multidimensional image data for visualization. Our proposed DR scheme is shown in Fig. 3(a) and includes two stages: (1) preprocessing (enhancement, coregistration, and image resizing); and (2) dimensionality reduction to obtain an embedded image for soft tissue segmentation. Figure 3(b) shows the steps for automated extraction of tumor boundaries [A and B in Fig. 3(b)] from the embedded and postcontrast images. The same procedure is used for fatty and glandular breast tissue. To test for overlap and similarity, we used the Dice similarity (DS) metric.<sup>42</sup> The DS measure is defined as

$$DS = 2 \frac{A \cap B}{A + B} = 2 \frac{TP}{2TP + FP + FN}, \tag{13}$$

where A is the tissue boundary in the embedded image, B is the tissue boundary in the postcontrast image, TP is the true positive, FP is the false positive, and FN is the false negative.

**II.F. Preprocessing steps**

**II.F.1. Image enhancement**

Since artifacts and noise can degrade radiological images and may make identification and diagnosis difficult. In particular, MRI images can have large  $B_1$  inhomogeneities that can obscure anatomical structures. To estimate true signal and correct  $B_1$ -inhomogeneity effects, we used a modified version of the “local entropy minimization with a bicubic spline model (LEMS),” developed by Salvado *et al.*<sup>43</sup> LEMS is based on modeling the bias field,  $\beta$ , as a bicubic spline and

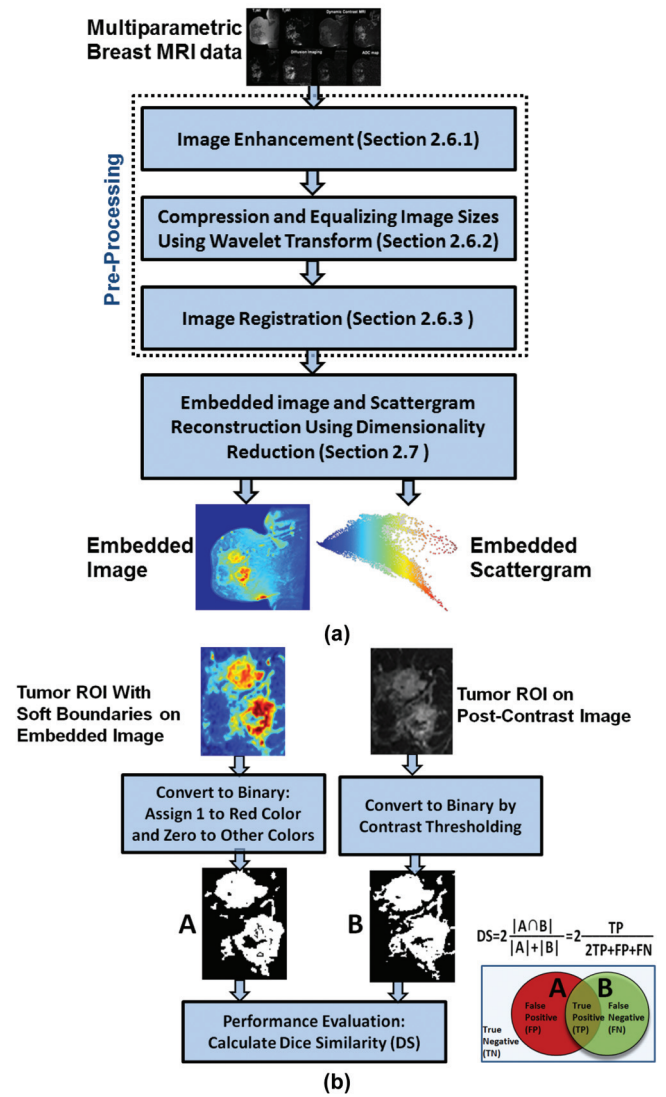


Fig. 3. (a) Proposed schema for multidimensional image data integration using the Non Linear Dimensional Reduction (NLDR) methods to construct the embedded image and soft tissue segmentation; (b) Proposed schema for extraction of tumor hard boundaries for direct comparison with the postcontrast image as the current standard technique. The same procedure is used for the other breast tissue (fatty and glandular).

the RF coil geometry as a sufficiently close rectangular grid of knots scattered across the image ( $X$ ). Initialization of LEMS begins with a fourth-order polynomial function estimation of the tissue voxel, where the background is excluded. Optimization of the bicubic spline model is performed in piecewise manner. LEMS first identifies a region within the image with the highest SNR and assigns it to Knot  $K_1$  to ensure that a good local estimate of the field  $\beta$  is obtained. LEMS then adjusts the signal at  $K_1$ , based on an  $8 \times 8$  neighborhood of knots, to locally minimize the entropy of  $X$  in  $K_1$  and its neighbors. LEMS repeats the same routine for other knots with high SNR and minimizes the entropy within the corresponding knot ( $K_i$ ) neighbor, as well as in prior knots ( $\{K_j\}_{1 \leq j < i}$ ) until either the maximum number of iterations are reached or the knot entropies do not change significantly.

**II.F.2. Equalizing image sizes using wavelet transform**

As image sizes can vary, there is a need for a method to equalize image sizes with less loss of spatial and textural information. If any image is smaller than the desired size (N), data interpolation should be used to upsize the image to N. There are several well-known interpolation methods, such as nearest-neighbor, bilinear, and bicubic. However, if any of the images are larger than the desired size (N), the image should be downsized, which causes loss of spatial and textural information. To avoid this problem, a powerful multiresolution analysis using wavelet transforms can be used.<sup>44,45</sup>

Wavelets are mathematical functions that decompose data into different frequency components, thereby facilitating the study of each component with a resolution matched to its scale.<sup>44</sup> The continuous wavelet transform of a square-integrable function,  $f(t)$ , is defined as

$$wf(s, t) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right), \tag{14}$$

where  $s$  and  $t$  are the scale (or frequency) and time variable, respectively. The function  $\psi(t)$ , is called a wavelet and must satisfy the admissibility condition, that is, it must be a zero-meant and square-integrable function.

In practical applications, the parameters  $s$  and  $t$  must be discretized. The simplest method is dyadic. By using this method, the fast wavelet transform (FWT) is defined as

$$wf[n, 2^j] = \sum_{m=0}^{N-1} f[m] \psi_j[m - n] = f[n] * \psi_j[n], \tag{15}$$

where  $\psi_j[n] = \psi(\frac{n}{2^j})$  and  $f[n]$  is a sequence with a length of  $N$  (discrete time function), and the  $*$  sign represents a circular convolution.

To implement a fast-computing transform, the FWT algorithm was used. At each level of decomposition, the FWT algorithm filters data with two filters, called  $h[n]$  and  $g[n]$ . The filter  $h[n]$ , a conjugate mirror filter, is a low-pass filter, and thus, only low-frequency (coarse) components can pass through it; conversely,  $g[n]$  is a high-pass filter to pass high frequency (detail) components. The dyadic wavelet representation of signal  $a_0$  is defined as the set of wavelet coefficients up to a scale  $J$  ( $=\log_2 N$ ), plus the remaining low-frequency information [ $1 \leq j \leq J$ ].<sup>44,45</sup>

For image processing, we first applied the 1D FWT to the rows of the image. Then, we applied the same transform to the columns and diagonals of each component. Therefore, we obtained three high-pass (detail) components corresponding to vertical, horizontal, and diagonal, and one approximation (coarse) component. Figure 4 outlines with block diagrams the procedure using the 1D and 2D FWT. By using the inverse FWT (IFWT), we could reconstruct the original image in a different scale and resolution (see Fig. 4) and when we reformatted, there was no loss of resolution.

