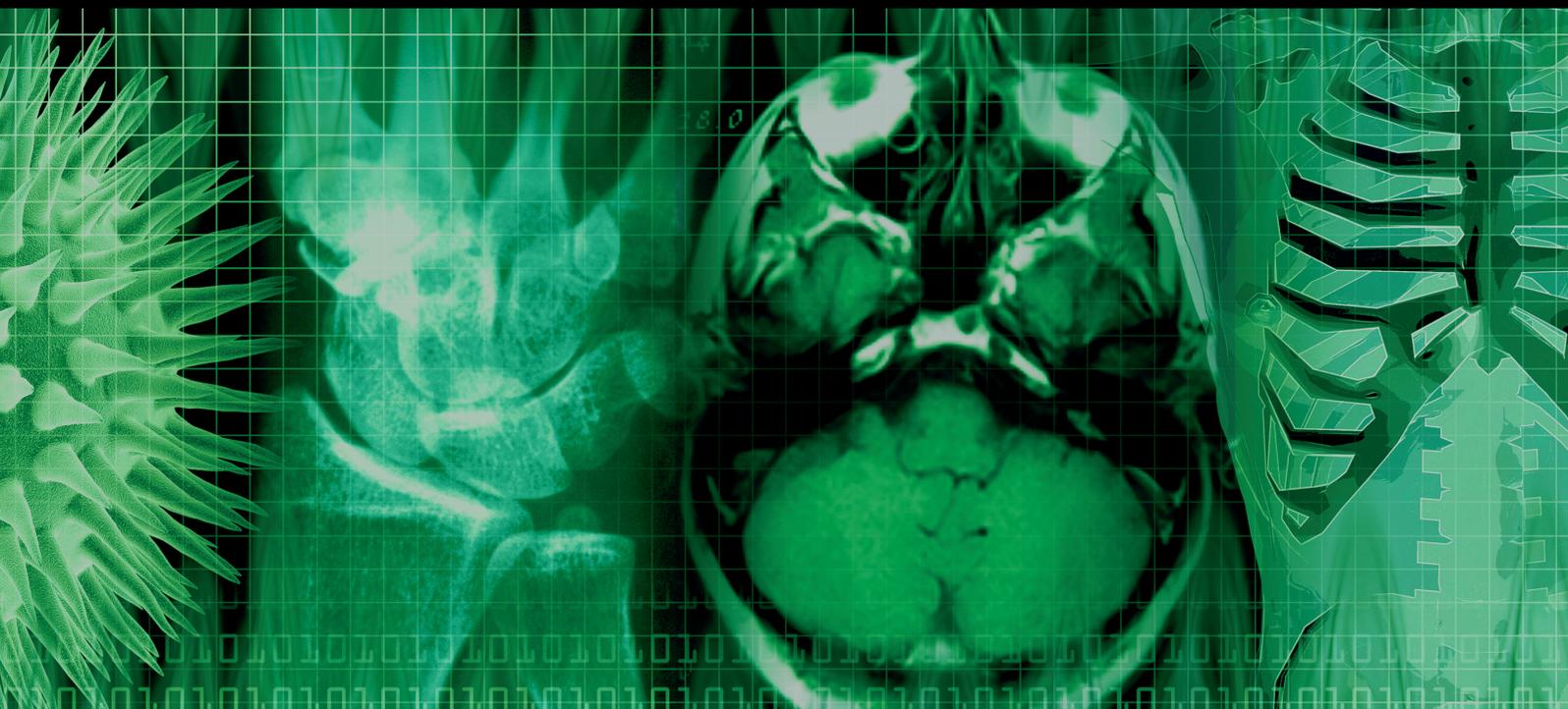


# Advanced Signal Processing Methods for Biomedical Imaging

Guest Editors: Juan Ruiz-Alzola, Carlos Alberola-López, and Carl-Fredrik Westin





---

# **Advanced Signal Processing Methods for Biomedical Imaging**

International Journal of Biomedical Imaging

---

## **Advanced Signal Processing Methods for Biomedical Imaging**

Guest Editors: Juan Ruiz-Alzola, Carlos Alberola-López,  
and Carl-Fredrik Westin



---

Copyright © 2013 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “International Journal of Biomedical Imaging.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Haim Azhari, Israel

K. Ty Bae, USA

Richard H. Bayford, UK

F. J. Beekman, The Netherlands

J. C. Chen, Taiwan

Anne Clough, USA

Carl Crawford, USA

Daniel Day, Australia

Eric Hoffman, USA

Jiang Hsieh, USA

M. Jiang, China

Marc Kachelrieß, Germany

Cornelia Laule, Canada

Seung W. Lee, Republic of Korea

A. K. Louis, Germany

Jayanta Mukherjee, India

Vasilis Ntziachristos, Germany

Scott Pohlman, USA

Erik L. Ritman, USA

Jay Rubinstein, USA

Peter Santago, USA

Lizhi Sun, USA

Kenji Suzuki, USA

Jie Tian, China

Michael W. Vannier, USA

Yue Wang, USA

Ge Wang, USA

Guo Wei Wei, USA

D. L. Wilson, USA

Sun K. Yoo, Republic of Korea

Habib Zaidi, Switzerland

Yantian Zhang, USA

Jun Zhao, China

Yibin Zheng, USA

Tiange Zhuang, China

Yu Zou, USA

# Contents

---

**Advanced Signal Processing Methods for Biomedical Imaging**, Juan Ruiz-Alzola, Carlos Alberola-López, and Carl-Fredrik Westin

Volume 2013, Article ID 696878, 2 pages

**A Novel Flexible Model for the Extraction of Features from Brain Signals in the Time-Frequency Domain**, R. Heideklang and G. Ivanova

Volume 2013, Article ID 759421, 12 pages

**Fast and Analytical EAP Approximation from a 4th-Order Tensor**, Aurobrata Ghosh and Rachid Deriche

Volume 2012, Article ID 192730, 9 pages

**The Smoothing Artifact of Spatially Constrained Canonical Correlation Analysis in Functional MRI**,

Dietmar Cordes, Mingwu Jin, Tim Curran, and Rajesh Nandy

Volume 2012, Article ID 738283, 11 pages

**A New GLLD Operator for Mass Detection in Digital Mammograms**, N. Gargouri, A. Dammak

Masmoudi, D. Sellami Masmoudi, and R. Abid

Volume 2012, Article ID 765649, 13 pages

**DCE-MRI and DWI Integration for Breast Lesions Assessment and Heterogeneity Quantification**, C.

Andrés Méndez, Francesca Pizzorni Ferrarese, Paul Summers, Giuseppe Petralia, and Gloria Menegaz

Volume 2012, Article ID 676808, 8 pages

## Editorial

# Advanced Signal Processing Methods for Biomedical Imaging

Juan Ruiz-Alzola,<sup>1,2</sup> Carlos Alberola-López,<sup>3</sup> and Carl-Fredrik Westin<sup>4</sup>

<sup>1</sup> Department of Signals and Communications, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

<sup>2</sup> Canary Islands Agency for Research, Innovation and Information Society, 35003 Las Palmas de Gran Canaria, Spain

<sup>3</sup> Image Processing Laboratory, ETSI Telecomunicación, University of Valladolid, 47011 Valladolid, Spain

<sup>4</sup> Laboratory for Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

Correspondence should be addressed to Juan Ruiz-Alzola; [jruiz@itccanarias.org](mailto:jruiz@itccanarias.org)

Received 23 December 2012; Accepted 23 December 2012

Copyright © 2013 Juan Ruiz-Alzola et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Signal processing is a long-established engineering field with broad applications in so different arenas such as multimedia (audio, still image, video, graphics, etc.), coding, communications, seismology, astronomy, biomedicine, artificial intelligence, and econometrics, to only name some of them. Signal processing stems from cross-fertilized grounds including mathematical analysis, algebra, numerical analysis, probability, statistics, information theory, discrete mathematics, cybernetics, and computer science. Advances in signal processing have frequently been pulled by needs in specific application domains, so that they are readily incorporated to the signal processing knowledge base for convenient use in distant areas of application.

Signal and image processing is ubiquitous in modern biomedical imaging, as it provides essential techniques for image construction, enhancement, coding, storage, transmission, analysis, understanding, and visualization from any of an increasing number of different multidimensional sensing modalities. As biomedical imaging is rapidly evolving, new and more powerful signal and image processing algorithms are required to meet the challenges imposed by modern huge multidimensional multimodal biomedical data, particularly in real clinical settings.

In this special issue, we present high-quality original research papers that address specific challenges for the biomedical imaging community and benefit of advanced and emerging methods in signal and image processing required for novel time-varying high-dimensional structural and functional biomedical imaging modalities. Interesting new ideas from other fields of application of signal and image processing have been highly welcome since their

practical relevance for biomedical imaging is clearly motivated.

In particular, this special issue features original research for different imaging modalities, such as generalized diffusion tensor imaging (GDTI), conventional diffusion weighted MRI (DW-MRI), dynamic contrast enhanced MRI (DCE-MRI), functional MRI (fMRI), X-ray mammography and time-frequency representations (TFR) of electrophysiological data such as electroencephalography (EEG), magnetoencephalography (MEG), or local field potentials (LFP). The medical application areas include central nervous system fiber tractography, analysis of episodic memory and motor brain activity, analysis of neural activity patterns, and breast cancer screening and detection. As for the signal and image processing methods, a brief list of topics addressed in this special issue include higher-order tensors, signal estimation and classification, clustering, and multivariate statistics, to name a few.

