STEP: Spatiotemporal enhancement pattern for MR-based breast tumor diagnosis

Yuanjie Zheng and Sarah Englander
Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Sajjad Baloch
Siemens Corporate Research, Princeton, New Jersey 08540

Evangelia I. Zacharaki
Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Yong Fan
Department of Radiology and BRIC, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina 27510

Mitchell D. Schnall
Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Dinggang Shen[a)]
Department of Radiology and BRIC, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina 27510

(Received 23 October 2008; revised 9 May 2009; accepted for publication 11 May 2009; published 15 June 2009)

The authors propose a spatiotemporal enhancement pattern (STEP) for comprehensive characterization of breast tumors in contrast-enhanced MR images. By viewing serial contrast-enhanced MR images as a single spatiotemporal image, they formulate the STEP as a combination of (1) dynamic enhancement and architectural features of a tumor, and (2) the spatial variations of pixelwise temporal enhancements. Although the latter has been widely used by radiologists for diagnostic purposes, it has rarely been employed for computer-aided diagnosis. This article presents two major contributions. First, the STEP features are introduced to capture temporal enhancement and its spatial variations. This is essentially carried out through the Fourier transformation and pharmacokinetic modeling of various temporal enhancement features, followed by the calculation of moment invariants and Gabor texture features. Second, for effectively extracting the STEP features from tumors, we develop a graph-cut based segmentation algorithm that aims at refining coarse manual segmentations of tumors. The STEP features are assessed through their diagnostic performance for differentiating between benign and malignant tumors using a linear classifier (along with a simple ranking-based feature selection) in a leave-one-out cross-validation setting. The experimental results for the proposed features exhibit superior performance, when compared to the existing approaches, with the area under the ROC curve approaching 0.97. © *2009 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3151811]

Key words: breast tumor diagnosis, feature extraction, classification, MRI, dynamic contrast enhancement

I. INTRODUCTION

Breast cancer is one of the major causes of cancer-related deaths among women, with annual mortality of over 400 000 worldwide. Like most cancers, its early detection can significantly improve therapeutic outcomes and increase the survival rates for the victims. Although biopsy has traditionally remained on the forefront in this regard, it is not ideally suited for periodic screenings due to the involvement of surgical procedures. Recently, magnetic resonance (MR) imaging has emerged as a method for *in vivo* detection, diagnosis, and characterization of breast tumors.[1] Studies have shown improved cancer detection in high risk populations by the introduction of yearly breast MRI into the screening regimen. This has led to widespread dissemination of screening breast MRI in high risk women.

Dynamic contrast-enhanced MR imaging (DCE-MRI) detects angiogenesis and forms the primary basis for detecting and characterizing breast cancer with MRI. It involves administration of a gadolinium-based contrast agent, followed by the acquisition of a temporal sequence of MRI images of the breast under investigation. High permeability of tumor capillaries allows the contrast agent to diffuse faster in a tumor, leading to better enhancement of a tumor relative to the surrounding breast tissue. This leads to higher sensitivity of DCE-MRI that is critical for accurate breast cancer diagnosis.[2]

In the DCE-MRI image, malignant and benign tumors have been found to exhibit three major types of spatiotemporal difference, i.e., (1) variations in their temporal enhancements (TEs), (2) variations in spatial and morphologi-

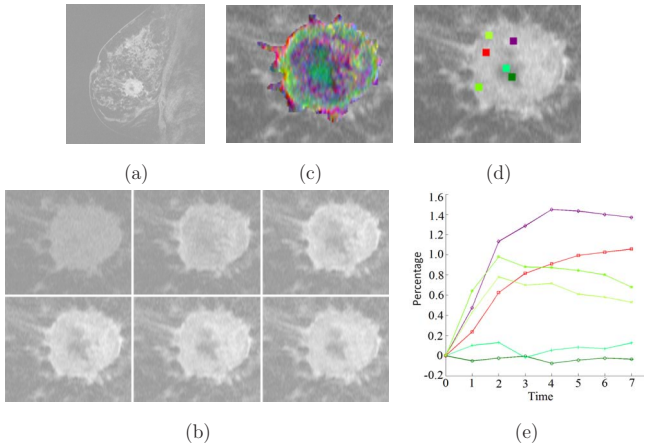


FIG. 1. Visualization of dynamic enhancement, spatial structure, and spatial variation in enhancement in a malignant tumor: (a) A complete slice, (b) temporal enhancement of tumor (from left to right and from top to bottom); (c) color-coded pixelwise enhancement, represented by three (PCA-based) major enhancement profiles in this tumor; [(d) and (e)] typical points and their corresponding enhancement profiles. The correspondence between points and profiles is specified by colors. In (e), the horizontal axis represents the acquisition times after the injection of the contrast agent, whereas the vertical axis is the percentage of temporal enhancement computed by Eq. (3).

cal structures, and (3) spatial variation in temporal enhancement. For instance, the dynamic response of malignant tumors shows strong early enhancement, followed by a rapid washout, with that of benign tumors indicating a slow and persistent increase in the enhancement.[3] On the other hand, spiculated borders and irregular shapes are prominent markers of malignancy, whereas smooth borders are typical characteristics of benignancy.[1] Furthermore, heterogeneous and peripheral enhancements are common in malignant tumors, whereas spatially homogeneous enhancements are often exhibited in benign tumors.[2,4,1] Although these spatial variations in temporal enhancement have been used by radiologists to better identify tumors[5,3] and to monitor their response to chemotherapy,[6] they have rarely been employed as the descriptive features for computer-aided diagnosis (CAD). The difference between benign and malignant tumors in terms of these spatiotemporal attributes may better be understood through Fig. 1, which illustrates a typical malignant tumor [Fig. 1(a)]. The morphological characteristics of this malignant tumor include rounded shapes, spiculated margins, and heterogeneous enhancements. The enhancement pattern (the term was coined in Ref. 6) is illustrated by color-coded rings shown in Fig. 1(c). These color codings of pixelwise enhancement are generated through principal component analysis (PCA) on the original pixelwise enhancement profiles. Three major components are retained and the corresponding scores are visualized (after normalization to the range [0,255]) with red, green, and blue, respectively. The color-coded pixelwise enhancement represents major variations in temporal enhancements within the tumor. Temporal profiles of some tumor voxels marked in Fig. 1(d) are given in Fig. 1(e), indicating variations in pixelwise enhancement within the tumor. These include various washout profiles, as well as a persistently increasing enhancement and a plateau enhancement.

For differentiation of benign from malignant tumors, one requires the extraction of features that best describe their spatiotemporal difference. Various features have been proposed in literature, which will be reviewed in Sec. II. However, their use is mostly limited in practice because either they describe only one type of difference or they are too coarse to contain rich information on the tumor.

In this article, we propose the spatiotemporal enhancement pattern (STEP) for capturing a comprehensive set of these features, including temporal enhancement, spatial structure, as well as spatial variation of pixelwise temporal enhancement.[7,8] The premise is that with the proposed STEP features, an MR CAD system can exhibit higher accuracy on breast cancer diagnosis due to richer representation of tumors. This is achieved by treating the serial contrast-enhanced images of tumor as a single spatiotemporal image and computing the STEP features. To this end, we use Fourier transformation and pharmacokinetic (PK) modeling to characterize the temporal enhancement, while moment invariants and Gabor texture features are used to characterize the spatial properties of the enhancements. In this article, we aim to highlight the importance of the STEP features in cancer diagnosis. We, therefore, use a linear classifier, along with a simple ranking-based feature selection method, to evaluate the performance of the STEP features in cancer diagnosis.

Since the STEP features are extracted from the tumor region, accurate segmentation of the tumor is important. In order to improve segmentation accuracy, we also propose a segmentation algorithm that is particularly suited to the problem under consideration. Based on energy minimization through graph cuts,[9,10] it refines a rough segmentation manually marked by a rater to yield an accurate representation.

This article is organized as follows. We first review the existing features used in breast tumor CAD in Sec. II. We then introduce the three main algorithmic components of our tumor diagnostic system in Sec. III, namely, (1) the segmentation algorithm for refining a coarse segmentation, (2) the extraction of the STEP features, and (3) the classification of the STEP features for cancer diagnosis using a linear classifier. A comparison of the proposed STEP features with existing features is presented in Sec. IV, where the importance of our segmentation refinement algorithm for tumor diagnosis is also demonstrated. We will finally conclude with a discussion of results in Sec. V.