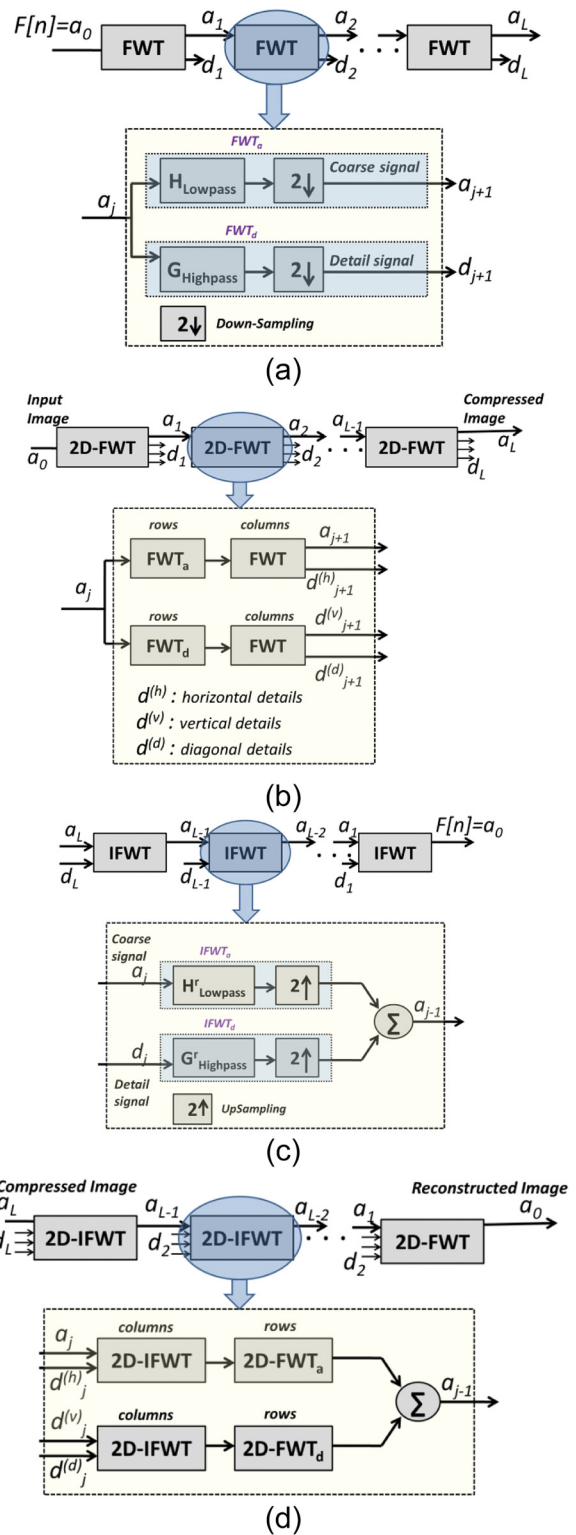


Fig. 4. Signal (a) and image (b) decomposition using Fast Wavelet Transform (FWT); Signal (c) and Image (d) reconstruction using Inverse FWT (IFWT).  $a_0$  is original signal (image).  $a_j$  and  $d_j$ , respectively, are the approximation (coarse or low pass) and detail (high pass) components corresponding to vertical, horizontal, and diagonal at decomposition level  $j$ .

**II.F.3. Coregistration methods**

For coregistration of the different modalities/parameters, we used a modified nonrigid registration technique developed by Periaswamy and Farid.<sup>46</sup> The method is based on both

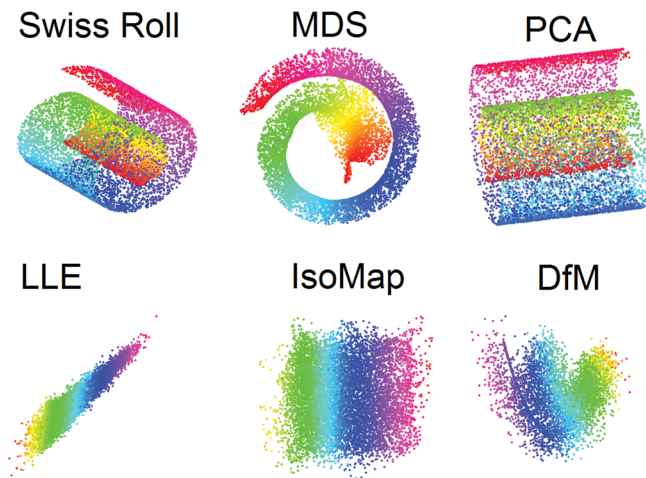


FIG. 5. Dimension reduction of the Swiss Roll (a) from 3 to 2 dimensions using MDS, PCA, LLE, Isomap and diffusion maps (DfM). Neighborhood size for LLE and Isomap, respectively, were 5 and 10. Sigma for DfM was 0.2. The linear methods (MDS and PCA) both failed to unfold the Swiss Roll in the reduced dimension, while all nonlinear methods (LLE, Isomap, and DfM) were able to unfold the data and preserve structure. The best result was obtained by Isomap.

geometric (motion) and intensity (contrast and brightness) transformations. The geometric model assumes a locally affine and globally smooth transformation. The local affine model is based on motion estimation. Model parameters are

estimated using an iterative scheme that uses nonlinear least-squares optimization to compute model parameters.<sup>47</sup> In order to deal with the large amount of motion and capture both large and small-scale transformations, a Gaussian pyramid was built for both source and target images. Model parameters were initially estimated at the coarsest level and were used to warp the source image at the next level of the pyramid and update the model parameters at each level of the pyramid. This multi-scale approach enabled registration of different images, such as T1WI, T2WI, and DWI. A set of derivative filters designed for multidimensional differentiation was used to decrease noise and to improve the resultant registration.

### II.G. Embedded image and scattergram reconstruction

After the preprocessing (e.g., image enhancement, wavelet-based compression, and image registration), the radiological images were used as inputs for DR methods. The embedded image was constructed by projecting the features (image intensities) from  $N$ -dimensional space to a one-dimensional embedding space, using the results from the DR methods [see Fig. 3(a)]. The resultant embedding points then reconstructed the embedded image by reforming the embedded data matrix from the size  $L \times 1$  to  $\frac{L}{2} \times \frac{L}{2}$ . If we map data from  $N$ -dimensional space into two-dimensional space, we