In “Fast and analytical EAP approximation from a 4th-order tensor,” A. Ghosha and R. Deriche deal with the complicated problem of inferring the microstructure of tissue or fiber bundles from DW-MRI. To this extent they propose a modified GDTI approach, where the higher-Order tensor (HOT) representation of the apparent diffusivity coefficient (ADC) is estimated, from the DW-MRI data, in such a way that they provide a closed-form approximation, using Hermite polynomials, to the ensemble average propagator (EAP) that overcomes the computational overload of other EAP estimation methods from GDTI. The EAP describes the probability of the diffusing particles, and, hence, the geometry of the EAP is a direct indicator of the microstructure of the

underlying tissue or fiber bundles. They provide experimental results on the fiber bundles estimated from real and synthetic brain datasets.

In “*DCE-MRI and DWI integration for breast lesions assessment and heterogeneity quantification*,” C. A. Méndez et al. address the quantification of breast tumor heterogeneity using jointly DW-MRI and DCE-MRI, which provide a measure of cellularity and an indication of blood volume, flow, and vascular permeability, respectively. To that end, they first make an affine followed by an elastic registration of both datasets and approach the tissue segmentation by clustering using a dissimilarity-based representation (DBR) of the DCE curves and the ADC maps, followed by K-means. Statistical testing is carried out for the ADC maps of the resulting clusters. They provide experimental results from 21 patients with primary ductal carcinoma.

In “*The smoothing artifact of spatially constrained canonical correlation analysis in functional MRI*,” D. Cordes et al. address the problem of the spatial specificity of the activation signal in fMRI due to smoothing artifacts that arise in an extension of the canonical correlation analysis (CCA) termed constrained CCA (cCCA), which is used as a statistic in fMRI to test for functional brain activation in a specific neighborhood. In particular, they investigate in detail the smoothing artifact that is associated with each spatial constraint in cCCA and provide a novel approach in order to correct the measure of activation for the smoothing artifact within a Bayesian testing framework. They also provide experimental results on six real cases with six normal subjects carrying out certain motor and episodic memory tasks.

In “*A new GLLD operator for mass detection in digital mammograms*,” N. Gargouri et al. propose an extension to the local binary pattern (LBP) operator used for texture classification in order to overcome some limitations that arise from the fact that LBP uses local gray-level differences. In particular they introduce the gray level and local Difference (GLLD) representation that, in addition to differences, also uses gray levels and investigate the efficiency of the GLLD-based approach as a method of feature extraction. They tackle the detection of abnormal masses on breast X-ray images by performing a manual region of interest (ROI) delineation, followed by the GLLD representation and standard classification by k-NN, support vector machines, and multilayer perceptrons. Experimental results are from 1000 ROIs from the Digital Database for Screening Mammography.

Last but not least, in “*A flexible model for feature extraction from brain signals in the time-frequency domain*,” R. Heideklang and G. Ivanova provide a different perspective since they deal with electrophysiological data. Images arise from the time-frequency representation (TFR) of these nonstationary signals that are obtained in the paper from the smoothed pseudo-Wigner-Ville distribution. A common problem of TFRs from electrophysiological data is inter- and intrasubject variability of the features to be extracted (curves in the TFR), which is here addressed by proposing the smooth natural Gaussian extension (snaGe) model, characterized by a sequence of multivariate Gaussians located on a parametric curve described by a cubic B-spline. By combining cubic B-splines with the Gaussian shape, they obtain a “sufficiently

smooth model which inherits both the splines’ flexibility and the Gaussian standard model’s robustness.” In order to fit the model to the data, after subsampling, an initial estimation of the initial parameters of the curve is obtained by dynamic programming, which is refined through iteration, and using Fréchet and other metrics for curve difference. The authors present results on real and synthetic EEG data, and they discuss the robustness and the sensitivity of the method to noise.

We hope that the readers enjoy this special issue.

Juan Ruiz-Alzola  
Carlos Alberola-López  
Carl-Fredrik Westin

## Research Article

# A Novel Flexible Model for the Extraction of Features from Brain Signals in the Time-Frequency Domain

R. Heideklang and G. Ivanova

Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Correspondence should be addressed to G. Ivanova; [givanova@informatik.hu-berlin.de](mailto:givanova@informatik.hu-berlin.de)

Received 20 July 2012; Revised 28 September 2012; Accepted 21 November 2012

Academic Editor: Juan Ruiz-Alzola

Copyright © 2013 R. Heideklang and G. Ivanova. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electrophysiological signals such as the EEG, MEG, or LFPs have been extensively studied over the last decades, and elaborate signal processing algorithms have been developed for their analysis. Many of these methods are based on time-frequency decomposition to account for the signals' spectral properties while maintaining their temporal dynamics. However, the data typically exhibit intra- and interindividual variability. Existing algorithms often do not take into account this variability, for instance by using fixed frequency bands. This shortcoming has inspired us to develop a new robust and flexible method for time-frequency analysis and signal feature extraction using the novel *smooth natural Gaussian extension (snaGe)* model. The model is nonlinear, and its parameters are interpretable. We propose an algorithm to derive initial parameters based on dynamic programming for nonlinear fitting and describe an iterative refinement scheme to robustly fit high-order models. We further present distance functions to be able to compare different instances of our model. The method's functionality and robustness are demonstrated using simulated as well as real data. The *snaGe* model is a general tool allowing for a wide range of applications in biomedical data analysis.

## 1. Introduction

Electrophysiological brain signals are widely studied to get insights into the inner function of the brain. The electroencephalogram (EEG), as an example, has been analyzed for decades and is particularly popular because of its noninvasiveness, wide availability, relatively small cost and excellent temporal resolution which enables capturing the fast neural dynamics. Because it is known that electrophysiological brain signals exhibit important spectral characteristics, frequency transforms are often applied. However, since in general brain signals do not possess the statistical property of stationarity, time-frequency transforms are of special interest. Such methods are able to represent a given signal jointly in the time-frequency domain, called its *time-frequency representation (TFR)*. Thereby the signal's spectral components can be analyzed in relation to its temporal dynamics. In a wide sense, TFRs can be interpreted as images containing complex pixel intensities in general.

An important feature of biological signals, and particularly brain signals, is their inter- and intraindividual

variability. That is, under fixed experimental conditions, the obtained signals exhibit heterogeneity not only between groups of subjects, but also between subjects within the same experimental group and even within the same subject between multiple experimental trials. Existing signal analysis techniques either do not take this issue into account adequately or usually treat it by defining frequency bands of interest rather than single frequencies, and similarly time intervals instead of sharp instants. However, this strategy requires a priori knowledge about the variability to appropriately set the interval widths, and it is imprecise because it blindly includes all information contained in that time/frequency region.

A good example is the time-frequency coherence analysis of two given input signals, for instance, by means of the cross short-time Fourier transform [1], the cross Wigner-Ville distribution [2], or the wavelet coherence [1, 3]. All of these methods relate both signals at fixed time/frequency (or scale) locations. Thus, if signal A exhibits the same neural activation as signal B, but signal A's pattern is shifted in frequency just a little, none of the abovementioned techniques will be able

to find the strong similarity of A and B. Although coherence estimation and other rigid strategies have been successfully applied “for more than 30 years” [4], this issue has inspired us to develop a general flexible method of *pattern analysis* and corresponding feature extraction in electrophysiological TFR data.

By abstracting from the TFR images and working with the representation of a TFR pattern, numerous applications in biomedical signal processing emerge. TFR patterns quantify neural activity and therefore extract useful features for subsequent analyses. The model proposed in this work goes even further by offering interpretable parameters. TFR patterns reduce dimensionality by representing neural activity in a wide spectrotemporal region by comparably few quantities. Pattern-based outlier detection has the potential to become a useful tool for data quality assurance. In ongoing studies we are employing the presented method, for instance, to estimate functional brain connectivity by means of a pattern-based approach.

In the following, we present the developed neuroinspired interpretable model which is able to capture general time-frequency patterns. We use solely EEG data for demonstrations here, but our method is applicable to general electrophysiological signals or even to other signals showing similar behavior. Section 2 is devoted to developing our idea by extending the multivariate Gaussian model. Algorithms for robustly fitting the novel model to time-frequency representations are presented. A strategy for finding an appropriate model order is given, and distance functions are defined which quantify (dis-)similarity of two given models. These methods are tested in Section 3, where real as well as simulated data are used to demonstrate our technique’s functionality and robustness.

## 2. Methods

While our technique is not restricted to specific time-frequency distributions, we employ the *smoothed pseudo Wigner-Ville distribution* [5] in this work. This is a quadratic transform estimating signal power in the time-frequency domain, whereby all quantities in this work are real numbers. The transform generates quite smooth TFRs, which means that neighboring pixel values are correlated. Although only positive values can be interpreted as signal power, the Wigner-Ville distribution introduces also negative values in general [6]. These data properties will be taken into account by our method.

We will refer to time-frequency representations as mappings  $y^{(\text{TFR})} : T \times F \rightarrow \mathbb{R}$  which estimate signal power for each point in the time-frequency domain.