II. PREVIOUS WORK

As explained in Sec. V, there are three main types of spatiotemporal difference between benign and malignant breast tumors. In the CAD of breast tumors, the features extracted from a breast tumor can be classified into three classes describing the three types of difference. Moreover, it is generally agreed that a combination of different types of features improves the diagnostic performance.[3,1,11]

The temporal enhancement pattern of a tumor is characterized by dynamic features,[3,1,12,2,13,14] which are typically derived from the statistical properties (such as the average or the maximum) of the time-intensity curves within the tumor region. Chen and Giger,[15] for instance, determined some representative temporal enhancement curves through clustering, from which they extracted some dynamic features such as maximum enhancement, uptake rate at the time point with maximum enhancement, and the washout rate between the time point with maximum enhancement and the final time point. They reported a value of 0.8 for the area under the receiver operating characteristic (ROC) curve with the washout rate for the classification of benign and malignant tumors. Tanner *et al.*,[16] on the other hand, derived the dynamic features from the standard deviation of the average signal. Dynamic features have also been captured through PK models,[17,6,18] which provide insight into the underlying physiology and are specifically used for the analysis of contrast agent uptake in DCE-MRI of breasts. The PK models were shown efficient in estimating the temporal enhancement and are very helpful in obtaining accurate segmentation and registration results. Qualitative dynamic measures may also be incorporated in these features by classifying curves as washout, persistent, and plateau.[1,5]

Architectural features are derived from the spatial tumor region to characterize the morphology of a tumor.[3,1,12,2,13] They typically describe the sharpness of the tumor margin, the shape of a tumor, the radial gradient,[2] the circularity, irregularity, margin gradients, and the variance of margin gradients for better spatial description of a tumor.[4,13] In addition, they may also rely on qualitative ratings of tumor shapes into round, oval, lobulated, irregular, and stellate, or of tumor margins into smooth, scalloped, irregular, and spiculated.[5,1,12]

The spatial variations in the temporal enhancement are described by various quantitative[2,4] and qualitative[11,3,1] approaches. In particular, the quantitative features may affect the diagnostic performance if they are too simple. Gilhuijs *et al.* used variance of uptake and change in variance of uptake,[2] whereas Chen *et al.* proposed enhancement-variance dynamic features.[4] Hayes *et al.*[6] utilized tumor region histograms of PK model parameters computed from pixelwise contrast agent uptake and then derived a number of local histogram descriptors such as mean, median, range, and skewness to describe the spatial variations in the temporal enhancement. Qualitative approaches, on the other hand, require classification of enhancements within a tumor by experts as homogeneous, heterogeneous, or rimlike.[11,3,1]

As explained above, the existing dynamic features and architectural features can only describe one type of spatiotemporal difference between the benign and malignant tumors. This limits their utility due to the lack of sufficient information. Although the features capturing the spatial variation in the temporal enhancement have potential to bear richer information, most of them are too simple. For tumor classification, the existing features may usually produce high sensitivity, but at the cost of low specificity. In contrast, the proposed STEP features are more comprehensive and are capable of simultaneously producing high sensitivity and specificity.

III. METHODS

In this section, we develop the STEP based cancer diagnosis using temporal (two-dimensional) 2D MR images. The proposed method consists of three main steps: Segmentation of the tumor region, extraction of the STEP features from the segmented region, and tumor classification into malignant or benign categories using the extracted features. Our segmentation algorithm refines a rough manual segmentation provided by a rater through a graph-cut based energy minimization. In order to achieve accurate segmentations, a new energy functional is proposed, which incorporates spatial as well as temporal properties of contrast-enhanced images. The STEP features are extracted from the segmented tumors by first normalizing spatially all the segmented tumor regions in order to eliminate scale variations. Then, enhancement modeling methods such as Fourier transformation and PK modeling are used to capture pixelwise temporal enhancement properties, as well as spatial description methods such as moment invariants and Gabor texture features to characterize the spatial variations in temporal enhancement. Tumor classification is performed using a linear classification method along with a simple ranking-based feature selection method. The diagnostic value of the STEP features is evaluated by ROC analysis and compared against commonly used dynamic and architectural features.

III.A. Segmentation

Tumor segmentation is the precise identification of the spatial domain of a tumor, where manual segmentation by an expert is generally regarded as a gold standard. Like any manual segmentation, expert segmentation is also prone to inaccuracy and high interobserver and intraobserver variability. It has been reported that nonenhancing tissue may be included in manually segmented tumors.[13] In addition, manual segmentation is time consuming and requires an examination of the entire series of enhanced data and/or the profile of pixels while carefully delineating the tumor boundary. This analysis of spatial as well as temporal profiles, therefore, limits the use and scope of manual segmentation in contrast-enhanced images.

In many cases, semiautomated algorithms are more efficient. They are usually initialized with a coarse boundary that is refined to yield an accurate representation of a tumor. With semiautomated algorithms, the time for manual delineation may be significantly reduced. It has been reported in Ref. 13 that the discrimination of benign and malignant breast tumors may be improved by refining the tumor segmentation.

In the following, we will introduce a new graph-cut based algorithm[9,10] for tumor segmentation based on the initial rough segmentation by manual rater. Our graph-cut based algorithm applies a Markov random field (MRF) and produces very accurate segmentation even in the presence of

noise and other confounding effects introduced in the imaging process. The MRF model plays the major role in this regard and introduces the notion of spatial smoothness of the segmentation when compared to the conventional clustering based methods.

III.A.1. Formulation of segmentation refinement algorithm

Our segmentation refinement algorithm operates on a 2D rectangular region Ω representing an MR image that contains a tumor. Segmentation refinement algorithm assigns a label l_p to each pixel $p \in \Omega$, thereby forming a partition $\Omega_t \cup \Omega_b = \Omega : \Omega_t \cap \Omega_b = \emptyset$, where the subscripts t and b denote the tumor and the background regions, respectively.

Typically, segmentation ensures that Ω_i exhibits similar statistical properties. However, the temporal enhancement exhibits large variability within a tumor region, as shown in Fig. 1. This prevents one from assuming a single distribution for each partition. To account for various enhancement types, we define multiple classes (labels) for both tumor and background. Let L_t and L_b be the set of labels for a tumor and the background, respectively, with $L = L_t \cup L_b$ being the set of labels for all classes. Any pixel p is regarded to be on a tumor if and only if $l_p \in L_t$. The idea of assigning multiple classes respectively for tumor and background makes our approach significantly different from Ref. 13, where a single class representation was used for each region, thus failing to give satisfactory results for contrast-enhanced images.

In order to derive the energy functional for segmentation refinement, four factors are considered. First, it is required that the pixelwise TEs within each class should be statistically similar to a compact distribution. Second, the generated labels should be spatially coherent, which allows compliance with the Markovian property of segmentation. This amounts to penalizing neighboring pixels $(p,q) \in \mathcal{N}$ from having different class assignments, where $\mathcal{N} \subset \Omega \times \Omega$ is a neighborhood system of pixels. Third, the segmentation boundary between different classes should be placed along pixels with high gradient of temporal enhancement. This encourages the assignment of neighboring pixels with significantly different enhancements to different classes. In this article, we use $\mathcal{N}_d \subset \mathcal{N}$ to represent the neighboring pixels (p,q) with different enhancements, i.e., $l_p \neq l_q \, \forall (p,q) \in \mathcal{N}_d$. Finally, the resulting boundaries between tumor and background are forced to be close to the rough manual segmentation, which places a certain level of confidence on the manual segmentation. In subsequent text, we use $(p,q) \in \mathcal{N}_{tb} \subset \mathcal{N}_d$ to denote pairs of neighboring pixels such that p belongs to the tumor and q lies in the background. These four criteria are used to formulate a new energy functional given in Sec. III A 2, which is minimized through a graph-cut algorithm.

Our segmentation refinement algorithm has multiple advantages when compared with the previous methods.[13,10] For example,[13] segments a region of interest (ROI) according to a maximum a posteriori (MAP) criterion, and then utilizes morphological closing and hole filling operations to fill the holes in the segmented region. In contrast, our algorithm operates simultaneously in feature (enhancement) space and in image space (due to the first two terms in the energy functional respectively), thus obtaining a spatially coherent segmentation eliminating the need of any post-processing. Hence, noisy image segmentation results, such as those in[10] do not occur. Moreover, our algorithm attracts the boundaries towards locations of high confidence, such as areas of high gradients in the feature image and manually delineated tumor contours.