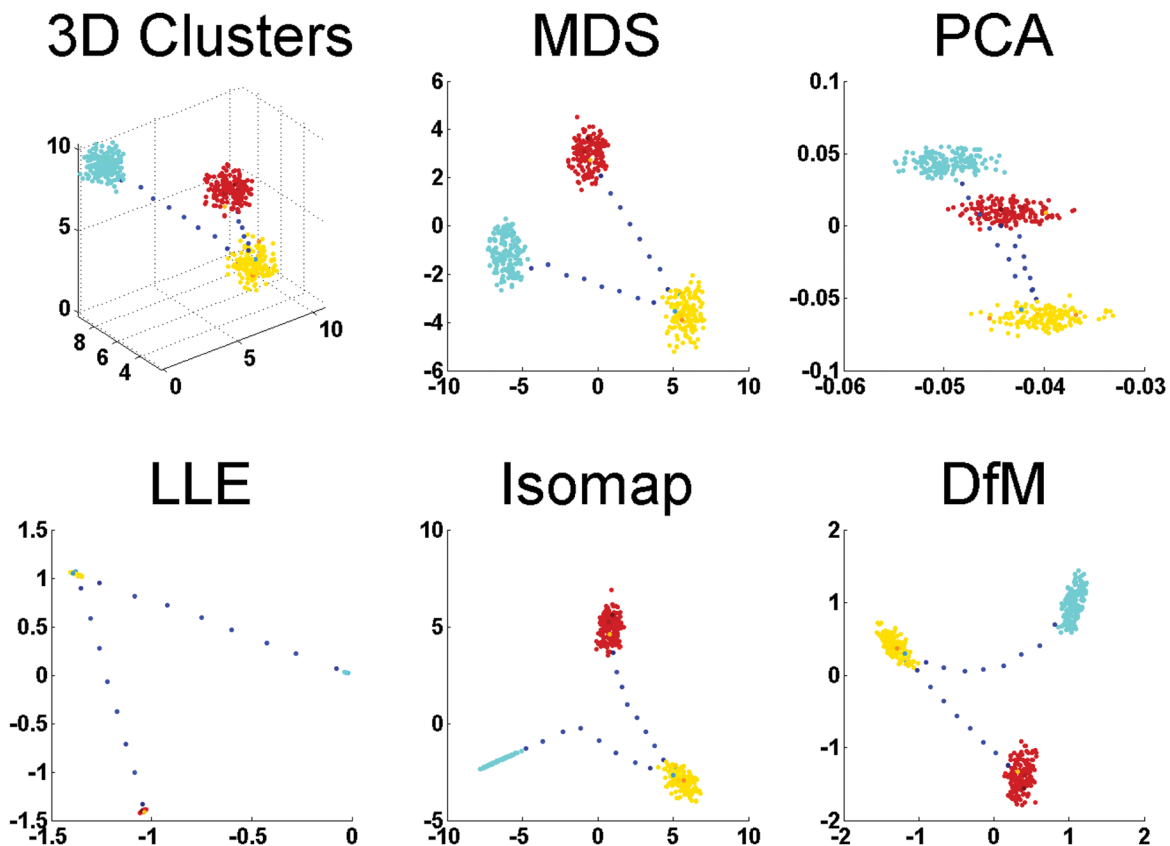


FIG. 6. Demonstration of dimension reduction of 3D clusters (a) from 3 to 2 dimensions using Multidimensional Scaling (MDS), Principal Component Analysis (PCA), Locally Linear Embedding (LLE), Isomap and diffusion maps (DfM). Sigma for DfM was 0.2. Neighborhood size for LLE and Isomap, respectively, were 5 and 10. In this example, all methods except LLE were able to preserve clusters in the reduced dimension. Isomap was not able to fully preserve structure of for the cyan color cluster in the embedding space. LLE was not able to preserve structure of all three clusters and converted the clusters to points in the embedding space.



will obtain an unfolded version of the data manifold with different clusters.<sup>21,22</sup>

## II.H. Multiparametric breast MRI segmentation

### II.H.1. Clinical subjects

Twenty-five women were scanned as part of a breast research study. All subjects signed written, informed consent, and the study was approved by the local IRB. Twenty-three patients had breast tumors (19 malignant and 4 benign) and two patients had no masses.

### II.H.2. Multiparametric MRI imaging protocol

MRI scans were performed on a 3T magnet, using a dedicated phased array breast coil with the patient lying prone with the breast in a holder to reduce motion. MRI sequences were: fat suppressed (FS) T2WI spin echo (TR/TE = 5700/102) and fast spoiled gradient echo (FSPGR) T1WI (TR/TE = 200/4.4, 2562, slice thickness, 4 mm, 1 mm gap); diffusion-weighted (TR/TE = 5000/90 ms,  $b = 0,500\text{--}1000,1282,ST = 6$  mm); and finally, precontrast and postcontrast images FSPGR T1WI (TR/TE = 20/4, matrix = 5122, slice thickness, 3 mm) were obtained after intravenous administration of a GdDTPA contrast agent [0.2 ml/kg(0.1 mmol/kg)]. The contrast agent was injected over 10 s, with MRI imaging beginning immediately after completion of the injection and the acquisition of at least 14 phases. The contrast bolus was followed by a 20cc saline flush. The DCE protocol included 2 min of high temporal resolution (15 s per acquisition) imaging to capture the wash-in phase of contrast enhancement. A high spatial resolution scan for 2 min then followed, with additional high temporal resolution images (15 s per acquisition) for an additional 2 min to characterize the wash-out slope of the kinetic curve. Total scan time for the entire protocol was less than 45 min.

### II.H.3. Multiparametric breast MRI segmentation and Comparison

The nonlinear dimensionality reduction (NLDR) methods and the embedded image and scattergram were applied to the breast MRI data. To differentiate tissue types and soft boundaries between them, a continuous RGB color code can be assigned to the embedded image. To perform a quantitative comparison between the embedded image and ground-truth, we used similarity measures based on regional overlap between hard boundaries. Ground truth was based on the current clinical standard in breast imaging, which is the post contrast MRI. We used the Dice similarity index, designed to find the overlapping regions between two objects (see Fig. 3). In our application, A and B are lesion areas obtained by ground truth (postcontrast image) and the embedded image, respectively. The lesion area for the postcontrast was obtained by thresholding of the contrast image. The threshold was obtained by evaluating the postcontrast MR image histogram, and using the mean and a 95% confidence interval. However, the embedded image returned a fuzzy boundary with an RGB color code and the hard boundary was obtained by converting the embedded image to a binary

image and assigning a “1” to the red color-coded pixels and a zero to the pixels with other colors. If A and B have full overlap, then the  $DS = 1.0$ . But if A and B do not intersect, then the  $DS = 0$ . To evaluate the DR methods, we applied them to segment and visualize breast tumors using multiparametric MRI data in a small series of patients and volunteers to discern the properties of each MR sequence in breast lesion identification, no classification was performed and under investigation.

## III. RESULTS

### III.A. Synthetic data

Figure 5 shows typical results for DR methods applied to the Swiss Roll manifold. Both linear methods (PCA and MDS) failed to unfold the manifold and preserve structure when mapping from 3 to 2 dimensions. The nonlinear method, Isomap, was able to unfold the Swiss roll with good results: but, both LEE and DfM methods were able, in part, to unfold the Swiss roll with most of the structure retained. In the 3D point cluster model, most of the linear and nonlinear DR methods, with the exception of LLE, were able to maintain the 3D cluster structure when mapped to the lower dimension, as demonstrated in Fig. 6. Notably, LLE converted each cluster to points. Finally, the sparse (nonuniform) manifold data revealed the main difference and the power of nonlinear DR methods compared to linear methods, as shown in Fig. 7. The linear methods failed to unfold the sparse data set, whereas, the nonlinear DR methods were able to preserve, in most cases, the structure of the sparse manifold. In particular, LLE performed the best, but Isomap and DfM were able to partially preserve the structure.

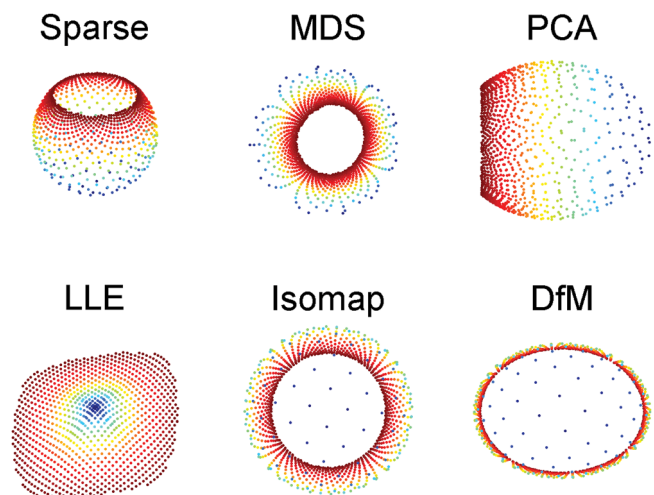


Fig. 7. Embedding of sparse data (a) from 3 to 2 dimensions using MDS, PCA, LLE, Isomap and diffusion maps (DfM). Neighborhood size for LLE and Isomap, respectively, were 5 and 10. Sigma for DfM was 0.2. Both linear methods (MDS and PCA) failed to preserve the sparse data structure in the reduced dimension. DfM was able to fully preserve the sparse data pattern in the embedding space. LLE successfully mapped the sparse data structure to the embedded space. Isomap also was able to preserve most of the data structure but was unable to correctly map the all the blue color to the embedding space.