Of the numerous ways to quantify TFR patterns, we choose to fit a parametric surface to the data. Because of the spatial correlation inherent in the data, traditional regression assumptions about independence of observations do not hold here [7]. This absence of strong gradients in the TFR images also invalidates most image feature extraction techniques, which are often based on edges and texture [8]. Our model, however, is especially designed for spatially correlated data; furthermore its parameters are interpretable. These quantities

are useful features which embody important information about the underlying signal and thereby considerably reduce data dimensionality. Using our method, feature extraction can be fully automated, and no training data are necessary. Nevertheless, a priori information can be incorporated fairly easily.

In the following, we propose an extension of the well-known Gaussian model for TFR analysis.

**2.1. The Gaussian Model.** The Gaussian model for multivariate data  $x \in \mathbb{R}^n$  is defined by

$$y^{(\text{G})}(x) = C + A \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (1)$$

with  $C \in \mathbb{R}$  being a constant additive offset,  $A \in \mathbb{R}$  the amplitude relative to the offset,  $\mu \in \mathbb{R}^n$  the constant  $n$ -dimensional mean vector, and  $\Sigma$  denoting a  $n \times n$  symmetric positive definite matrix. Positive definiteness ensures that the argument to the exponential function is always negative; additionally we know that  $e^x$  is bounded by zero and one for negative  $x$ . Therefore, the exponential factor scales the final amplitude between 0 and  $A$  relative to the offset  $C$ . The term  $(x - \mu)^T \Sigma^{-1} (x - \mu)$  is also known as the squared Mahalanobis distance of  $x$  with respect to  $\mu$  and  $\Sigma$ .

Because in our context this function represents arbitrary data in contrast to statistical distributions,  $\mu$  will also be called the *position vector*, and  $\Sigma$  is the *spread matrix*, its entries  $\sigma_{ij}$  are denoted *spread parameters*.

Gaussian models are quite *robust* in various ways. Firstly, the model will be shaped like a peak for all possible parameter values by imposing the constraint that  $\Sigma$  (and thus also  $\Sigma^{-1}$ ) is symmetric positive definite. Thereby, the model will never be able to completely “degenerate.” Because the model is not flexible enough to fit small local variations of an expected pattern, the Gaussian model is relatively insensitive to local data outliers and is also unsusceptible to overfitting. An additional aspect of robustness is that extreme peak deformations are directly reflected in extreme parameter values. Thereby degenerated models can be easily detected or may even be prevented by imposing parameter constraints.

**2.1.1. Interpretability.** The Gaussian model is well-suited to extract bivariate peaks from brain signals’ TFR data, reflecting short intervals of neural excitement in a specific frequency range. An instance of the above described surface, in the bivariate case, is fully identified by its parameter vector

$$p = (C, A, \mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T. \quad (2)$$

The absolute peak height  $C + A$ , the peak position  $(\mu_1, \mu_2)^T$ , and the peak orientation can be derived from the parameter vector. Further relevant quantities are the temporal peak onset, peak offset, and peak duration (as the difference of the previous two).

**2.2. Extending the Gaussian Model: The snaGe Model.** As already mentioned, the Gaussian model’s robustness comes at the cost of inflexibility. While some local effects in TFR data

can be appropriately explained by (1), more general patterns of activation do not follow peak-like shapes, as will be shown later. Thus a generalization of the Gaussian model would be desirable, especially concerning the ability to represent *patterns of activation* rather than just independent “events” in the spectrotemporal domain. At the same time, a generalized method should maintain maximum robustness in order not to degenerate easily and to prevent overfitting. The so-called Gaussian mixture modeling (GMM) is a straightforward extension [9], but this method still assumes (multiple-) peak-shaped data. In the following, the smooth natural Gaussian extension (*snaGe*) model is presented as a flexible extension of the multivariate Gaussian model.

Before giving a formal definition, we explain the idea in an intuitive way, guided by Figure 1. The  $n$ -variate Gaussian model can be described by its  $(n + 1)$ -dimensional peak point  $P = (\mu_1, \dots, \mu_n, A)$  and its spread parameters controlling the exponential flattening relative to  $P$  along the independent variables’ dimensions. Now the idea is to not use only one, but  $K$  peak points  $P^{(i)} \in \mathbb{R}^{n+1}$ ,  $i = 1 \dots K$ , interpolated by a smooth  $(n + 1)$ -dimensional curve of peaks. A way to think of the surface in Figure 1 modeled by *snaGe* is to “shape” it by sliding an  $n$ -variate Gaussian model along the curve, its peak point being connected to the curve and thus varying in height ( $A$ ) and position (vector  $\mu$ ). Thereby complex smoothly “bent” patterns of data with varying amplitude (dependent variable) can be captured. The term *snaGe* is inspired by these snake-like forms. Analogously to the Gaussian model, an  $n \times n$  spread matrix determines the model’s shape, which is a surface for  $n = 2$ .

The tradeoff between robustness and flexibility can be controlled by choosing the number of peak points  $K$ . Using many peak points will allow for good fits to complex patterns but will also increase the danger of overfitting. Choosing a small  $K$  yields a robust model, but its ability to capture complex patterns will be limited. By setting  $K = 1$ , *snaGe* reduces to the traditional Gaussian model as a special case.

Regarding the standard Gaussian model, each function value  $y^{(G)}(x)$  is fully determined by the Mahalanobis distance of the point  $x$  to the unique mean point  $\mu$ . But since the *snaGe* model offers infinitely many mean points, the question arises how to calculate the surface values. This issue will be addressed in the next section, where a formal definition is given.

**2.2.1. Formal Definition.** Let the “number of peak points” be denoted by  $K \in \mathbb{N}$ ,  $K \geq 1$ . Let  $\mu : [1, K] \rightarrow \mathbb{R}^n$  denote a smooth  $n$ -dimensional curve of “means,” and let  $A : [1, K] \rightarrow \mathbb{R}$  be a smooth one-dimensional curve of “amplitudes” along  $\mu$ . Let further the “offset”  $C \in \mathbb{R}$ , and let  $\Sigma \in \mathbb{R}^{n \times n}$  be a symmetric positive definite matrix. Define

$$\begin{aligned} \tilde{y}^{(snaGe)} : \mathbb{R}^n \times [1, K] &\longrightarrow \mathbb{R} \\ \tilde{y}^{(snaGe)}(x, u) &:= C + A(u) \exp\left(-\frac{1}{2}(x - \mu(u))^T \right. \\ &\quad \left. \times \Sigma^{-1}(x - \mu(u))\right). \end{aligned} \quad (3)$$

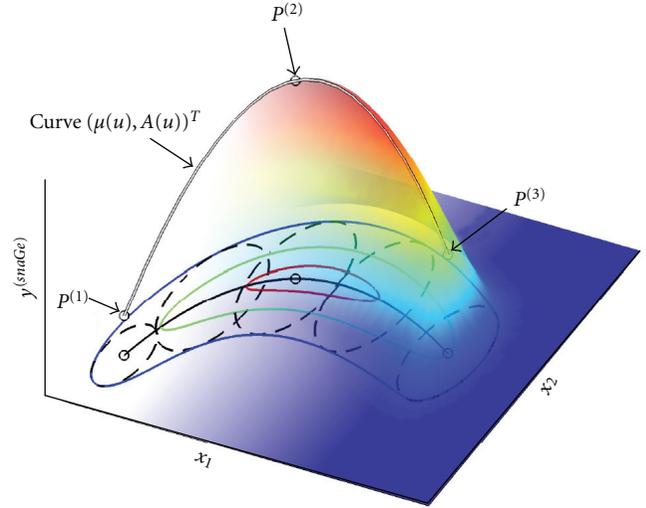


FIGURE 1: Example of an instance of the *snaGe* model, in the bivariate case. A surface plot is shown as well as its two-dimensional projection (colored contours).  $K = 3$  three-dimensional points were smoothly interpolated to yield a “curve of peaks” (curve connecting the circles). This curve’s 2d projection is  $\mu(u)$ , the black line. A surface is determined by the spread parameters (dashed 2d ellipses), controlling the shape of the exponential flattening to both sides of the curve.

Note that  $\tilde{y}^{(snaGe)}$  is a family of traditional Gaussian models, parameterized by the curve parameter  $u$ . In order to construct a function which is independent of  $u$ , we define

$$u^*(x) := \arg \max_{u \in [1, K]} \left| \tilde{y}^{(snaGe)}(x, u) - C \right|. \quad (4)$$

The *snaGe* model is then given by

$$\begin{aligned} y^{(snaGe)} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ y^{(snaGe)}(x) &:= \tilde{y}^{(snaGe)}(x, u^*(x)). \end{aligned} \quad (5)$$

The function  $u^*(x)$  defines that traditional Gaussian model which assigns to  $x$  the largest absolute amplitude relative to the offset  $C$  among all members of the Gaussian family  $\tilde{y}^{(snaGe)}$ . This is necessary to cope with positive as well as negative  $A(u)$ . In TFR analysis,  $A(u)$  can be restricted to only positive values (see Section 2.3.1), in which case (5) in fact simplifies to

$$y^{(snaGe)+}(x) := \max_{u \in [1, K]} \tilde{y}^{(snaGe)}(x, u). \quad (6)$$

For  $y^{(snaGe)}$  as well as  $y^{(snaGe)+}$  the  $\max()$  function is necessary in case  $\mu(u)$  is a “near self-intersecting” curve, in the sense that  $\|\mu(u_1) - \mu(u_2)\|$  is small, but  $|A(u_1) - A(u_2)|$  is large. Figure 2 illustrates such an exemplary scenario.