III.A.2. The proposed energy functional

Our segmentation refinement algorithm is based on the minimization through graph cut[10] of an energy functional that is composed of four terms corresponding to the criteria described above,

$$E = \sum_{p \in \Omega} E_1(l_p) + \lambda_1 \sum_{(p,q) \in \mathcal{N}} E_2(l_p, l_q) + \lambda_2 \sum_{(p,q) \in \mathcal{N}_d} E_3(l_p, l_q) + \lambda_3 \sum_{(p,q) \in \mathcal{N}_{tb}} E_4(l_p, l_q), \quad (1)$$

where factors λ_1, λ_2, and λ_3 control the relative importance of the four energy terms. The individual energy terms in Eq. (1) correspond to the four factors explained in Sec. III A 1.

E_1 captures the agreement between the model and the original image, enforcing statistical similarity on TEs of pixels within each class. It is usually referred to as the data term. In this article, we assume a Gaussian distribution model as in Ref. 9, which is estimated from the current label assignments in the corresponding classes. E_1 is, therefore, defined as

$$E_1(l_p) = 1 - Pr(\bar{F}(p) | \bar{\mu}_{l_p}, \bar{\sigma}_{l_p}), \quad (2)$$

where the latter term measures the likelihood of the feature vector $\bar{F}(p)$ belonging to the class l_p. The feature vector $\bar{F}(p)$ is composed of the relative temporal enhancement values of p and characterizes the dynamic response. The relative temporal enhancement for a pixel p is defined as

$$C(p,t) = \frac{I(p,t) - I(p,0)}{I(p,0)}, \quad t = 1, \ldots, T-1, \quad (3)$$

where $I(p,t)$ denotes the intensity of p at a scanning time t. T is the total number of time sections (e.g., Ref. 6). $\bar{F}(p) = [C(p,1), \ldots, C(p,T-1)]^T$, therefore, represents a vector composed of $T-1$ features. Each class l_p is represented by a multivariate Gaussian distribution with mean $\bar{\mu}_{l_p}$ and covariance Σ_{l_p}, which is a diagonal matrix if the components of $\bar{F}(p)$ are independent. We have

$$Pr(\bar{F}(p) | \bar{\mu}_{l_p}, \bar{\sigma}_{l_p}) = \frac{1}{(2\pi)^{(T-1)/2} |\Sigma_{l_p}|^{1/2}} \exp^{-1/2 (\bar{F}_p - \bar{\mu}_{l_p})' \Sigma_{l_p}^{-1} (\bar{F}_p - \bar{\mu}_{l_p})}. \quad (4)$$

$E_2(l_p, l_p)$ is introduced to impose the spatial smoothness constraints on the assigned labels, making the resulting labels to be spatially coherent. It may simply be defined as

$$E_2(l_p,l_q) = 1 - \delta(l_p - l_q), \tag{5}$$

where δ is the Kronecker delta function.

E_3 attracts the boundary toward the regions of high gradients and is defined as

$$E_3(l_p,l_q) = g_1(\|\bar{F}(p) - \bar{F}(q)\|), \tag{6}$$

where $\|\bar{F}(p)-\bar{F}(q)\|$ is the $L2$-norm of the difference of the feature vectors at pixels p and q with different labels. Function g_1 is defined as

$$g_1(\zeta) = \frac{1}{\zeta + 1}. \tag{7}$$

E_4 is defined to force the boundary to lie in the vicinity of manual delineations

$$E_4(l_p,l_q) = g_2(\xi \cdot D_{p,q}), \tag{8}$$

where

$$g_2(\zeta) = \frac{\zeta}{\zeta + 1} \tag{9}$$

and $D_{p,q}$ is the distance from the center point between pixels p and q to the manually delineated boundary, and ξ is a control parameter for $D_{p,q}$ on E_4 which needs to be specified empirically. We choose $\xi=10$ in our experiments.

The graph construction and energy minimization methods in graph cut[9,10] can be directly used to minimize the energy functional of Eq. (1), although there are two additional terms E_3 and E_4 in Eq. (1). By combining E_2, E_3, and E_4 into one pairwise term E_2', Eq. (1) becomes

$$E = \sum_{p \in \Omega} E_1(l_p) + \sum_{(p,q) \in \mathcal{N}} E_2'(l_p,l_q), \tag{10}$$

which is very similar to the energy functional in Refs. 9 and 10 and can be optimized directly by the graph-cut algorithm. $E_2'(l_p,l_q)$, itself, is written as

$$E_2'(l_p,l_q) = \begin{cases} \lambda_1 E_2(l_p,l_q) + \lambda_2 E_3(l_p,l_q) + \lambda_3 E_4(l_p,l_q) & \text{if } (p,q) \in \mathcal{N}_{tb} \\ \lambda_1 E_2(l_p,l_q) + \lambda_2 E_3(l_p,l_q) & \text{if } (p,q) \in \mathcal{N}_d \quad \text{and} \quad (p,q) \notin \mathcal{N}_{tb} \\ \lambda_1 E_2(l_p,l_q) & \text{if } (p,q) \in \mathcal{N} \quad \text{and} \quad (p,q) \notin \mathcal{N}_d \end{cases} \tag{11}$$

using the definitions of each individual energy term. As explained in Sec. III A 1, $\mathcal{N}_d$ and $\mathcal{N}_{tb}$ are determined by comparing the labels of each pair of pixels in $\mathcal{N}$.

Given the manual rough segmentations by a rater, we apply the graph-cut based refinement algorithm within a rectangular region around the initial contours. The rectangular region is determined by expanding the bounding box of the rough segmentations with 10 pixels both horizontally and vertically. The initialization of our graph-cut based algorithm is carried out by classifying the tumor area and the background area in the rectangular region into multiple classes using a k-means clustering algorithm. The energy functional of Eq. (1) is then minimized through the expansion move algorithm[9] due to its small computational cost and high accuracy in segmentation. Once the algorithm converges, all the segmented regions labeled as tumor in the initialization are regarded as tumor in the final segmentation.

Representing the tumor and background pixels by several classes in segmentation is very important for getting accurate results due to (1) the amount of spatial variations in temporal enhancements of the tumor region as discussed in Sec. I and (2) the possible presence of multiple tissue types partitioning the background region. While various data clustering approaches may be used for the determination of the number of classes, such as minimum description length criteria,[19] we have found experimentally that three classes per region are sufficient for accurate segmentation. This is also in accordance with the types of temporal enhancement patterns that are exhibited by a tumor as mentioned in Sec. I, namely, washout, persistent, and plateau.

The three parameters $\{\lambda_k, k=1,\ldots,3\}$ in Eq. (1) determine the relative importance of each term and are set empirically. In our experiments, the choice of $\lambda_1=2$, $\lambda_2=1.2$, and $\lambda_3=1$ yielded excellent results. Moreover, for density estimation in Eq. (2), we employed the maximum likelihood estimates.

We compare our approach with expert segmentation in Fig. 2. In the figure, the top, middle, and bottom rows, respectively, show manual initializations, algorithm refined

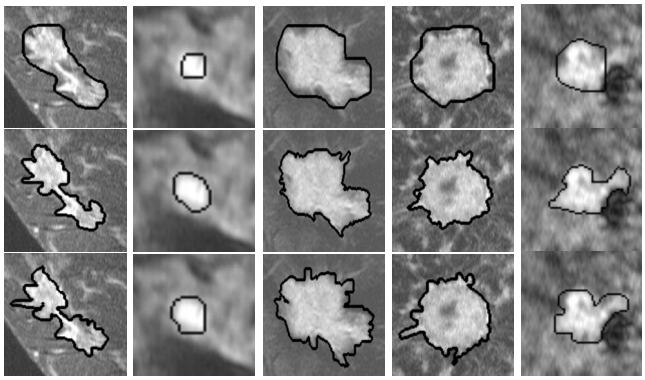


FIG. 2. Top panel: Rough manual segmentations by a rater. Middle panel: Corresponding refined segmentation results. Bottom panel: Corresponding manual segmentations by an expert.

segmentations, and carefully drawn manual segmentations by an expert (which are treated as the ground truth). It may be observed that the refined segmentations are close to the precise manual segmentations by the expert. To quantitatively assess the performance of our segmentation refinement algorithm, we measure the distance of the calculated segmentation curve to the carefully segmented manual curve. The distance between two contours A and B is defined as the average of the distances of contour A to B and contour B to A. The distance of contour A to B is the average distance of all points $a \in A$ to the closest point $b \in B$. Based on our entire dataset, the mean and standard deviation of these distances were found to be 4.10 and 5.62 pixels, which are significantly better than those for the rough segmentation (7.67 and 8.24 pixels).