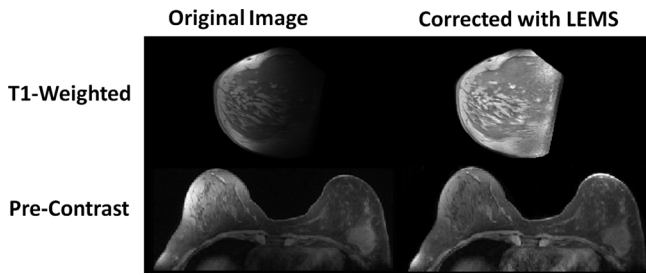


FIG. 8. Correction of  $B_1$  inhomogeneity in the MRI data with the local entropy minimization with a bicubic spline model (LEMS) method: Shown are original and corrected images, respectively, for T1-weighted images. After correction, better visualization of breast tissues is noted and they are isointense across the image, compared with the images on the left.

**III.B. Multiparametric breast MRI preprocessing steps**

Based on our encouraging results from the synthetic data, we applied the nonlinear methods to the multiparametric breast data using our hybrid integration scheme outlined in Fig. 3. The preprocessing of the T1WI (sagittal and axial) before and after  $B_1$  inhomogeneity correction is shown in Fig. 8. The coefficients of variation (COV) for fatty tissue, as an example, before applying LEMS, were 98.6 and 67.2, respectively, for T1WI and precontrast MRI images. The COV improved to 49.3 and 39.2, after correction.

Figure 9 demonstrates the wavelet compression and decompression methods applied to a DWI ( $b = 500$ ) image.

The original size was  $256 \times 256$ , and was resized (compressed and decompressed) to  $64 \times 64$ . The reconstructed image was identical to the original image with no errors. Finally, registration of the breast MRI was achieved using a locally affine model, and typical results are shown in Fig. 10. The mean square error between the predicted intensity map of the reference image and original reference image was 0.0567.

**III.C. Multiparametric breast MRI segmentation**

Figures 11 and 12 demonstrate the power of using nonlinear embedding of multiparametric breast MRI in several dimensions, for example, in Fig. 11, there are nine dimensions: T1WI; T2WI; precontrast; postcontrast; DWI (4-b values) and ADC maps into a single image that provided a quantitative map of the different tissue types for abnormal (malignant) and normal tissue. All images were resized to  $256 \times 256$  using the methods outlined in Fig. 4. After scaling the intensity of the embedded images (range of 0–1000), the lesion tissue appeared red, with normal fatty tissue appearing blue. The embedded scattergram was useful in classifying the different tissue types (see Figs. 11 and 12). The Dice similarity index between the lesion contours defined by the embedded image demonstrated excellent overlap ( $>80\%$ ) with the DCE-MRI-defined lesion. Table I summarizes the Dice similarity index on 25 patients (19 malignant and four benign cases, and two cases with no masses).

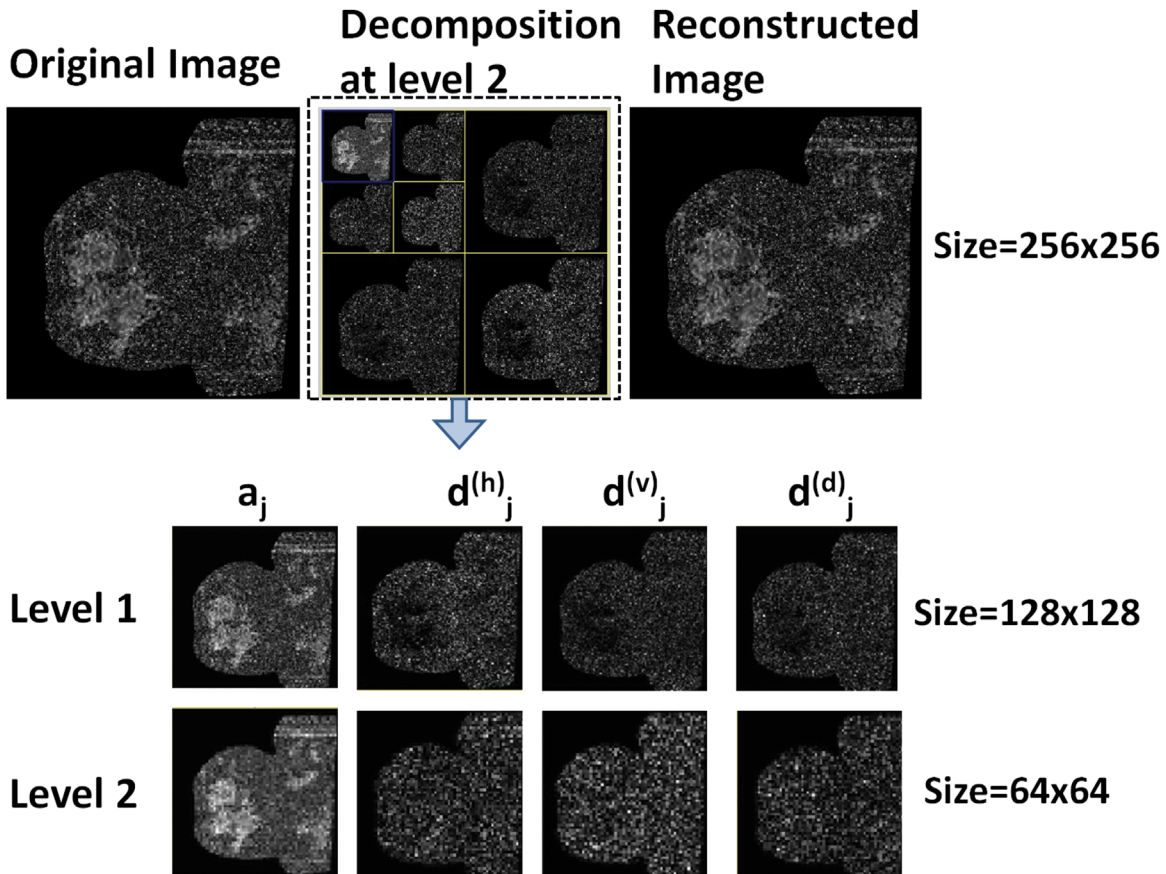


FIG. 9. Typical diffusion-weighted image ( $b = 500$ ) in the original size  $256 \times 256$ , after compression ( $64 \times 64$ ) and decompression (reconstructed image:  $256 \times 256$ ). For compression, 2D biorthogonal spline wavelets were used.  $d_j^{(h)}$ ,  $d_j^{(v)}$  and  $d_j^{(d)}$ , respectively, are detail components corresponding to vertical, horizontal, and diagonal.  $a_j$ , is the approximation (coarse) component at decomposition level  $j$ .