We assume that the diagonal entries of the spread matrix  $\Sigma$ , that is, the spread along each dimension, are sufficient to control a TFR pattern’s “width.” Therefore, we fix offdiagonal entries to zero for the sake of robustness. Thereby, the Mahalanobis distance in (3) reduces to the weighted Euclidean distance.

The two curves  $A(u)$  and  $\mu(u)$  are yet to be defined in terms of discrete parameter values. For good parameter interpretability, we choose to form both curves by interpolating  $K$  points  $P^{(i)} \in \mathbb{R}^{n+1}$ ,  $i = 1 \dots K$  by B-splines of degree  $\leq 3$ , which yields the curve:

$$P : [1, K] \longrightarrow \mathbb{R}^{n+1}, \quad P(u) = \begin{pmatrix} P_1(u) \\ \vdots \\ P_n(u) \\ P_{n+1}(u) \end{pmatrix} = \begin{pmatrix} \mu_1(u) \\ \vdots \\ \mu_n(u) \\ A(u) \end{pmatrix} = \begin{pmatrix} \mu(u) \\ A(u) \end{pmatrix}. \quad (7)$$

By combining cubic B-splines with the Gaussian shape, we obtain a sufficiently smooth model which inherits both the splines' flexibility and the Gaussian standard model's robustness. Our model further inherits the B-splines' *local control* property; that is, varying a  $P^{(i)}$  will affect the model only in the  $P^{(i)}$ 's vicinity. Additionally, the degree of flexibility can be adapted to the data at hand by varying  $K$  from 1 (single Gaussian peak) to arbitrary flexibility with  $K > 1$ .

An instance of the  $n$ -variate *snaGe* model with  $K$  points (or of order  $K$ ) is fully represented by the parameter vector of length  $1 + n + (n + 1)K$ :

$$p = \left( C, \sigma_{11}, \sigma_{22}, \dots, \sigma_{nn}, P_1^{(1)}, \dots, P_{n+1}^{(1)}, \dots, P_1^{(K)}, \dots, P_{n+1}^{(K)} \right)^T. \quad (8)$$

While the offset  $C$  and the spread parameters are mainly responsible for the prediction of data values, the curve  $P(u)$  interpolating the  $P^{(i)}$  is directly interpretable as it models the main *path of peaks* in the data.

In the next section we show how to fit the model to data in a robust manner.

**2.3. Fitting the Model to TFR Data.** Given a time-frequency representation  $y^{\text{TFR}}(x)$ ,  $x \in T \times F$ , we aim to find a parameter vector  $p$  so that the respective model  $y^{(\text{snaGe})}$  fits the data "best," in the sense that it minimizes a cost function. We use the sum of squared differences of the data and the modeled surface:

$$\text{SSE}(p) = \sum_{\substack{x=(t,f)^T \\ t \in T, f \in F}} \left( y^{\text{TFR}}(x) - y^{(\text{snaGe})}(x) \right)^2. \quad (9)$$

The parameters  $p$  are implicitly represented by  $y^{(\text{snaGe})}$  in the above formula. This quantity is also called the "sum of squares due to error" which is zero if the model perfectly fits the data. SSE is a nonlinear function dependent on  $p$ . Using squared differences pronounces outliers, but these are not expected to occur frequently in our smooth TFR data. In order to find a locally minimum solution of SSE, a nonlinear least squares algorithm implemented in the

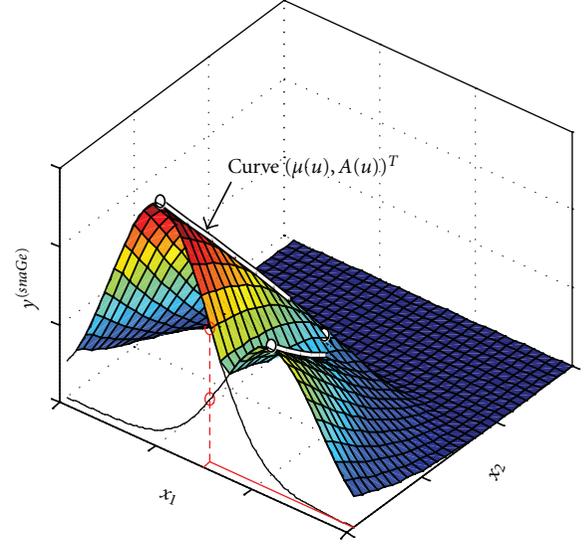


FIGURE 2: Depiction of the maximum rule. The curve of means  $\mu(u)$  is sharply bent. The surface value for a point  $(x_1^*, x_2^*)$  (coordinates depicted by red lines) is chosen among all possible traditional Gaussians (black graphs visualize 1D cuts of two of these) by the maximum rule (red circles), see (6).

MATLAB Optimization Toolbox, *lsqnonlin* [10], is employed. Given an initial parameter vector and the cost function SSE, this optimizer produces a sequence of models  $p_i$ . The iteration hopefully converges to a  $p^*$  with minimum cost, that is, best resemblance between model and data.

We attempt to provide advantageous starting conditions for the optimizer by preprocessing the TFR and by obtaining an initial parameter vector  $p_0$  which is expected to be close to the optimum with respect to SSE. Moreover, an iterative refinement scheme is proposed to be able to robustly fit models of high order. The process of fitting is outlined by Figure 4.

**2.3.1. TFR Preprocessing.** TFR data are badly scaled, showing differences of several orders of magnitude in values of time, frequency, and signal power, which affects optimization performance [11]. To address this problem, *lsqnonlin* offers a way to take into account typical values for each dimension for gradient estimation. Also concerning this issue, any Euclidean distance operating on TFR data in our algorithms is weighted appropriately. Furthermore, smooth objective functions are desirable so that the low-order Taylor approximations used during optimization resemble the cost function in a relatively large neighborhood around the current point. To this end, the TFR images are smoothed and subsampled, which has the additional benefit of faster cost function evaluations. Finally, since negative values in the time-frequency domain are not interpretable, they are usually set to zero.

**2.3.2. Estimation of Initial Parameters.** Generally, nonlinear cost functions may exhibit multiple local extreme values. Since *lsqnonlin* performs local optimization, a start parameter vector should be chosen such that it already lies in the basin

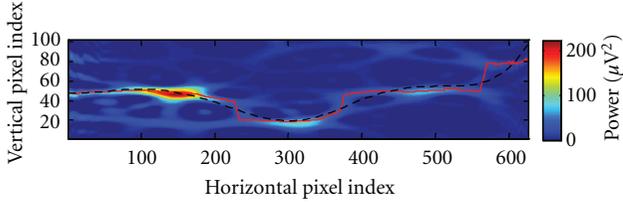


FIGURE 3: Illustration of an optimal horizontal image path (red line) found by a dynamic programming algorithm. The sum of signal power values along the red line is higher or equal to that of any other horizontal path. The trajectory is smoothed (black) for the robust extraction of the initial peak points' time-frequency coordinates. The background TFR image is computed from simulated data, whose three consecutive peaks of activity are correctly connected by the path.

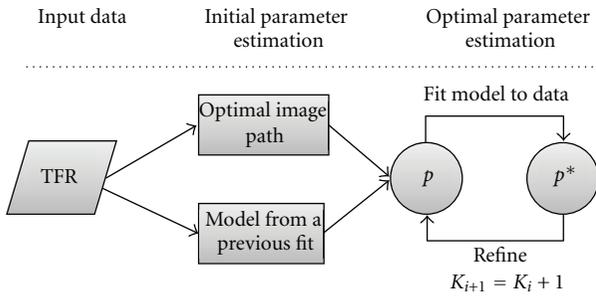


FIGURE 4: Flow chart of the fitting process. Initial parameters are derived from the TFR data, or a previously fitted model is used to determine  $p$ . The model is fitted using a nonlinear optimization algorithm. If desired, the optimal model can be iteratively refined up to  $K_{\max}$ .

of the cost function's (unknown) global minimum, that is, a vector whose cost is already low. To this end, the constant offset and the spread parameters are initialized to  $C = 0$ ,  $\sigma_{11} = (t_{\max} - t_{\min})/5$  and  $\sigma_{22} = (f_{\max} - f_{\min})/5$  if the underlying TFR's time and frequency axes are bounded by  $t_{\min}$ ,  $t_{\max}$  and  $f_{\min}$ ,  $f_{\max}$ , respectively. One may choose any sensible values alike, but the spreads should not be initialized too small in order to obtain a generalizable model. Yet, more thought has to be put in choosing the number and coordinates of the peak points  $P^{(i)}$ . In the following we propose a way to compute initial estimates of the time, and frequency coordinates of all  $P^{(i)}$  directly from the data.

A reasonable approximation to the unknown optimal curve of peaks can be found by tracing a path through the TFR image  $y^{(\text{TFR img})}$  from left to right, which runs through areas of high pixel intensity. More precisely, the sum of all pixel intensities along this path should be as high as possible. This is an optimization problem in turn, yet its solution can be computed in quadratic time complexity (provided that the path's slope is bounded) by a dynamic programming algorithm, see [12]. Because a global optimum is guaranteed to be found, this strategy is insensitive to local outliers and noise. To this end, a similar approach as described in [13] is employed. Following the notation therein, we define our

energy function to be equal to the TFR values themselves, that is,  $e(I) = I$ . A horizontal path  $s^*$  (called *seam* in [13]) which maximizes (this is in contrast to [13], where *minimum* energy seams are computed) this simple cost function is found by dynamic programming. See Figure 3 for an example. Additionally, the paths are constrained by imposing an upper bound  $k$  on their slopes. This value depends on the time-frequency resolution here and once more represents a compromise between robustness and flexibility.