Graph-cut based interactive segmentation techniques have recently become quite popular. Some recent works[20,21] utilize image gradient information at segmentation boundaries (as in E_3) as well as some prior knowledge (as in E_4) in the form of a rough segmentation. For example, the strategy for employing image gradient information in Ref. 20 is very similar to E_3 in our method. However, our method differs from most of the previous methods in the way prior segmentation is employed. References 20 and 21, for instance, treat certain prelabeled pixels (which are definitely foreground or background) as hard constraints, whereas our method minimizes the distance between the arbitrary segmentation boundaries and the ones from rough segmentation. This constrains the refined segmentation edges from drifting too far from the coarse segmentation edges, which is not explicitly ensured in Refs. 20 and 21.

III.B. Extraction of the STEP features

In order to compute the STEP features for characterizing a tumor that is represented by a sequence of 2D images, it is necessary to normalize the segmented tumor regions before modeling pixelwise temporal enhancements and their spatial variations. Normalization eliminates scale difference, thereby eliminating variations in tumor sizes from the cancer diagnosis. Pixelwise temporal enhancement is then captured through enhancement modeling techniques, such as Fourier transformation and PK modeling. Their spatial variations are eventually characterized by employing spatial description methods, such as moment invariants and Gabor texture features. The idea is not only to get a representation of tumor dynamics but also to exploit the structural difference in the texture of malignant and benign tumors.

III.B.1. Tumor normalization

We utilize Procrustes analysis to normalize 2D tumor regions. It involves eigendecomposition of the covariance matrix associated with the distribution of pixels in a given tumor region. Tumor regions are then rotated and scaled to ensure that (1) their principal directions are aligned with a reference coordinate space and (2) their "most significant eigenmodes" become identical to a predefined size (such as 35 mm). Later, we will employ rotation-invariant features for capturing spatiodynamic properties of a tumor. This will apparently make the normalization redundant. However, it is still required for the generalization of the proposed framework. Figure 3 shows some malignant and benign tumor samples before and after normalization, where bilinear interpolation is used for subpixel interpolation.

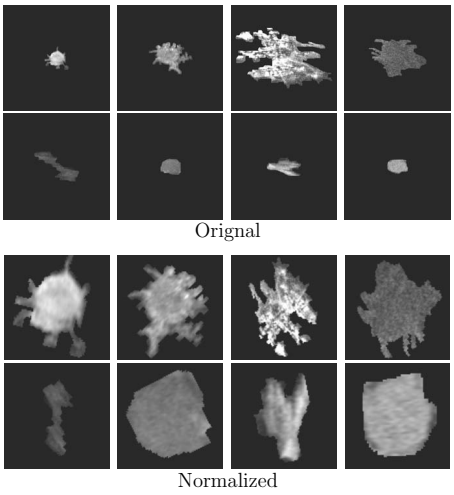


FIG. 3. Examples of tumors before and after normalization. For each panel, the tumors in the top row are malignant, whereas those in the bottom row are benign.

III.B.2. Temporal enhancement modeling

To capture the temporal response of various tissue types, we adopt and extensively compare two different signal representation techniques, namely, the general Fourier representation and a specific model-based PK representation of contrast agent uptake.[6]

Fourier transformation. Out of many available signal representations, such as Fourier transformation, discrete wavelets, and wavelet packets,[22] we choose Fourier transform to characterize the temporal enhancement curves of breast tissues. A pixelwise 1D discrete Fourier transform (DFT) is performed on the enhancement curve of each pixel p, $C(p,t)$ $(t=1,\ldots,T-1)$ [as defined in Eq. (3)], thus obtaining $T-1$ DFT coefficients for each pixel p. Consequently, for a given tumor, each DFT coefficient yields a distinctive temporal enhancement map, which collectively represents the frequency content of the corresponding temporal enhancements. Figure 4 shows the temporal enhancement map of the first DFT coefficient for the tumors in Fig. 3, from which we can see that the malignant tumors have more spiculated border and more heterogeneous enhancements and therefore are well differentiated from the benign tumors. In practice, we select N_t ($N_t=3$ in our results) enhancement maps corresponding to the lower order DFT coefficients.

Pharmacokinetic model. PK modeling provides insight into the underlying physiology and is specifically used for the analysis of contrast agent uptake in DCE-MRI of breasts.[6,23,17,24] With a few parameters, one may exploit PK modeling to capture the essential properties of contrast enhancements. Although several PK models have recently been

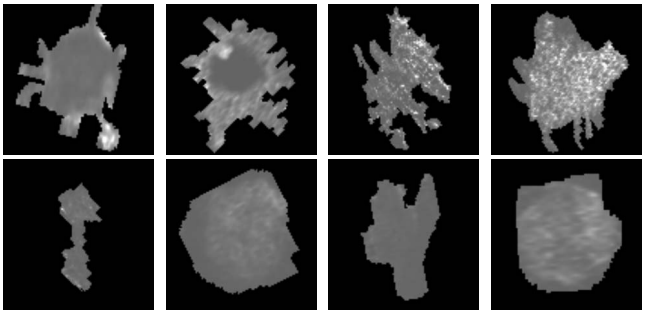


FIG. 4. Temporal enhancement maps of the first DFT coefficient for the tumors given in Fig. 3.

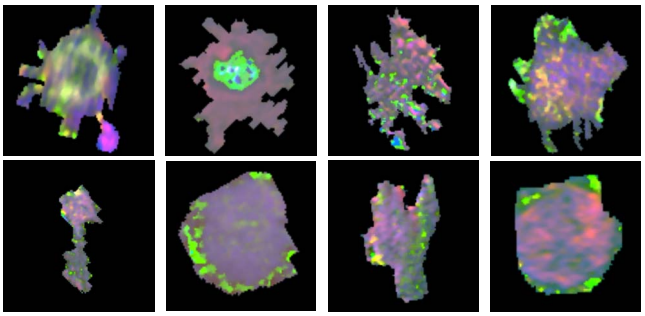


FIG. 5. Temporal enhancement maps of PK model parameters for the tumor samples given in Fig. 3. For the purpose of display, three temporal enhancement maps of each tumor sample (corresponding to A, a, and b) are integrated as different color components.

proposed,[23,17,24,18,25] we employ the model proposed in Ref. 24 for its ability to effectively capture the temporal enhancement properties. This PK model is of two compartment (i.e., blood plasma and EES) and is derived from the physics of the underlying process takes into account the observation that the relative signal increase is proportional to the concentration of contrast enhancement in the extracellular space of breast tissue. Mathematically, the relative signal increase based on the PK model is given by

$$\hat{C}(p,t) = \frac{A_p}{a_p - b_p}(e^{-b_p t} - e^{-a_p t}), \tag{12}$$

where $\hat{C}(p,t)$ is an estimate of the true signal increase $C(p,t)$ and A_p, a_p, and b_p are reparametrizations of the compartmental variables of pixel p (see details in Ref. 24).

Instead of independently estimating model parameters for each individual pixel,[17] all parameters (A_p, a_p, and b_p) for the entire breast image may be estimated jointly, thus increasing spatial consistency of the parameters. We follow the work of Schmid *et al.*[17] to incorporate a Gaussian Markov random field prior into the estimation of model parameters in Eq. (12). In other words, signal variations at different times are assumed to be modulated by the PK model plus some Gaussian noise. Thus, Bayesian inference principle may then be effectively exploited to estimate the unknown model parameters $\forall_p \in \Omega$. The Bayesian inference scheme in Ref. 17 assumes that each pixel is composed of different tissue types, and the PK on a pixel is determined as the average computed through the portion of each tissue at this pixel. It iteratively estimates the portion values for each pixel and the PK model parameters for each tissue.