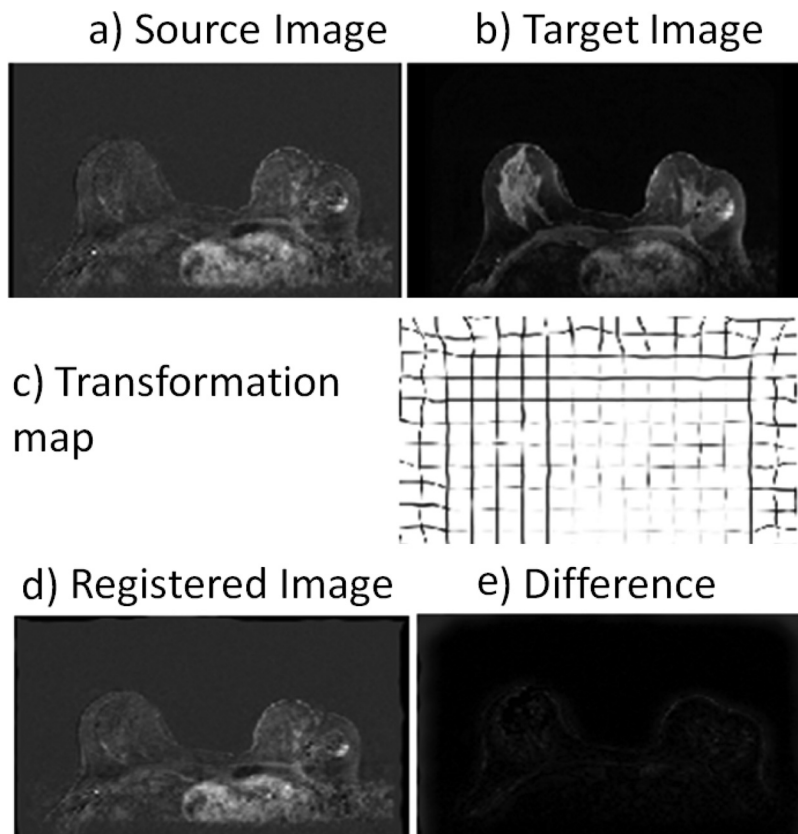


FIG. 10. Demonstration of the coregistration of the MRI data using a locally affine model: (a) the T2WI source image; (b) the T1WI reference image; (c) warping map; (d) the final coregistered image and (e) the difference image obtained from subtracting the reference image from the registered image.

Comparison of the computational loads for the NLDR methods indicated that the load for Isomap was very high compared to DfM and LLE. For example, for breast MRI images with a matrix size of  $256 \times 256$ , the computation times for DfM, Isomap, and LLE were 19.2, 508.4, and 78.6 s, respectively, using an Intel quad-core with 8 Gb RAM and a Nvidia Quadro 4000 (256 cores, 2 GB) video card. To determine the robustness of the NLDR methods to their input parameters, we varied the different input parameters. These parameters are defined in Isomap and LLE as the neighborhood size [ $K$ ], and defined in DfM as the sigma parameter. Figure 13 shows how the Dice similarity index changed with each of the NLDR parameter [neighborhood size [ $K$ ] in Isomap and LLE; sigma in DfM] variations. As shown, the neighborhood size  $K$  for Isomap should be chosen to be greater than 10, and, for LLE, it should vary from 20 to 60. However, the DfM is very sensitive to sigma in the Gaussian kernel, and, based on our investigations, sigma should range between 100 and 1000, but a sigma of at least 100 would be sufficient for complex data, such as MRI.

#### IV. DISCUSSION

In this paper, we developed a novel hybrid scheme using NLDR methods to integrate multiple breast MRI data into a single embedded image. The resultant embedded image enabled the visualization and segmentation of breast lesions from the adjacent normal breast tissue with excellent overlap and similarity. To determine which DR method to use, we compared the performance of both linear and nonlinear DR

methods using synthetic and multiparametric breast MRI. In the synthetic data, the three nonlinear DR methods, namely DfM, Isomap, and LLE, outperformed the linear methods (PCA and MDS) in all categories. The NLDR methods, in general, were able to segment and visualize the underlying structure of the Swiss Roll, point clouds, and, more importantly, sparse data. This is important because the sparse data represents a real-world scenario and the NLDR methods were able to unfold the underlying structure. This led us to choose the NLDR methods for the breast data. Multiparametric breast data represent complex high-dimensional data, and no single parameter conveys all the necessary information. Integration of the breast parameters is required and our methods provide an opportunity to achieve such integration. Our results demonstrate that when the NLDR methods are applied to breast MRI data, the DR methods were able segment the lesion and provide excellent visualization of the tissue structure.

This mapping of high-dimensional data to a lower dimension provides a mechanism by which to explore the underlying contribution of each MR parameter to the final output image. Indeed, each one of these methods is designed to preserve different data structures when mapping from higher dimensions to lower dimensions. In particular, DfM and Isomap primarily emphasize the global structure within the multidimensional feature space and are less sensitive to variations in local structure. However, LLE has greater sensitivity to the variations in local structure and is less sensitive to variations in global structure. In general, DfM and Isomap returned images similar to the embedded images from LLE,

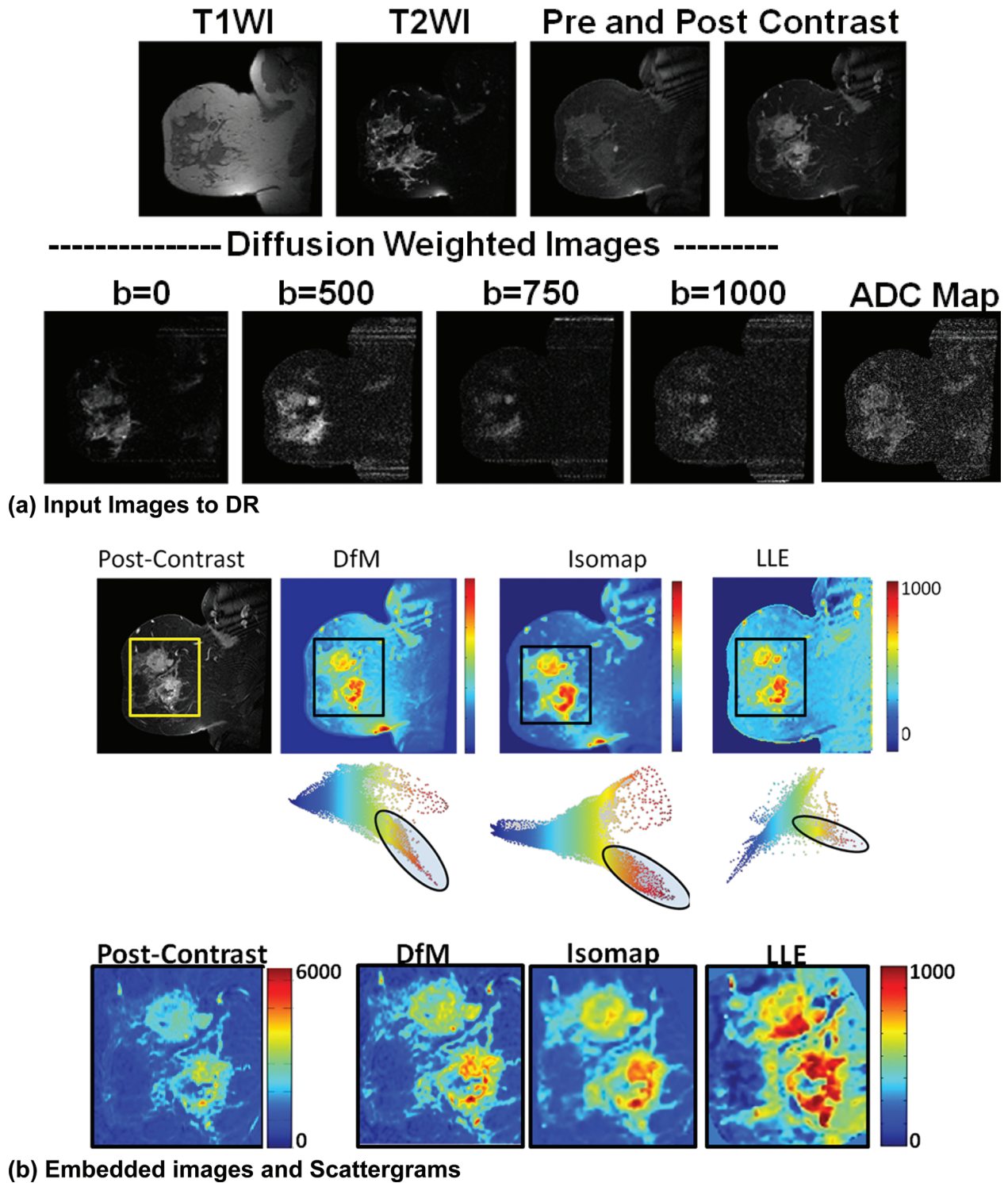


FIG. 11. (a) Typical multiparametric MRI data and (b) resulting embedded images and scattergrams for a malignant breast case from three NLDR algorithms Diffusion Maps (DfM), Isometric feature mapping (ISOMAP), Locally Linear Embedding (LLE). The lower scattergram shown is derived from Isomap. Clear demarcation of the lesion and surrounding breast are shown.