Given the found path  $s^*$  and the desired model order,  $K$  evenly spaced samples are subsequently drawn from a smooth approximating curve to obtain estimates of the first two coordinates of the  $P^{(i)}$ ,  $i = 1 \dots K$ . We choose to empirically set the  $P^{(i)}$ 's last components, interpretable as amplitudes relative to the initial constant offset  $C = 0$ , to  $\max(y^{(\text{TFR})}) - C = \max(y^{(\text{TFR})})$ .

Once a parametric representation of the data is available, its accuracy can be improved in a step-wise manner, as is presented in the following section.

**2.3.3. Iterative Refinement.** As already stated, the number of points  $P^{(i)}$  controls the model's robustness which complements its ability to resemble complex patterns. Therefore, the demand for a near-optimal initial parameter vector increases with the model order  $K$ . Employing the optimal image path method described in the previous section yields a "reasonable" estimation, but sampling  $K$  equidistant points  $P^{(i)}$ ,  $i = 1 \dots K$  from the resulting curve is a simplification. In fact, it can be observed that the optimizer tends to concentrate the  $P^{(i)}$  in time-frequency regions of high signal variability. For low model order  $K$ , this shortcoming of our initial parameter estimation algorithm can be compensated easily by the optimization algorithm, but it may become a problem for increasingly flexible models. For this reason we propose an iterative scheme.

- (1) Find initial parameters  $p_0$  by means of an optimal path (see Section 2.3.2) for a first, robust model of low order  $K_0$ . Let  $i := 0$ .
- (2) Fit the model to the data to obtain optimal parameters  $p_i^*$ .
- (3) Construct the optimal curve of peaks  $P(u)$  by interpolation of the peak points (see Section 2.2.1).
- (4) Obtain the  $K_{i+1} = K_i + 1$  peak points for a refined model by uniformly sampling (with respect to the spline's sites  $u_j$ ) the curve computed in the previous step.
- (5) Construct the parameter vector  $p_{i+1}$  of the refined model from  $p_i$  by replacing the old  $K_i$  peak points with the  $K_{i+1}$  new ones.
- (6) If  $K_i < K_{\max}$ , let  $i := i + 1$  and continue with step 2.

The curve found in steps 2 and 3 will exhibit smaller gradient magnitude, that is, traversal speed, in areas of high signal variability than in other regions. We aim at maintaining the curve-defining points' optimum distribution found by the fitting algorithm and at enhancing the model's

flexibility mainly in these areas. It turns out that simply by uniformly sampling the fitted curve (step 4) we obtain a new interpolated curve which retains these properties.

An application of this algorithm is demonstrated in Section 3.1, where the resulting sequence of nested models is evaluated.

**2.4. Model Distance.** In this section we propose two functions for calculating the distance between two models  $y_1^{(snaGe)}, y_2^{(snaGe)}$  regarding (dis-)similarity of shape. Distance measures are necessary, for instance, to quantify how well the data exhibit an expected pattern. We will also employ these functions to assess our model's robustness.

Distance functions which are based solely on the curve of peaks  $P(u)$  were found to be quite effective. Other possibilities include parameter vector distances and pixel-wise differences of signal power of the models' generated data. By comparison, curve-based distance functions have the advantage of being able to interrelate models of different orders. Additionally, they are not influenced by the less informative parameters (offset  $C$  and the entries of  $\Sigma$ ).

A popular distance measure for parametric curves is the *Fréchet distance* [14]. In the continuous case, the Fréchet distance of two parametric curves  $P_1(u)$  and  $P_2(u)$  is defined by

$$D_{F\_max}(P_1, P_2) = \inf_{\alpha, \beta} \max_u d(P_1(\alpha(u)), P_2(\beta(u))). \quad (10)$$

Here,  $\alpha(t)$  and  $\beta(t)$  are monotone reparameterizations of the two curves, and  $d(\cdot)$  denotes (weighted) Euclidean distance. In words, we search for those reparameterizations which make the curves the most similar with respect to maximum point-wise Euclidean distance along the curves. This maximum for these reparameterizations is returned as the two curves' continuous Fréchet distance. In practice, the *discrete* Fréchet distance is frequently applied, whose computation is based on dynamic programming once more [15]. In the discrete case, an additional distance function  $D_{F\_sum}(P_1, P_2)$  can be obtained by replacing the max function with a *sum* over  $u$ . That way,  $D_{F\_sum}$  represents an average distance, being less prone to outliers in the curves.

### 3. Results

**3.1. Real Data.** We demonstrate the workflow to determine the appropriate model order  $K$  by fitting a TFR of real EEG data in this section. Typically we determine the necessary model complexity by fitting data with good signal to noise ratio (SNR) in order to prevent the overestimation of  $K$ . For example, one possibility to achieve sufficient data quality is to average several TFRs which are expected to show similar patterns. The averaged TFR of real EEG data shown in Figure 5 will guide the following explanations. The depicted brain signals located in the lower frequency bands were recorded from the temporal brain region during a face recognition experiment. These data exhibit a pattern of activity which is too complex to be captured by a traditional Gaussian model.

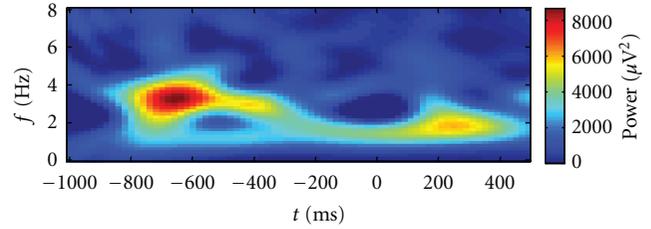


FIGURE 5: TFR of real EEG data which is fitted by iteratively refined models. Multiple local peaks are visible which are connected by a path of increased activity. This forms a complex pattern.

The iterative scheme described in Section 2.3.3 is employed to fit models of increasing flexibility to the high-quality data. An optimal path (see Section 2.3.2) estimates the initial parameters for the first, least flexible model of order  $K_0$ . A minimum value of  $K_0 = 3$  is necessary to model bent patterns. Since the appropriate  $K$  is still unknown, a sufficiently large number  $K_{max} = 7$  is chosen for the refinement. At each refinement stage the respective model is evaluated, and in the end the most suitable

$$K^* \in \{K_0, \dots, K_{max}\}, \quad K_0 \leq K^* \leq K_{max} \quad (11)$$

is chosen as the model order for future fittings on lower-quality data. Model evaluation is realized by three measures. These are the cost function value (SSE, see Section 2.3), the coefficient of determination  $R^2$  and its adjusted version  $R_{adj}^2$  [16]. Although the use of quantities based on the coefficient of determination is discouraged for nonlinear models [17], they are applied here nonetheless for two reasons. They are found to perform well for our purposes, and the proposed alternatives (AIC [18] and BIC [19]) are not easily applicable here. This is because the assumption of normally distributed residuals often does not hold, which is supported by a highly significant Shapiro-Wilk test [20] at  $\alpha = 5\%$  yielding  $p < 10^{-3}$  for this experiment.

Figures 6, 7, and 8 illustrate the results.

The results show that for the data at hand a model of order  $K^* = 4$  is sufficient to capture the variability.

Having determined the maximum model complexity on high-quality data, such a model can now be fitted to the rest of the data. If the TFRs are not expected to vary substantially, like when fitting a model to signals from several nearby sensors, a previous fit may serve as the initial model. However, if, for instance, multiple data segments of the same sensor should be fitted, the TFRs' patterns may vary strongly. In this case, initial parameters should be chosen depending on the data by using the method of optimal paths described in Section 2.3.2. Since in this example  $K^* = 4$  is a quite moderate number, the iterative refinement may also be skipped. However, in general we would start with  $K_0 = 3$  or  $K_0 = 4$  and refine up to the determined  $K^*$ , as proposed in Section 2.3.3.