Figure 5 shows the temporal enhancement maps of PK model parameters for the tumor samples in Fig. 3. Each PK model parameter gives rise to a temporal enhancement map, leading to a total of $N_t=3$ temporal enhancement maps corresponding to the three PK model parameters. This has allowed us to display these three parameter maps as a single RGB color map (for illustration purposes only). We can see that the enhancement maps in Fig. 5 bear rich information for distinguishing the malignant tumors from the benign tumors in the sense of the heterogeneous appearance of the malignant tumors and the homogeneous appearance of the benign tumors.

III.B.3. Spatial description of temporal enhancement maps

Once the tumor dynamics are modeled and the temporal enhancement maps are constructed, they are utilized to capture the spatial variations within a tumor. Since the orientation of a tumor sample is not related to its type, the features representing morphological and spatial structure should be rotation invariant. It may be seen from Fig. 4 that the shape and the local variation in the temporal enhancements alone are sufficient for distinguishing between malignant and benign tumors. Accordingly, we employ rotation-invariant moment features (referred to as rotation invariants in literature)[26] to capture the global structure, and Gabor rotation-invariant texture[27,28] to capture the local spatial behavior. Although several other rotation invariants exist in the literature,[29,30] we utilize[26] them due to their simplicity and easy interpretation.

Moment invariants. Image moments are computed as the particular weighted averages of pixel intensities to explain some global spatial distribution of image intensities. Hu's seven moment invariants[26] are, for instance, defined as a polynomial equation of some scale-normalized centralized moments. With regard to our work, the two-dimensional centralized moments of an $M \times N$ temporal enhancement map, denoted as $f(x,y)$, $(x,y) \in \Omega$, are defined as

$$m_{\alpha\beta} = \sum_{x=1}^{N}\sum_{y=1}^{M}(x-\bar{x})^{\alpha}(y-\bar{y})^{\beta}f(x,y), \tag{13}$$

where

$$\bar{x} = \frac{\Sigma_{x=1}^{N}(x\Sigma_{y=1}^{M}f(x,y))}{\Sigma_{x=1}^{N}\Sigma_{y=1}^{M}f(x,y)}$$

and

$$\bar{y} = \frac{\Sigma_{y=1}^{M}(y\Sigma_{x=1}^{N}f(x,y))}{\Sigma_{x=1}^{N}\Sigma_{y=1}^{M}f(x,y)}.$$

The two-dimensional scale-normalized centralized moment $\eta_{\alpha\beta}$ is defined as

$$\eta_{\alpha\beta} = m_{\alpha\beta}/m_{00}^{\gamma}, \tag{14}$$

where $\gamma=(\alpha+\beta)/2+1$, $\forall(\alpha+\beta)\geq 2$.

This readily leads to $H_m=7$ moment invariants[26] for each of N_t temporal enhancement maps. Consequently, each tumor sample is represented by $H_m \times N_t$ moment-invariant features.

Rotation-invariant gabor textures. Texture features of an image explain the local spatial properties of the image intensities. Gabor rotation-invariant features[27,28] are efficient texture descriptors. They are extracted in three steps. First, each Gabor function $g(x,y;\omega,\theta)$ with radial frequency ω and orientation θ is convolved with a temporal enhancement map $f(x,y)$ to obtain a Gabor feature $q(x,y;\omega,\theta)$ at each location (x,y). Subsequently, for each radial frequency ω and orientation θ, the Gabor features at different locations (x,y) in the temporal enhancement map $f(x,y)$ are averaged to obtain the average Gabor feature $Q(\omega,\theta)$. Since the average Gabor feature $Q(\omega,\theta)$ of a specific ω is a periodic function of θ with a period of π, the average Gabor features should be made rotation invariant. By calculating the magnitudes of Fourier coefficients on $Q(\omega,\theta)$, we obtain a number of rotation-invariant features for each ω.[28]

Thus, for K orientations θ within a period of π, K magnitudes of Fourier coefficients are obtained as rotation-invariant texture features for each frequency ω. If Z radial frequencies are used, one finally obtains $H_g=K\times Z$ rotation-invariant Gabor texture features, which we refer to as the STEP features for tumor classification. We used $K=8$ and $Z=4$ in our experiments.

By including both moment invariants and Gabor texture features, we obtain a total of $N_t\times(H_m+H_g)$ features for characterizing the STEP. These features capture the spatiotemporal profile of DCE images and may be used for tumor classification as explained in Sec. III C.

III.C. Tumor classification

The STEP features have been derived to potentially provide a rich representation of the spatiotemporal response of tumors to the contrast agent. However, the number of the STEP features is large (equals to 117 when $N_t=3$), and not all features have equal discriminative power in tumor classification. Thus, in the learning phase, the most discriminative features are selected, whereas the redundant features are discarded. To identify the significance of the STEP features in improving the classification performance, a linear classifier, along with a simple ranking-based feature selection method, is applied. Although advanced feature selection methods[31] and nonlinear classifiers[32] exist, we have observed that in the current study the computationally efficient linear classifier performs very well.

III.C.1. Feature selection

For classification purposes, some STEP features might be less effective, irrelevant, and redundant for classification. Therefore, it is important to select a small set of most discriminating features in order to improve the generalization ability and the performance of the finally constructed classifier.

A simple ranking-based feature selection method is used to select a small set of effective features for tumor classification. It first computes a ranking score for each feature according to its discriminative power, and then selects the top ranked features through a greedy algorithm in a set of best features. Classification is subsequently carried out via cross validation in an iterative manner. In each iteration, a feature in the left out features is found that yields best increase in the classification rate and is added to the set of best features. The process continues until there is no increase in the classification rate.

In particular, we use t score, denoted by t_{score}, to measure the capability of each individual feature of separating malignant and benign samples. The larger absolute t score indicates better distinction by a particular feature for classification. Since the sample size for benign and malignant can be different, t score is computed by the following equation:

$$t_{\text{score}} = \frac{\mu_+ - \mu_-}{\sqrt{\frac{(N_+-1)\sigma_+^2+(N_--1)\sigma_-^2}{N_++N_--2}\left(\frac{1}{N_+}+\frac{1}{N_-}\right)}}, \tag{15}$$

where μ_+ and σ_+^2 are the mean and variance, respectively, of a particular feature in all positive training samples (malignant samples) and μ_- and σ_-^2 are the mean and variance of the same feature in all negative training samples (benign samples). N_+ and N_- are the numbers of positive and negative training samples.

The above ranking-based feature selection method considers only the individual performance of the features and might be problematic if there are strong correlations among features. In such a case, a more sophisticated feature selection method[33] may be used. For the application of interest, this simple ranking method yields satisfactory performance.

III.C.2. Classification

We use the linear discriminant analysis[34] with Fisher linear discriminant rule to perform the tumor classification. The classification method is based on the fact that distributions (of the linear combination of the selected features), which have a greater variance between the two classes and smaller variance within each class, are easier to separate. With the trained classifier $\bar{W}$, a test sample $\bar{F}$ may be classified according to its score $\bar{W}^T\bar{F}$. Specifically, it is classified positive if $(|\bar{W}^T\bar{F}-\bar{W}^T\bar{\mu}_+|-|\bar{W}^T\bar{F}-\bar{W}^T\bar{\mu}_-|)<b$, where we set $b=0$ in our experiments, and negative otherwise. Here, $\bar{\mu}_+$ and $\bar{\mu}_-$ represent the mean values of feature vectors in the positive and negative training groups, respectively, and b is the standard threshold used to separate the classes in equal distance from the (projected) class means. A nonzero threshold value may be used in order to increase sensitivity or specificity depending on the application. It may be chosen manually, or determined by optimizing the classification rate over all training samples.

For the assessment of the classifier, a leave-one-out cross-validation strategy is employed, where a classifier is constructed with all but one sample. The "left out" sample is then treated as a test sample to be classified accordingly. The process is repeated until all samples are selected as the left out sample. Classification rate is finally computed as the mean of correctly classified samples.