using the breast MRI data. Therefore, each of the methods was able to differentiate the breast lesion cluster within the scattergram, suggesting that, for computer-assisted diagnosis, either of the three DR methods (DfM and Isomap for global structures and LLE for local structures) could be applied to better assist the radiologist in decision-making. This was confirmed by the congruence between each embed-

ded image and breast DCE MRI with the Dice similarity metric. The Dice similarity metric showed excellent results with a little variation in the segmented lesion areas between the embedded and DCE MRI. One potential advantage of using an embedded image created from the breast MRI would be to create tissue masks for automatic overlay onto functional MRI parameters, such as ADC or sodium maps to



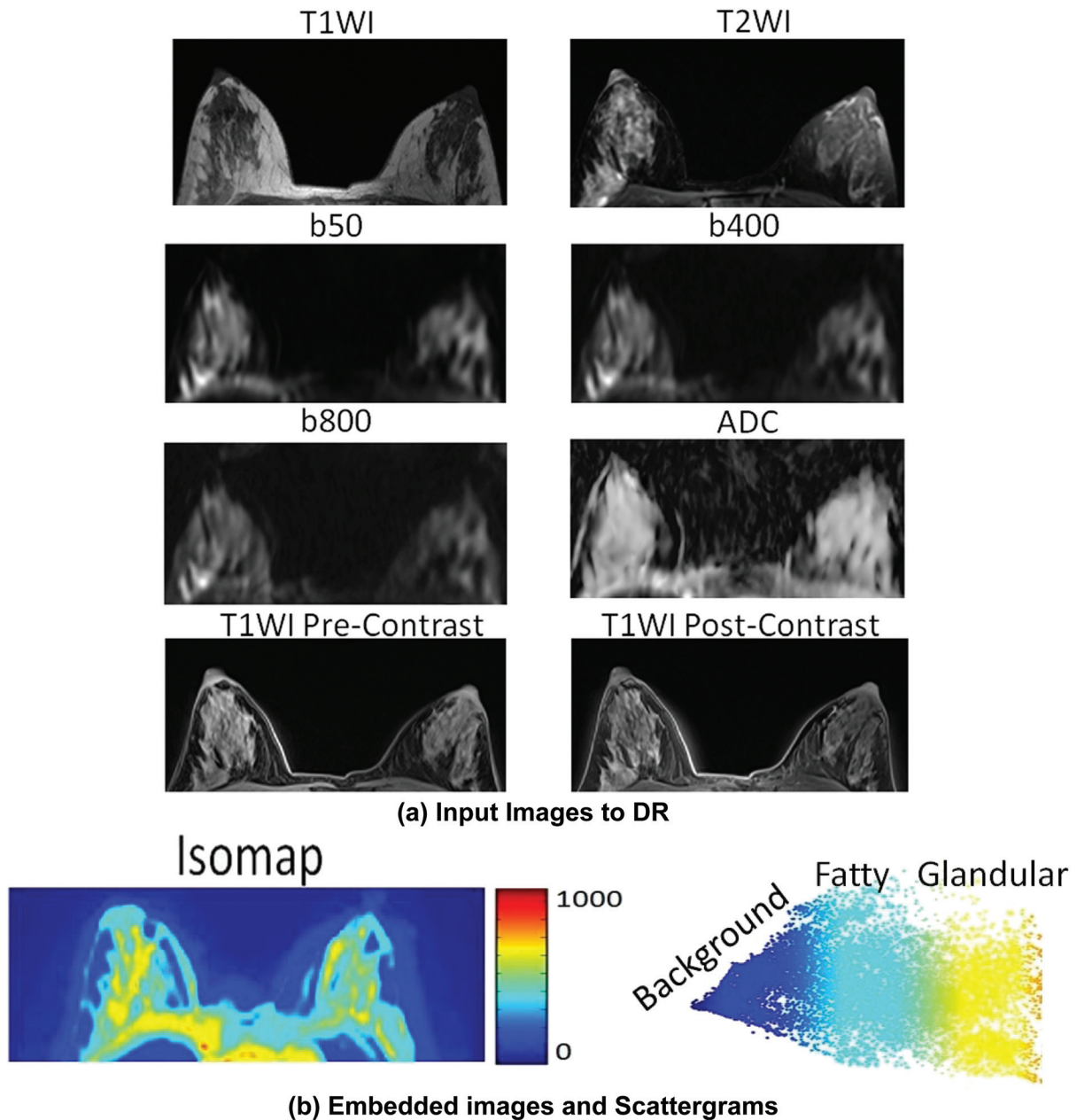


FIG. 12. (a) Typical axial multiparametric MRI data from a patient with no breast lesion. (b) resulting embedded images and scattergrams demonstrating the separation of fatty and glandular tissue in the embedded image, the scattergram shown is derived from Isometric feature mapping (ISOMAP).

develop an automatic quantification system that could be used to monitor treatment response.<sup>12</sup>

Other applications for DR methods with MRI data that have been published include the differentiation of benign

TABLE I. Dice similarity metric.

	Benign (4 cases)			Malignant (19 cases)		
	DfM (%)	Isomap (%)	LLE (%)	DfM (%)	Isomap (%)	LLE (%)
Mean	81.18	80.61	81.49	87.40	86.70	85.50
Median	82.08	79.97	82.74	86.72	86.15	86.31
SD	6.35	8.51	6.91	5.81	6.83	8.57

Note: 2 cases did not have masses and are not shown. Isometric feature mapping (ISOMAP), Locally Linear Embedding (LEE) Diffusion Mapping (DFM).

and malignant tissue on magnetic resonance spectroscopy (MRS) data. Tiwari *et al.*<sup>48</sup> used DR methods to separate the different peaks of metabolites (choline, citrate, etc.) into different classes and then overlaid the results onto the T2WI of the prostate with increased sensitivity and specificity. However, these investigators only utilized the frequency domain of the spectra and not the entire dimensionality of the prostate data set, that is, combined MRI and MRS. Nevertheless, they did demonstrate the power of DR methods consistent with our results. Other recent applications for DR methods include use on diffusion tensor magnetic resonance images (DT-MRI). DT-MRI cannot be analyzed by commonly used linear methods, due to the inherent nonlinearity of tensors, which are restricted to a nonlinear submanifold of the space in which they are defined. To overcome this problem, Verma