**3.2. Synthetic Data.** We want to assess our model's robustness by simulating data and measuring how strongly the model is affected by additive Gaussian noise. To this end, artificial data

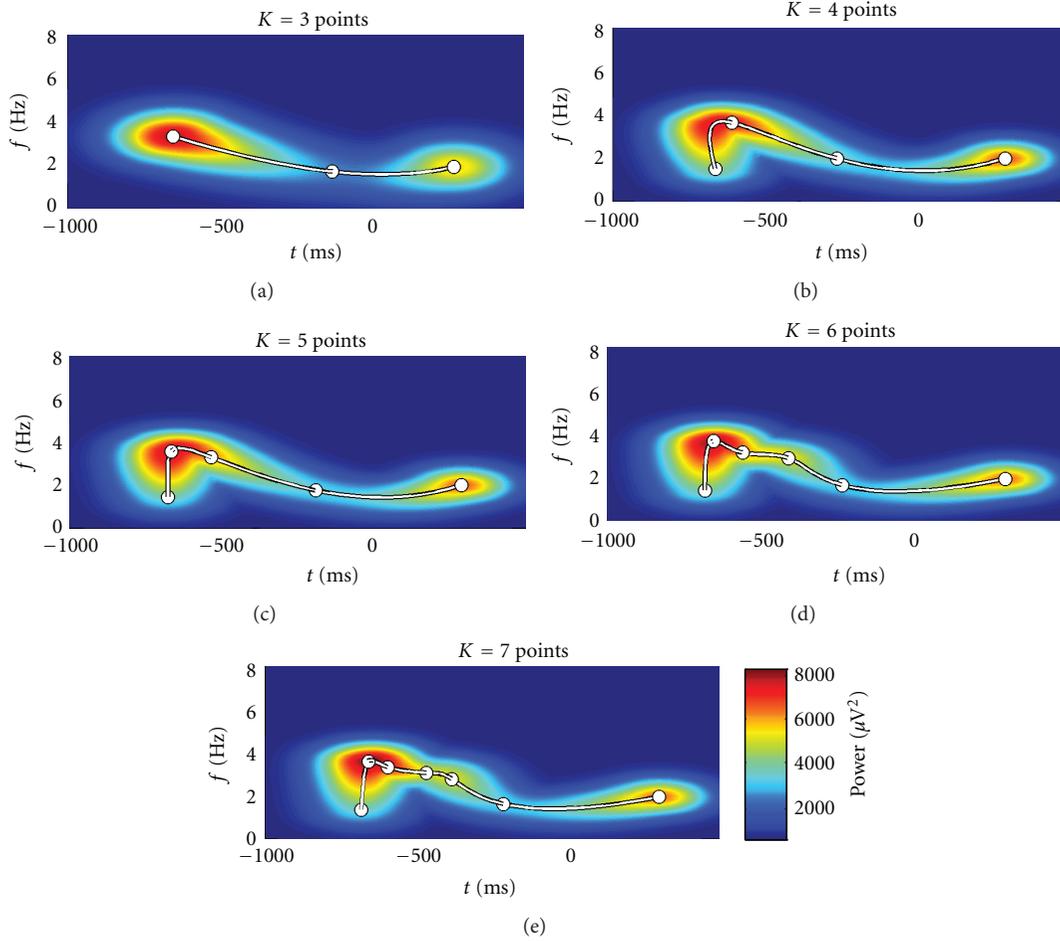


FIGURE 6: Surfaces of 5 iteratively refined models. Increasing the model order  $K$  clearly enhances the models' flexibility.

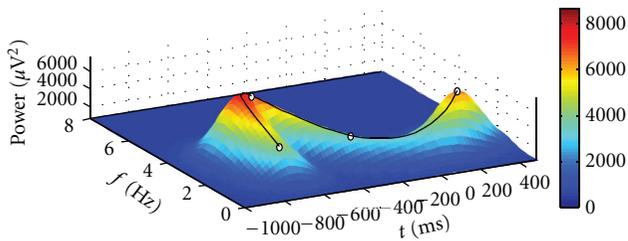


FIGURE 7: Three-dimensional plot of the fitted model of order  $K = 4$ . The black line represents the curve of peaks.

are created in the time domain, and their TFRs are computed to which our model will be fitted.

**3.2.1. Description of the Simulated Data.** We created a signal consisting of three consecutive oscillations, representing an alpha-theta-alpha EEG pattern at 10 Hz/4 Hz/10 Hz respectively over a time span of 2.5 seconds. The simulated sample rate is 250 Hz. A plot is shown in Figure 9. These data are quite challenging for our model because three distinct peaks

emerge in the TFR which could be more appropriately modeled by a mixture of independent Gaussian peaks. However, we want to demonstrate the flexibility of the *snaGe* model which should also be able to cope with patterns of this form.

For the following experiments we chose to start the fitting with a model of order  $K = 5$  to account for the pattern's complexity and perform one refinement step. Initial parameters are estimated by finding optimal paths, which means that no a-priori information about the known optimal model is passed to the fitting procedure other than the number of  $P^{(i)}$  to use. We define the optimal model by fitting the noise-free simulation in the same way. The distance measures from Section 2.4 are used to determine how well the simulated pattern is found.

**3.2.2. Noise Experiment.** In this experiment we added Gaussian noise, which is appropriately filtered with respect to the sampling frequency, to the simulated data in the time domain. Signals exhibiting signal to noise ratios of  $-15$  dB up to  $+10$  dB were generated in steps of 2.5 dB. At each SNR, ten distinct noise realizations are created to obtain representative results. This independent noise in the time domain will produce correlated noise in the time-frequency domain due

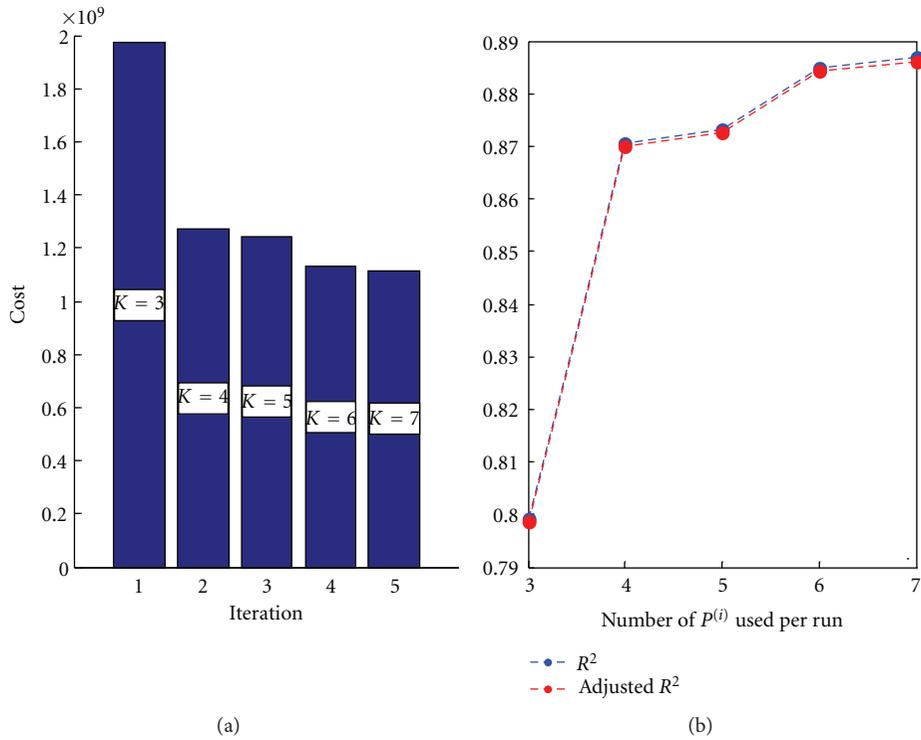


FIGURE 8: Goodness of fit for each iteration. (a) Cost function value (SSE). (b) Coefficient of determination  $R^2$  and adjusted  $R^2$ . Apparently, the data require the flexibility of a model with at least  $K = 4$ . Adding further points does not improve the model as much anymore and therefore  $K^* = 4$  is a reasonable choice.

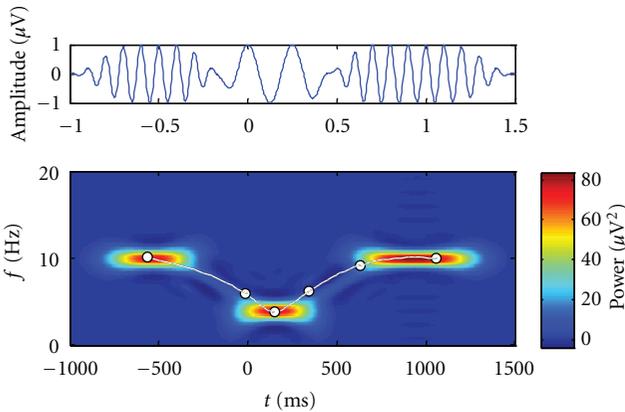


FIGURE 9: The simulated signal in the time domain and in the time-frequency domain showing a complex alpha/theta/alpha oscillation pattern. The white line with dots represents a fitted model which will serve as the reference model for the experiments.

to smoothing. Therefore, the pattern shown in Figure 9 will be distorted. This experiment serves to assess how strongly our algorithm is affected by pattern variability, respectively, to investigate its robustness. Small pattern distortions should ideally only slightly alter the optimal model, reflecting its robustness and avoidance of overfitting. We further want to find out to what degree our model is able to find the simulated pattern at all.

We note here that adding noise increases the TFRs' maximum amplitudes exponentially which strongly affects the

comparability of different models. Without normalization, one would observe an exponentially decreasing distance for increasing signal to noise ratio. But this would merely reflect the decreasing data amplitudes and contain no information about the quality of fit. However, normalizing maximum data values are not an appropriate option either, because, for negative SNRs, this would keep the noise constant while exponentially shrinking the pattern's pixel intensities. Even if the optimal model was perfectly recovered from the noisy simulation, high distances would arise. Only if signal power is excluded from model distance estimation, the returned values are useful representatives of how well the pattern was found. The *snaGe*'s robustness to noise with respect to the pattern's power is therefore not regarded here. This is done by setting the third dimension of the path of peaks  $P(u)$  to zero during Fréchet distance computation.

Figure 10 visualizes the results. Both curves of mean distance consistently decrease with improving data quality. Convergence to the optimal model seems to require high signal to noise ratios. At 7.5 dB, the distance measures' variances fall off, reflecting the point of reliable pattern extraction. Apparently, the noise and interferences introduced in this experiment considerably impair the fitting process. In Figure 11, this issue is exemplarily investigated. At the positive SNR of 5 dB, where distance variances across the noise realizations are still high, the fit which exhibits the largest distance is plotted. The pattern was in fact found, but only in a different way than was expected. This leads to high Fréchet distances. Nevertheless, this example shows that the impact different kinds of noise may have on the fitting process.