IV. EXPERIMENTAL RESULTS

IV.A. Testing data

Images used in this study were acquired from patients with breast tumors in a 1.5 T scanner (Siemens Sonata) or a 3 T scanner (Siemens Trio). A specially designed surface breast coil array was employed to achieve high SNR as well as minimal level of distortion. The imaging protocol included bilateral fat suppressed *T*2 weighted images in the sagittal plane and a slab interleaved 3D fat suppressed spoiled gradient echo prior to and after the injection of contrast. A rapid bolus injection of 0.1 mmol/kg Gadopentetate dimeglumine (Omniscan; GE health, NJ), followed by a 10 ml saline flush was administered in all cases. Sequential postcontrast acquisitions were acquired for approximately 6 min following the contrast injection. The spoiled gradient echo sequence had a minimum spatial resolution of 20 cm over a 512×256 matrix and a minimum time of 90 s in the sagittal plane and slice thickness of 2–3.5 mm. One slice can contain either 384×384 pixels of 0.47×0.47 mm^2, 512×512 pixels of 0.35×0.35 mm^2, or 896×896 pixels of 0.22×0.22 mm^2, depending on the scanners or the protocols. In total, there were 36 subjects used in our experiments, including 22 malignant and 14 benign cases. All of these samples were histologically verified.

IV.B. Experiments

A series of experiments was performed to evaluate the effectiveness of the proposed segmentation refinement algorithm and the STEP features for tumor classification, as outlined next.

- Comparison of spatiotemporal models. In the first set of experiments, we carried out a comparison between DFT and PK representations of the temporal profiles in terms of tumor classification rates. For both signal representations, expert-segmented tumor samples were used. Initially, only moment invariants were employed to describe spatial variation in temporal enhancement features. DFT was found to perform better than PK modeling in terms of classification rates. Therefore, in all subsequent experiments, we proceeded with DFT as the appropriate choice for temporal representation.

  In the second step, we also introduced local Gabor texture features in the spatial representation of temporal enhancement, in addition to global moment invariants. Again expert-segmented tumor samples were employed for feature classification. It was observed that the introduction of local features, such as Gabor, yielded improved performance. We, therefore, conclude that (1) DFT is better suited to capture the temporal enhancement due to its completeness of representation, and (2) both moment invariants and Gabor texture features are essential for capturing the spatial variations in temporal enhancement.
- Evaluation of segmentation refinement algorithm. Based on the selection of the STEP features, the third experiment was performed to investigate the accuracy of our segmentation refinement algorithm in the context of tumor classification. In the breast tumor dataset, several inexperienced raters were asked to roughly delineate the tumors. The segmentation refinement algorithm was then applied on the roughly segmented tumor samples. It was found that the segmentation refinement improved the tumor classification performance.
- Validation of the STEP features. In the final set of experiments, we analyzed the performance of the STEP features by comparing them with the previously proposed features, such as the combinations of dynamic features, architectural features, and enhancement variation features. In order to demonstrate the stability of the classification rates for the STEP features with respect to the sample size, we carried out a comparison of the rates when training sets of different sizes were taken into account.

It is worth noting that the classification performances in all experiments were cross validated using a leave-one-out strategy as mentioned above. We employed binomial ROC curve fit through the ROCKIT algorithm[35] to determine sensitivity, specificity, and accuracy of our classifier. ROC curves measure false positive rate and true positive rate as the discrimination threshold is varied for a classifier. A large value of the area under the curve (AUC) corresponds to better performance of the corresponding classifier. Sensitivity may then be computed as TP/(TP+FN), specificity as TN/(TN+FP), and accuracy as (TP+TN)/(TP+TN+FP+FN), where TP, TN, FN, and FP denote true positive, true negative, false negative, and false positive, respectively. Herein, true positive (TP) denotes malignant tumor classified as malignant, while false positive (FP) corresponds to malignant tumor classified as benign. TN and FN are defined similarly.

In order to determine the significance level of the assessment based on the difference in AUC values, we performed a t-test to evaluate the p value (p_{val}) between each set of selected features and our STEP features under the null hypothesis that there exists no AUC value difference between them. To obtain the statistics of AUC values for each set of features, we performed a leave-one-out testing. Each time, we dropped off one subject and with the left subjects we obtained an ROC curve. This step was repeated until all subjects were dropped off. Finally, we got the same number of AUC values as the number of subjects. Note that for each computation of an ROC curve, the classification algorithm of Sec. III C is performed with another leave-one-out process.

The ROC curves explained above were only used for computing the p values in order to compare two sets of fea-

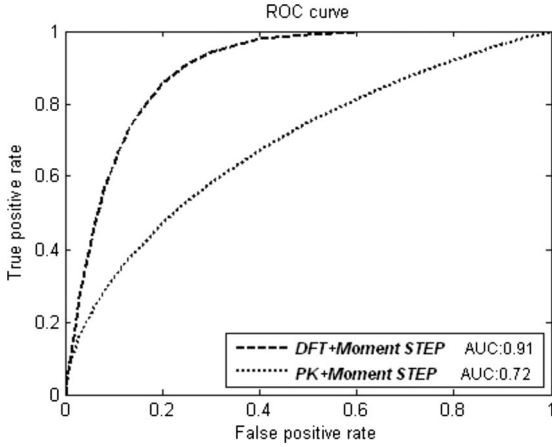


FIG. 6. ROC curves of various classifiers employing DFT or PK modeling as a temporal enhancement descriptor: $p_{val}=0.008$.

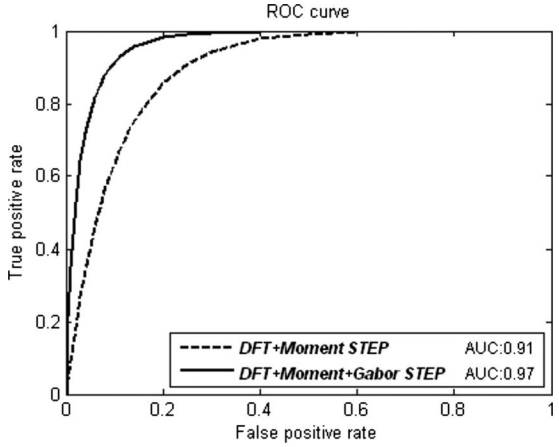


FIG. 7. ROC curves for tumor classification using the STEP features with and without local Gabor texture features: $p_{val}=0.019$.

tures. To show the performance of various types of features, we used the ROC curves computed from all subjects.

IV.B.1. Comparisons of DFT and PK modeling in enhancement description

In this experiment, we compare the performance of DFT with that of a PK model for tumor classification. Moment invariants calculated on expert-segmented tumor samples are used to describe the spatial variation in temporal enhancement features.

ROC curves given in Fig. 6 show that the tumor classification performance based on DFT for enhancement description is superior to that for PK modeling, although the PK model is specifically designed for characterizing the contrast enhancement in the DCE-MRI.[17] A possible explanation is the generality and completeness of DFT, which presumably leads to a richer description. On the other hand, three PK model parameters may not be sufficient to accurately represent the underlying signal variations. Some recent work by Armitage *et al.*[36] highlighted this limitation of PK models. In addition, their correlations may also not be ideal for the problem under consideration. Due to these reasons, we use DFT as an enhancement descriptor in all subsequent experiments.

IV.B.2. Effectiveness of local Gabor texture features

Based on the results of the previous experiment, we choose DFT for characterizing the temporal enhancement and moment invariants for capturing global spatial variations in temporal enhancement. In this experiment, we will investigate improvement in performance by the inclusion of Gabor features for capturing local variations in temporal enhancement. Again, we use expert-segmented tumor regions for training and testing purposes.

ROC curves given in Fig. 7 illustrate that the inclusion of local spatial descriptors improves the tumor classification by efficiently exploiting the local variations between malignant and benign tumors. It is, therefore, concluded that the combination of global and local features, along with temporal enhancement described by DFT, should be used for tumor classification.

IV.B.3. Effectiveness of segmentation refinement algorithm in tumor classification

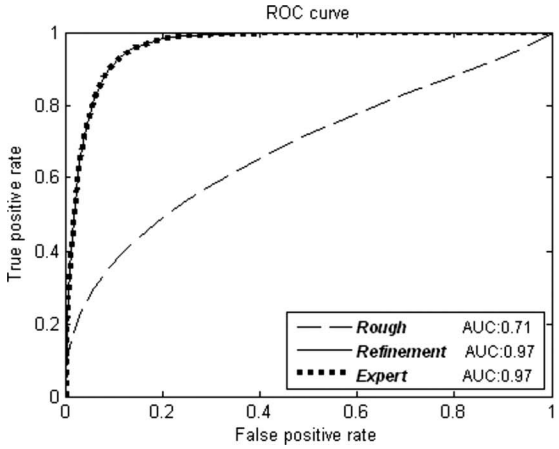
Validation of the segmentation refinement algorithm is carried out by comparing the performance of the tumor classification using rough manual segmentations, the refined segmentations calculated by our graph-cut based algorithm, and the accurate expert-segmentations. Having established the importance of the STEP features that are composed of DFT based temporal enhancement, and global and local spatial variations in temporal enhancement, we proceed with this combination.