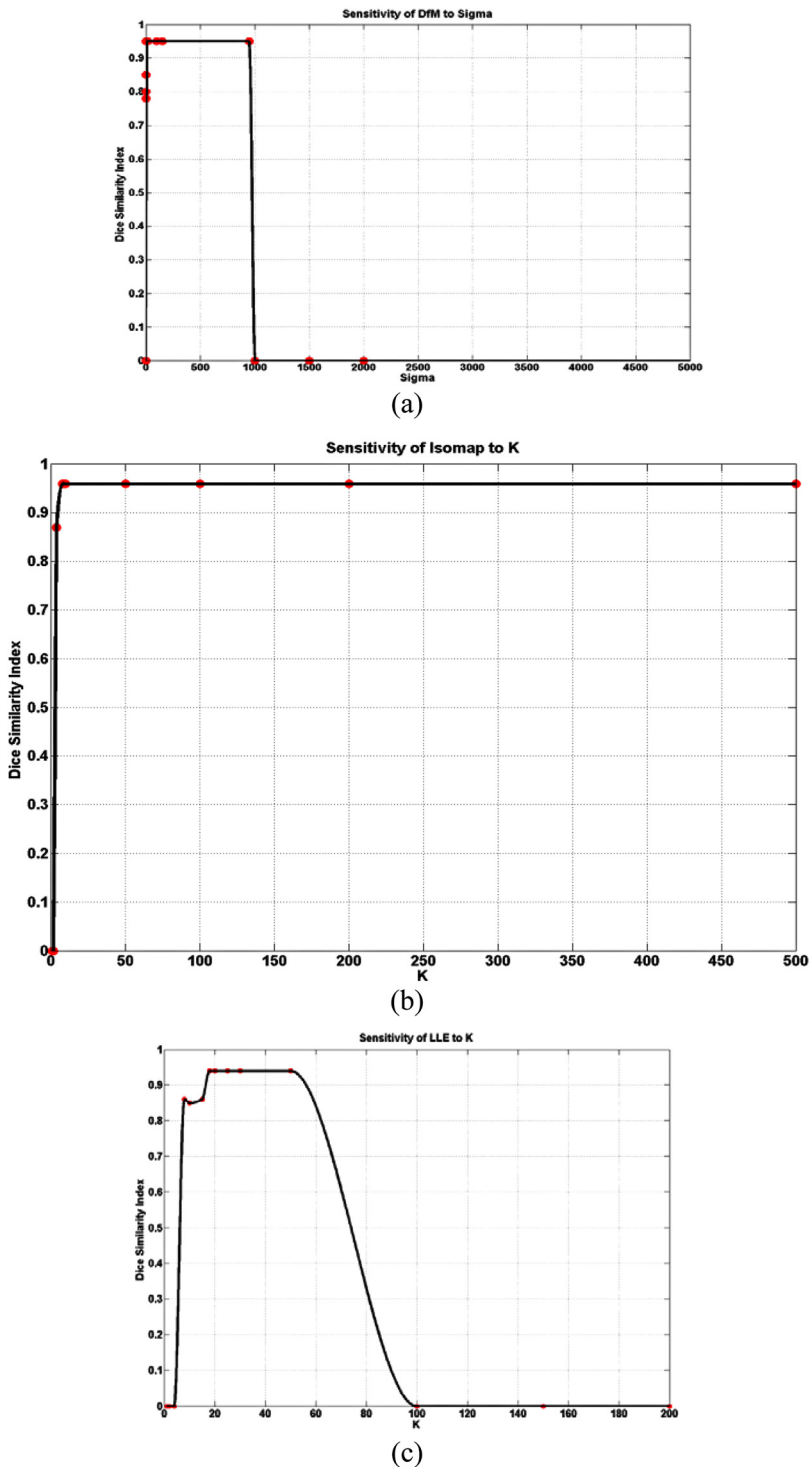


FIG. 13. Demonstration of the sensitivity of NLDR methods to control parameters: (a) effects of Sigma value on the dice similarity index between Diffusion Maps (DfM) based embedded image and the postcontrast; (b) and (c) effects of neighborhood size ( $K$ ) on the dice similarity index between the embedded image and the postcontrast, respectively, for Isomap and LLE. Example input MRI data for the NLDR methods are shown in Fig. 11.

*et al.*<sup>50</sup> and Hamarneh *et al.*<sup>49</sup> calculated the geodesic distance using Isomap to perform tensor calculations and colorize the DTI tensor data. DT-MRI represents tensor data and step 3 of the Isomap and other NLDR algorithms (embedded space reconstruction) cannot be applied to integrate diffusion tensor images. Therefore, the full use of any DR techniques on DTI data is not feasible. Compared to previous work, in this manuscript, we compared six types of DR techniques (three linear and three nonlinear techniques) using both real and synthetic data and use the power of the DR methods to segment breast tissue and create an embedding image of

each parameter. Future work is ongoing to discern the importance of each parameter in the MRI data space. In another report, Richards *et al.*,<sup>51</sup> demonstrated that DfM outperformed PCA (linear DR) in the classification of galaxy redshift spectra when used as input in a regression risk model. In addition, DfM gave better visualization of the reparameterized red-shift separation in the embedded image, but both methods were able to separate the first few components to identify the different subgroups, which are consistent with our results. Similarly, we have demonstrated a method by which to reconstruct an embedded image from the higher-

dimension MRI data space into a reduced (embedded) dimension to use for tumor segmentation and visualization. Moreover, to deal with the high computational load of DR, we tested the stability of each method in relation to the input parameters to determine the optimal range for correct segmentation as defined by the current clinical imaging standard and found ranges that can be utilized for as starting points for other applications using breast MRI or other related data sets. In summary, by combining different MRI sequences, using dimensionality reduction and manifold learning techniques, we developed a robust and fully automated tumor segmentation and visualization method. This approach can be extended to facilitate large-scale multiparametric/multimodal medical imaging studies designed to visualize and quantify different pathologies.<sup>52</sup>

## ACKNOWLEDGMENTS

The authors thank the reviewers and editorial support of Mary McAllister, MS. This work was supported in part by NIH R01CA100184, P50CA103175, 5P30CA06973, Breast SPORE P50CA88843, Avon Foundation for Women:01-2008-012, U01CA070095, and U01CA140204.

<sup>a)</sup>Presented in part at the 52nd AAPM Annual Meeting in Philadelphia, PA.

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: mikej@mri.jhu.edu; Telephone: 410-955-7483; Fax: 410-614-1948.

<sup>1</sup>M. Filippi and R. I. Grossman, "MRI techniques to monitor MS evolution: the present and the future," *Neurology* **58**(8), 1147–1153 (2002).

<sup>2</sup>M. A. Jacobs, V. Stearns, A. C. Wolff, K. Macura, P. Argani, N. Khouri, T. Tsangaris, P. B. Barker, N. E. Davidson, Z. M. Bhujwalla, D. A. Bluemke, and R. Ouwerkerk, "Multiparametric magnetic resonance imaging, spectroscopy and multinuclear ((2)(3)Na) imaging monitoring of preoperative chemotherapy for locally advanced breast cancer," *Acad. Radiol.* **17**(12), 1477–1485 (2010).

<sup>3</sup>B. A. Landman, A. J. Huang, A. Gifford, D. S. Vikram, I. A. Lim, J. A. Farrell, J. A. Bogovic, J. Hua, M. Chen, S. Jarso, S. A. Smith, S. Joel, S. Mori, J. J. Pekar, P. B. Barker, J. L. Prince, and P. C. van Zijl, "Multiparametric neuroimaging reproducibility: A 3-T resource study," *Neuroimage* **54**(4), 2854–2866 (2011).

<sup>4</sup>M. W. Vannier, R. L. Butterfield, D. L. Rickman, D. M. Jordan, W. A. Murphy, and P. R. Biondetti, "Multispectral magnetic resonance image analysis," *Radiology* **154**, 221–224 (1985).

<sup>5</sup>M. A. Jacobs, R. A. Knight, H. Soltanian-Zadeth, Z. G. Zheng, A. V. Goussev, D. J. Peck, J. P. Windham, and M. Chopp, "Unsupervised segmentation of multiparameter MRI in experimental cerebral ischemia with comparison to T2, Diffusion, and ADC MRI parameters and histopathological validation," *J. Magn. Reson. Imaging* **11**(4), 425–437 (2000).

<sup>6</sup>M. A. Jacobs, R. Ouwerkerk, A. C. Wolff, V. Stearns, P. A. Bottomley, P. B. Barker, P. Argani, N. Khouri, N. E. Davidson, Z. M. Bhujwalla, and D. A. Bluemke, "Multiparametric and multinuclear magnetic resonance imaging of human breast cancer: Current applications," *Technol. Cancer Res. Treat.* **3**(6), 543–550 (2004).

<sup>7</sup>M. Seitz, A. Shukla-Dave, A. Bjartell, K. Touijer, A. Sciarra, P. J. Bastian, C. Stief, H. Hricak, and A. Graser, "Functional magnetic resonance imaging in prostate cancer," *Eur. Urol.* **55**(4), 801–814 (2009).