To get a better feel for the average ability to fit the pattern under the influence of noise, see Figure 12. At each noise level, the ten fitted models are averaged by computing the mean parameter vector. Shown is a sequence of mean models which progressively look more similar to the true pattern. In fact, the average fitting capability concerning both the positioning of the peak points in the time-frequency domain and the estimation of surface values is better than expected after having studied Figure 10. Apparently, although the mean distances are still decreasing at negative signal to noise ratios, they are already small enough for successful pattern extraction on average. An example is the subplot corresponding to SNR = 0 dB in Figure 12, which already clearly resembles the simulated pattern. This experiment shows that interferences between the desired signal and additive noise affect the fitting process quite strongly in the worst case. Positive signal to noise ratios of at least 7.5 dB are found to be necessary for reliable pattern extraction in this investigation. However, successful data modeling is also possible at lower SNRs, as is seen in the average case.

#### 4. Discussion

The *snaGe* model is especially suited for time-frequency representations of electrophysiological signals because of their (expected) nonnegativity, smoothness and their patterns following a path of peaks. However, our robust model is able to cope with data which do not exactly meet these requirements.

In order to retain robustness, we imposed several restrictions on our model, like neglecting offdiagonal spread parameters and holding the spread matrix constant over the curve of peaks. An interesting question remains how the model's flexibility and robustness would be affected if these constraints were dropped. Little effort would be necessary to include the stated extensions.

As is typical for nonlinear optimization problems, the choice of initial parameters is crucial to obtain satisfying results. Therefore, a-priori knowledge about the optimal model can be incorporated by starting the optimization with a model which was previously fitted to similar data. Moreover, an algorithm based on an optimal path is developed to estimate initial model parameters directly from the data. Its robustness stems from the guarantee to find the globally optimal path. However, this method is limited to *positive* peak polarity by trying to *maximize* the path's average amplitude. Additionally, the technique will not be able to find initial models which exhibit multiple contemporary components. In such a case, the nonlinear optimization algorithm, which was found to work well, must compensate. Further strategies for the estimation of initial parameters would be desirable. In particular, the extraction of the curve interpolation points  $P^{(i)}$  from the optimal path possesses potential for improvement.

An open question is how we should deal with the spatial correlation of both the dependent variable and the residuals in a statistical inferential context. Further work is necessary to facilitate statistical testing, for instance, to assess the null

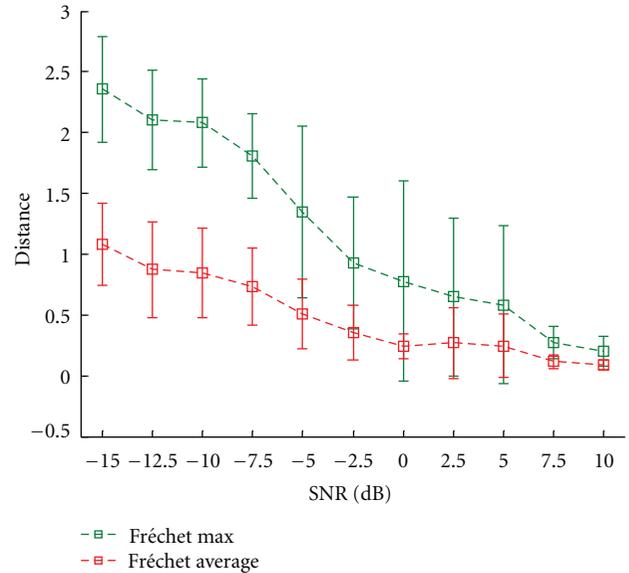


FIGURE 10: Mean distance to the optimal model versus level of correlated noise in the time domain. Error bars represent one standard deviation.

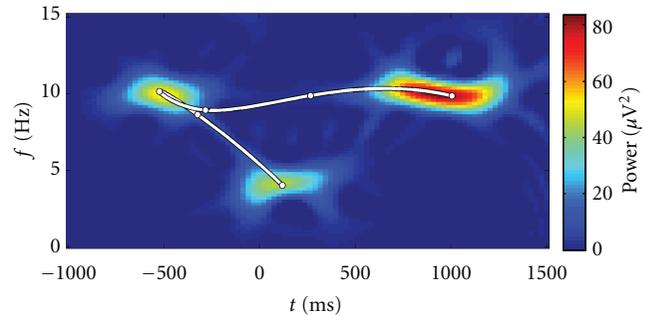


FIGURE 11: Fitted model (white curve) at SNR = 5 dB, which has the largest distance  $D_{F\_sum}$  to the target pattern across all ten noise instances. The noisy TFR is shown in the background. The signal was successfully extracted, yet a high-distance results due to a different connection of the three peaks compared to the pattern.

hypothesis that an expected pattern is not contained in the data.

Concerning the presented measures of model distance, the Fréchet distances were found to be very useful to assess model similarity in our experiments involving simulated noise. Their distinct advantage is their independence of model order and the disregard of the less interpretable parameters. On the other hand, spurious high distances could be observed when in fact the pattern was found. This can be attributed to the fact that the simulated data exhibit three independent peaks, which is a violation of the *snaGe*'s assumption of a connected path of peaks. Therefore, a combination of Fréchet values and a pixel-wise distance function based on the models' generated data seems advantageous.

When applied to noisy time signals, the *snaGe* model adapts too well to the corresponding smooth time-frequency representations. Because the optimized cost function does

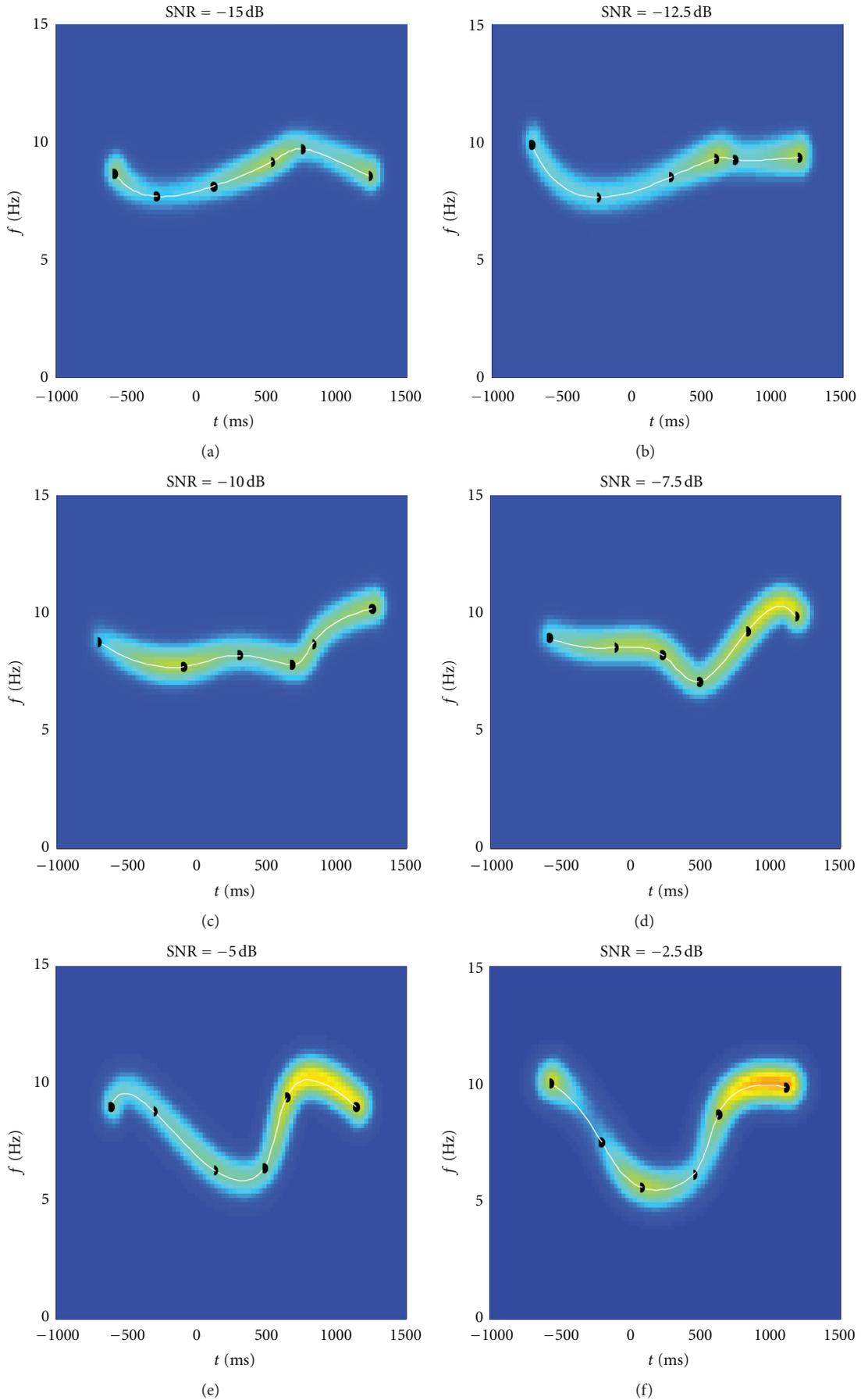


FIGURE 12: Continued.