As indicated by ROC curves in Fig. 8, the segmentation refinement algorithm improves the tumor classification rates, which is consistent with the results reported in Ref. 13. Additionally, it may be observed that the classification results on the refined segmentations are comparable to those on expert segmentations, as reflected by the level of agreement in the corresponding ROC curves in Fig. 8.

IV.B.4. Performance of the STEP features

The performance of the STEP features is analyzed by way of (1) a comparison between the tumor classification results for the STEP features and numerous existing features and (2) the examination of the effect of the sample size on classification accuracies.

To compare the performance of our STEP features with commonly used dynamic and architectural features in tumor diagnosis, we selected some existing state-of-the-art features. The STEP features were extracted using DFT to describe temporal enhancement, and both moment invariants and local Gabor texture features to capture spatial variations of temporal enhancement. In this experiment, all tumor regions were obtained by the segmentation refinement algorithm described previously.



	p_{val}
Rough & Refinement	0.024
Rough & Expert	0.018
Refinement & Expert	0.462

FIG. 8. ROC curves for tumor classification using rough segmentation, computer refined segmentation, and expert segmentation. Notice that ROC curves of "computer refined" and "expert" segmentations overlap.

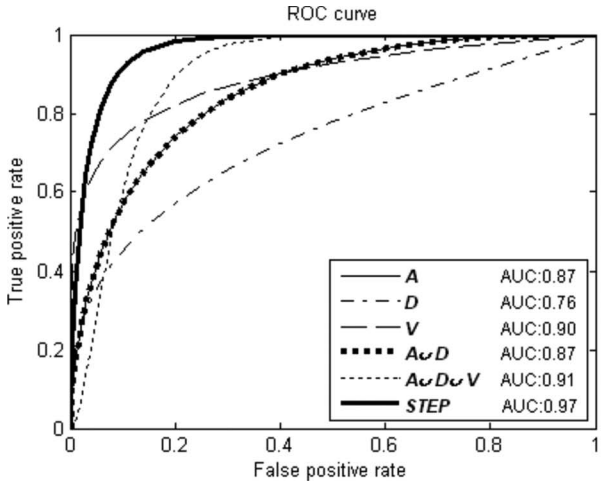
For comparison, we considered the dynamic features proposed in Refs. 4, 18, and 16 as they were shown efficient for representing the temporal enhancement properties. Specifically, standard deviation of enhancement (D_1), maximum washout (D_2),[16] maximum uptake (D_3), uptake rate (D_4) corresponding to the maximum uptake, washout rate (D_5) between the time point with the maximum uptake and the final time point,[4] and the parameters (D_6) and (D_7) of Hayton–Brady pharmacodynamic model of the enhancement curve[18] were taken into account. These features were computed from the average of the intensities in the tumor region at each time point.

For the architectural features, we considered the compactness (A_1),[16] the circularity (A_2),[4] the irregularity (A_3), the eccentricity (A_4), the rectangularity (A_5), and the entropy of radial length distribution (A_6) (Ref. 13) of a tumor, as they can adequately describe the architectural properties of tumors.

Features that describe the spatial variation in temporal enhancement are the variance of uptake (V_1), change in variance of uptake (V_2), margin gradient (V_3), variance of margin gradient (V_4), variance of radial gradient histogram (V_5),[2] the maximum variation in enhancement (V_6), the enhancement-variance increasing rate (V_7), the enhancement-variance decreasing rate (V_8), and the enhancement variance (V_9) at the first postcontrast scan.[4]

These features were combined as $A=\{A_1,A_2,\ldots,A_6\}$, $D=\{D_1,D_2,\ldots,D_7\}$, and $V=\{V_1,V_2,\ldots,V_9\}$ to represent three different characterizations of the tumor response. The classification performance of the individual features $\{A,D,V\}$, the joint features $\{A\cup D,A\cup D\cup V\}$, and the STEP features were compared using the same feature selection and leave-one-out classification procedure as explained in Sec. III C.

The fitted ROC curves and their AUC values for all feature types are given in Fig. 9. The classification accuracy, sensitivity, and specificity for the best subset of features (chosen during features selection) are listed in Table I. The feature types are shown in the first column and the corresponding best subset is shown in the last column. For instance, for the STEP features (shown in the last row of Table



	p_{val}
A & STEP	0.018
D & STEP	0.004
V & STEP	0.027
AuD & STEP	0.012
AuDuV & STEP	0.033

FIG. 9. ROC curves for tumor classification based on different combinations of features. The corresponding AUC values are also provided. Notice that ROC curves of A and $A\cup D$ completely overlap.

TABLE I. The best classification accuracy among all sets of selected features and the corresponding sensitivity and specificity for different sets of features used in tumor classification.

Feature	Accuracy (%)	Sensitivity	Specificity	Selected features
A	86.1	0.91	0.78	A_3, A_5, A_6
D	63.9	0.68	0.57	D_6
V	88.9	0.91	0.86	V_3, V_5
$A \cup D$	86.1	0.91	0.79	A_3, A_5, A_6
$A \cup D \cup V$	88.9	0.91	0.86	A_3, A_5, V_3, V_5
STEP	97.2	0.95	1.00	3(moment)+1(Gabor)

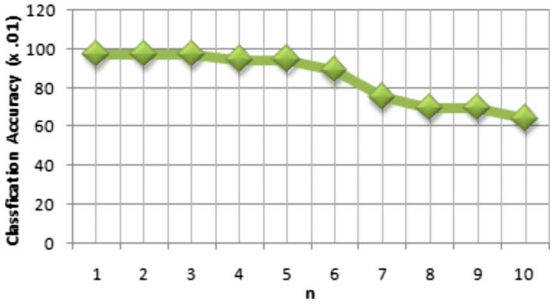


FIG. 10. Classification accuracy of the STEP features with respect to the number of samples left out in our leave-n-out cross-validation experiments. Here, n ranging from 1 to 10 is tested.

I), three moment invariants and one local Gabor texture feature are selected as the best subset. In particular, these features were the sixth moment invariant of the second DFT coefficient map, the first moment invariant of the first DFT coefficient map, and the second moment invariant of the fourth DFT coefficient map, respectively. The Gabor texture feature was calculated from the first DFT coefficient map. This suggests that it is mostly the low frequency components of the temporal enhancement that are most discriminating. It is reasonable considering the fact that the temporal enhancements are usually smooth over time.

For the five different combinations of the features in Table I, one can see that the selected features include the irregularity (A_3), the rectangularity (A_5), the entropy of radial length distribution of the architectural features (A_6), the margin gradient (V_3), and the variance of radial gradient histogram (V_5). In particular, the irregularity (A_3) and the margin gradient (V_3) are actually the main features used by radiologists for identifying malignant tumors. Also, the variance of radial gradient histogram (V_5) describes the spatial variation in the temporal enhancements, which radiologists always use to identify suspicious malignant tumors.

Combining general architectural and dynamic features ($A \cup D$) did not improve performance in terms of the AUC and the best classification rate when compared to individual architectural features. A possible explanation might be the use of the simple ranking-based feature selection method, which may not optimally combine both architectural and dynamic features. On the other hand, the spatial variations in the contrast enhancement were found to be much more informative. Their combination with the architectural and dynamic features improves the performance, as evident from the classification rates given in Table I and the AUC values shown in Fig. 9. This demonstrates that the dynamics, morphology, and spatial variation in temporal enhancement all play an important role for distinguishing between malignant and benign tumors.

The STEP features yielded superior performance in all experiments, with improvements in AUC, classification accuracy, sensitivity, and specificity. Since both moment invariants and local Gabor texture features were selected as the STEP features, the importance of both global and local variations in contrast enhancement seems evident.