<sup>8</sup>Y. Guo, Y. Q. Cai, Z. L. Cai, Y. G. Gao, N. Y. An, L. Ma, S. Mahankali, and J. H. Gao, "Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging," *J. Magn. Reson. Imaging* **16**(2), 172–178 (2002).

<sup>9</sup>R. H. El Khouli, M. A. Jacobs, S. D. Mezban, P. Huang, I. R. Kamel, K. J. Macura, and D. A. Bluemke, "Diffusion-weighted imaging improves the diagnostic accuracy of conventional 3.0-T breast MR imaging," *Radiology* **256**(1), 64–73 (2010).

<sup>10</sup>R. H. El Khouli, K. J. Macura, I. R. Kamel, M. A. Jacobs, and D. A. Bluemke, "3-T dynamic contrast-enhanced MRI of the breast: Pharmacokinetic

parameters versus conventional kinetic curve analysis," *AJR, Am. J. Roentgenol.* **197**(6), 1498–1505 (2011).

<sup>11</sup>X. Chen, M. O. Moore, C. D. Lehman, D. A. Mankoff, T. J. Lawton, S. Peacock, E. K. Schubert, and R. B. Livingston, "Combined use of MRI and PET to monitor response and assess residual disease for locally advanced breast cancer treated with neoadjuvant chemotherapy," *Acad. Radiol.* **11**(10), 1115–1124 (2004).

<sup>12</sup>M. A. Jacobs, A. C. Wolff, R. Ouwerkerk, S. Jeter, E. Gaberialson, H. Warzecha, D. A. Bluemke, R. L. Wahl, and V. Stearns, "Monitoring of neoadjuvant chemotherapy using multiparametric, <sup>23</sup>Na sodium MR, and multimodality (PET/CT/MRI) imaging in locally advanced breast cancer," *Breast Cancer Res. Treat.* **128**(1), 119–126 (2011).

<sup>13</sup>K. Doi, "Overview on research and development of computer-aided diagnostic schemes," *Semin Ultrasound CT MR* **25**(5), 404–410 (2004).

<sup>14</sup>R. H. El Khouli, M. A. Jacobs, and D. A. Bluemke, "Magnetic resonance imaging of the breast," *Semin Roentgenol.* **43**(4), 265–281 (2008).

<sup>15</sup>D. Thorwarth, X. Geets, and M. Paiusco, "Physical radiotherapy treatment planning based on functional PET/CT data," *Radiother. Oncol.* **96**(3), 317–324 (2010).

<sup>16</sup>T. Twellmann, A. Meyer-Baese, O. Lange, S. Foo, and T. W. Nattkemper, "Model-free visualization of suspicious lesions in breast MRI based on supervised and unsupervised learning," *Eng. Applic. Artif. Intell.* **21**(2), 129–140 (2008).

<sup>17</sup>D. Cobzas, P. Mosayebi, A. Murtha, and M. Jagersand, "Tumor invasion margin on the Riemannian space of brain fibers," *Med. Image Comput. Comput. Assist. Interv.* **12**(Pt 2), 531–539 (2009).

<sup>18</sup>A. Meyer-Baese, T. Schlossbauer, O. Lange, and A. Wismueller, "Small lesions evaluation based on unsupervised cluster analysis of signal-intensity time courses in dynamic breast MRI," *Int. J. Biomed. Imaging* **2009**, 326924 (2009).

<sup>19</sup>P. Niyogi and N. K. Karmarkar, "An Approach to data reduction and clustering with theoretical guarantees," *Proc. Int. Conf. Mach. Learn.* **17**, 679–686 (2000).

<sup>20</sup>H. S. Seung and D. D. Lee, "COGNITION: The manifold ways of perception," *Science* **290**(5500), 2268–2269 (2000).

<sup>21</sup>J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**(5500), 2319–2323 (2000).

<sup>22</sup>S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**(5500), 2323–2326 (2000).

<sup>23</sup>L. Tong and Z. Hongbin, "Riemannian manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(5), 796–809 (2008).

<sup>24</sup>J. M. Lee, *Riemannian Manifolds: An Introduction to Curvature* (Springer, New York, 1997).

<sup>25</sup>I. T. Jolliffe, *Principal Component Analysis* (Springer, New York, 2002).

<sup>26</sup>R. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.* **7**, 179–188 (1936).

<sup>27</sup>R. Breiger, S. Boorman, and P. Arabie, "An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling," *J. Math. Psychol.* **12**(3), 328–383 (1975).

<sup>28</sup>M. Partridge and R. A. Calvo, "Fast dimensionality reduction and simple PCA," *Intell. Data Anal.* **2**(1–4), 203–214 (1998).

<sup>29</sup>A. Sehad, A. Hadid, H. Hocini, M. Djeddi, and S. Ameur, "Face recognition using neural networks and eigenfaces," *Comput. Their Appl.* **13**(454), 253–257 (2000).

<sup>30</sup>A. Santhanam and M. Masudur Rahman, Moving Vehicle Classification Using Eigenspace. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, IROS 2006: 3849–3854 (2006).

<sup>31</sup>J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(5), 572–575 (1998).

<sup>32</sup>M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.* **14**, 585–591 (2002).

<sup>33</sup>M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(3), 377–391 (2006).

<sup>34</sup>L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.* **4**, 119–155 (2003).

- <sup>35</sup>B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a Kernel eigenvalue problem," *Neural Comput.* **10**(5), 1299–1319 (2006).
- <sup>36</sup>R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proc. Natl. Acad. Sci. U.S.A.* **102**(21), 7426–7431 (2005).
- <sup>37</sup>B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *Appl. Comput. Harmon. Anal.* **21**(1), 113–127 (2006).
- <sup>38</sup>Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Sci. Comput. (USA)* **26**(1), 313–338 (2004).
- <sup>39</sup>T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.* **43**(1), 59–69 (1982).
- <sup>40</sup>E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.* **1**(1), 269–271 (1959).
- <sup>41</sup>T. F. Cox, *Multidimensional Scaling*, 1st ed. (Chapman and Hall, London, 1994).
- <sup>42</sup>L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
- <sup>43</sup>O. Salvado, C. Hillenbrand, Z. Shaoxiang, and D. L. Wilson, "Method to correct intensity inhomogeneity in MR images for atherosclerosis characterization," *IEEE Trans. Med. Imaging* **25**(5), 539–552 (2006).
- <sup>44</sup>S. G. Mallat, "A theory for multiresolution signal decomposition—The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989).
- <sup>45</sup>S. G. Mallat, *A Wavelet Tour of Signal Processing*, The Sparse Way, 3rd ed. (Academic Press, Burlington, MA, 2008).
- <sup>46</sup>S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Trans. Med. Imaging* **22**(7), 865–874 (2003).
- <sup>47</sup>B. K. Horn, *Robot Vision* (MIT, Cambridge, MA, 1986).
- <sup>48</sup>P. Tiwari, M. Rosen, and A. Madabhushi, "A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS)," *Med. Phys.* **36**(9), 3927–3939 (2009).
- <sup>49</sup>G. Hamarneh, C. McIntosh, and M. S. Drew, "Perception-based visualization of manifold-valued medical images using distance-preserving dimensionality reduction," *IEEE Trans. Med. Imaging* **30**(7), 1314–1327 (2011).
- <sup>50</sup>R. Verma, P. Khurd, and C. Davatzikos, "On analyzing diffusion tensor images by identifying manifold structure using isomaps," *IEEE Trans. Med. Imaging* **26**(6), 772–778 (2007).
- <sup>51</sup>J. W. Richards, P. E. Freeman, A. B. Lee, and C. M. Schafer, "Exploiting low-dimensional structure in astronomical spectra," *Astrophys. J.* **691**(1), 32–42 (2009).
- <sup>52</sup>M. A. Jacobs, "Multiparametric magnetic resonance imaging of breast cancer," *J. Am. Coll. Radiol.* **6**(7), 523–526 (2009).