In order to understand the influence of the size of the training set on the performance of the STEP features, we employed a series of leave-n-out validations by varying n from 1 to 10. A leave-n-out strategy is similar to the leave-one-out validation explained above except that each time n subjects are dropped instead of just one. Note that all the dropped subjects are treated as the test data and the left subjects are considered as the training set. A larger value of n means a smaller training set and vice versa. With the given dataset, the classification rates as a function of n are plotted in Fig. 10. As seen from the figure, the STEP features result in high rates when the size of training set is large and lead to a degradation when the size gets too small. This is true in general for feature classification, where correlations between the samples may be lost due to inadequate sample size. It can be seen that the performance is consistently maintained above 90% until $n=5$ (1/7th of the size of the actual dataset or seven-fold cross validation), which indicates the robustness of the proposed classification method. Since the slope of the curve at $n=5$ is not high, one can safely regard that it does not overestimate the true classification rate.

V. CONCLUSION

This article illustrates that temporal enhancement, architectural structure, and spatial variations of pixelwise temporal enhancement within a tumor area are distinguishing markers for malignant and benign tumors. Although these properties have widely been used by radiologist in tumor diagnosis, they are captured intuitively and are used independently in most computer-aided diagnosis systems, resulting in a loss of useful information for tumor diagnosis. Although features for capturing spatial variations in temporal enhancement were proposed in the past, their primitive nature remains confined to the modeling of simple variations while leaving out more complex behaviors.

Accordingly, we have proposed a framework for extracting the STEP for completely characterizing these three properties of tumors. Although in this article we have confined ourselves to DFT or PK modeling for capturing pixelwise temporal enhancements and to moment invariants and Gabor texture features for characterizing their spatial variations, it should be noted that the utility of this framework is not limited to this particular combination of features. In general, other advanced spatiotemporal analysis techniques may also

be incorporated. For accurate tumor classification, it is important to employ an effective feature selection method (with cross validation) in order to select the best set of features. Results presented in this article show that the extracted features should be acquired as richly as possible; then the feature selection algorithm can discard less discriminating features. This idea is further strengthened by noticing that it is not always straightforward to determine *a priori* the relative importance of different features, and their *ad hoc* selection (a common practice in breast tumor diagnosis algorithms) does not lead to optimal performance.

In short, we have presented a method for capturing the STEP features for tumor classification. Experimental results show that the STEP features exhibit better performance than combined dynamic, architectural features, and features on spatial variations in enhancements. We have also shown that refinement of rough segmentation by manual rater can improve tumor classification, with classification rates approaching those for expert segmentations. In the future, we plan to extensively evaluate our method on larger breast MR image datasets.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support of this work through National Cancer Institute grant 1R21CA140841.

[a]Electronic mail: dgshen@med.unc.edu

[1]M. D. Schnall *et al.*, "Diagnostic architectural and dynamic features at breast MR imaging: Multicenter study," Radiology **238**, 42–53 (2006).
[2]K. G. A. Gilhuijs *et al.*, "Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging," Med. Phys. **25**, 1647–1654 (1998).
[3]B. K. Szabo *et al.*, "Dynamic MR imaging of the breast: Analysis of kinetic and morphologic diagnostic criteria," Acta Radiol. **44**, 379–386 (2003).
[4]W. Chen *et al.*, "Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics," Med. Phys. **31**, 1076–1082 (2004).
[5]M. D. Schnall and D. M. Ikeda, "Lesions diagnosis working group report," J. Magn. Reson Imaging **10**, 982–990 (1999).
[6]C. Hayes, A. R. Padhani, and M. O. Leach, "Accessing changes in tumor vascular function using dynamic contrast-enhanced magnetic resonance imaging," NMR Biomed. **15**, 154–163 (2002).
[7]Y. Zheng, S. Englander, M. D. Schnall, and D. Shen, "Step: Spatial-temporal enhancement pattern, for mr-based breast tumor diagnosis," *Proceedings of the Fourth IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007* (IEEE, New York, 2007), pp. 520–523.
[8]Y. Zheng, S. Baloch, S. Englander, M. D. Schnall, and D. Shen, "Segmentation and classification of breast tumor using dynamic contrast-enhanced mr images," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2007*, Lecture Notes in Computer Science, Vol. 4792 edited by N. Ayache, S. Ourselin, and A. Maeder (Springer, Berlin, 2007), pp. 393–401.
[9]Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1222–1239 (2001).
[10]R. Zabih and V. Kolmogorov, "Spatially coherent clustering using graph cuts," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, CVPR 2004* (IEEE, New York, 2004), Vol. 2, pp. 437–444.
[11]B. K. Szabo *et al.*, "Neural network apporach to the segmentation and classification of dynamic magnetic resonance images of the breast: Comparison with empiric and quantitative kinetic parameters," Acad. Radiol. **11**, 1344–1354 (2004).
[12]D. M. Ikeda, "Progress report from the american college of radiology breast MR imaging lexicon committee," Magn. Reson. Imaging **9**, 295–302 (2001).
[13]C. Tanner *et al.*, "Classification improvement by segmentation refinement: Application to contrast-enhanced MR-mammography," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2004*, Lecture Notes in Computer Science Vol. 3216 edited by C. Barillot, D. R. Haynor, and P. Hellier (Springer, Berlin, 2004), pp. 184–191.
[14]J. Levman, T. Leung, P. Causer, D. Plewes, and A. L. Martel, "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines," IEEE Trans. Med. Imaging **27**, 688–696 (2008).
[15]W. Chen and M. L. Giger, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," Med. Phys. **33**, 2878–2887 (2006).
[16]C. Tanner *et al.*, "Does registration improve the performance of a computer aided diagnosis system for dynamic contrast-enhanced MR mammography," *Proceedings of the Third IEEE International Symposium on Biomedical Imaging: Macro to Nano—ISBI 2006* (IEEE, New York, 2006), pp. 466–469.
[17]V. J. Schmid, B. Whitcher, A. R. Padhani, N. J. Taylor, and G. Z. Yang, "Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging," IEEE Trans. Med. Imaging **25**, 1627–1636 (2006).
[18]X. Chen *et al.*, "Simultanous segmentation and registration of contrast-enhanced breast MRI," in *IPMI'05*, Lecture Notes in Computer Science, Vol. 3565, edited by G. E. Christensen and M. Sonka (Springer, Berlin, 2005), pp. 126–137.
[19]J. Rissanen, "Modelling by shortest data description," Automatica **14**, 465–471 (1978).
[20]Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," *International Conference on Computer Vision* (IEEE, New York, 2001), pp. 105–112.
[21]C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," ACM Trans. Graphics **23**, 309–314 (2004).
[22]S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 83–88 (1999).
[23]P. S. Tofts *et al.*, "Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: Standardized quantities and symbols," J. Magn. Reson Imaging **10**, 223–232 (1999).
[24]P. Hayton *et al.*, "Analysis of dynamic MR breast images using a model of contrast enhancement," Med. Image Anal. **1**, 207–224 (1996).
[25]N. Hylton, "Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker," J. Clin. Oncol. **24**, 3293–3298 (2006).
[26]M. K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Inf. Theory **8**, 179–187 (1962).
[27]T. N. Tan, "Texture feature extraction via visual cortical channel modelling," *Proceedings of the 11th IAPR International Conference on Pattern Recognition, 1992, Conference C: Image, Speech and Signal Analysis* (IEEE, New York, 1992), Vol. III, pp. 607–610.
[28]T. N. Tan, "Rotation invariant texture features and their use in automatic script identification," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 751–756 (1998).
[29]J.-L. Chen and A. Kundu, "Rotation and grey-scale transform invariant texture identification using wavelet decomposition and hmm," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 208–214 (1994).
[30]G. Eichmann and T. Kasparis, "Topologically invariant texture descriptors," Comput. Vis. Graph. Image Process. **41**, 267–281 (1988).
[31]H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining* (Springer, Berlin, 1998).
[32]V. Vapnik, *Statistical Learning Theory* (Wiley Interscience, New York, 1998).
[33]I. Guyon *et al.*, "Gene selection for cancer classification using support vector machines," Mach. Learn. **46**, 389–422 (2002).
[34]P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).
[35]C. E. Metz *et al.*, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[36]P. Armitage, C. Behrenbruch, M. Brady, and N. Moore, "Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast," Med. Image Anal. **9**, 315–329 (2005).