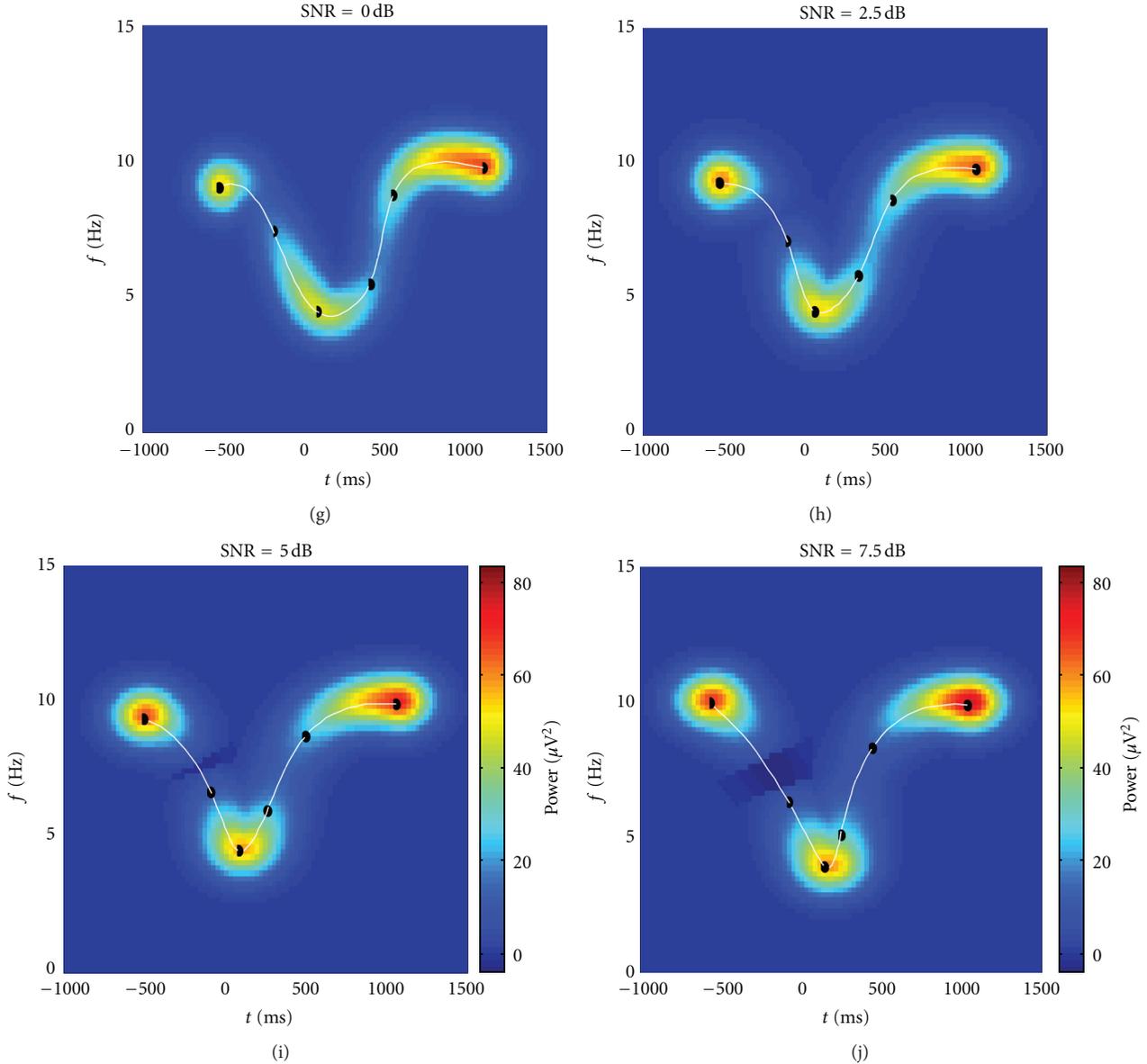


FIGURE 12: Mean *snaGe* models per simulated SNR. Models were averaged across the ten noise realizations by direct parameter vector averaging. The curve of peaks (white line) is shown as well as the models' predicted data  $y^{(snaGe)}(t, f)$ . Compare with Figure 9.

not take into account information about the expected pattern, the model simply tries to capture the TFR data as accurately as possible. Data preprocessing and TFR interference suppression are therefore extremely important. Adding penalty terms to the cost function and/or providing explicit initial parameters are ways to point the optimizer in the right direction. However, even without specifying a-priori knowledge, the model was able to find the simulated pattern for low signal to noise ratios in the average case.

## 5. Conclusion

The analysis of time-frequency representations of electrophysiological signals calls for flexible methods accounting for inter- and intraindividual data variability. We present

the flexible, robust, and interpretable model *snaGe*, which extends the established Gaussian model. Its ability to extract 3D features from time-frequency representations of electrophysiological data is demonstrated. However, the model applies to general multivariate data which exhibit similar behavior.

In this work, several techniques to improve the model fitting performance are described. We show how to estimate start parameters directly from the data. An iterative scheme to refine optimized models is proposed so that high-order models can be robustly fitted.

Experiments with real as well as simulated data demonstrate the *snaGe* model's robustness and flexibility. Under the influence of severe noise, the developed technique is best suited for patterns which are too complex to be appropriately

captured by a Gaussian model, but still simple enough to facilitate robust fits.

To summarize, due to its robustness and flexibility the *snaGe* model possesses the potential to become a beneficial tool for practical EEG/MEG analysis, including functional brain connectivity analysis, outlier detection, time-frequency denoising, and feature extraction.

## References

- [1] Y. Zhan, D. Halliday, P. Jiang, X. Liu, and J. Feng, "Detecting time-dependent coherence between non-stationary electrophysiological signals—a combined statistical and time-frequency approach," *Journal of Neuroscience Methods*, vol. 156, no. 1-2, pp. 322–332, 2006.
- [2] L. B. White and B. Boashash, "Cross spectral analysis of non-stationary processes," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 830–835, 1990.
- [3] J. P. Lachaux, A. Lutz, D. Rudrauf et al., "Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence," *Neurophysiologie Clinique*, vol. 32, no. 3, pp. 157–174, 2002.
- [4] T. H. Sander, T. R. Knösche, A. Schlögl et al., "Recent advances in modeling and analysis of bioelectric and biomagnetic sources," *Biomedical Engineering*, vol. 55, no. 2, pp. 65–76, 2010.
- [5] P. Flandrin and W. Martin, "Pseudo-Wigner estimators for the analysis of nonstationary processes," in *Proceedings of the 2nd IEEE ASSP Spectrum Estimation Workshop*, pp. 181–185, Tampa, Fla, USA, 1983.
- [6] L. Cohen, "Time-frequency distributions—a review," *Proceedings of the IEEE*, vol. 77, pp. 941–981, 1989.
- [7] C. M. Beale, J. J. Lennon, J. M. Yearsley, M. J. Brewer, and D. A. Elston, "Regression analysis of spatial data," *Ecology Letters*, vol. 13, no. 2, pp. 246–264, 2010.
- [8] M. S. Nixon and A. S. Aguado, *Feature Extraction & Image Processing*, Academic Press, 2008.
- [9] D. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometric Recognition*, 2008.
- [10] The MathWorks Inc., MATLAB Optimization Toolbox, 2010.
- [11] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, 1999.
- [12] S. P. Bradley, A. C. Hax, and T. L. Magnanti, "Dynamic programming," in *Applied Mathematical Programming*, Addison-Wesley, 1977.
- [13] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, 2007.
- [14] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk, "Fréchet distance for curves, revisited," in *Algorithms-ESA*, pp. 52–63, 2006.
- [15] T. Eiter and H. Mannila, "Computing discrete Fréchet distance," Tech. Rep. CD-TR 94/64, TU Vienna, Vienna, Austria, 1994.
- [16] H. Theil, *Economic Forecasts and Policy*, North-Holland, 1961.
- [17] A. N. Spiess and N. Neumeyer, "An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach," *BMC Pharmacology*, vol. 10, article 6, 2010.
- [18] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the 2nd International Symposium on Information Theory*, vol. 1, pp. 267–281, 1973.
- [19] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [20] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3, pp. 591–611, 1965.

## Research Article

# Fast and Analytical EAP Approximation from a 4th-Order Tensor

**Aurobrata Ghosh and Rachid Deriche**

*ATHENA Research Team, INRIA Sophia Antipolis Méditerranée, 2004 Route des Lucioles, BP 93,  
06902 Sophia Antipolis Cedex, France*

Correspondence should be addressed to Aurobrata Ghosh, aurobrata.ghosh@inria.fr

Received 20 July 2012; Accepted 2 December 2012

Academic Editor: Carl-Fredrik Westin

Copyright © 2012 A. Ghosh and R. Deriche. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Generalized diffusion tensor imaging (GDTI) was developed to model complex apparent diffusivity coefficient (ADC) using higher-order tensors (HOTs) and to overcome the inherent single-peak shortcoming of DTI. However, the geometry of a complex ADC profile does not correspond to the underlying structure of fibers. This tissue geometry can be inferred from the shape of the ensemble average propagator (EAP). Though interesting methods for estimating a positive ADC using 4th-order diffusion tensors were developed, GDTI in general was overtaken by other approaches, for example, the orientation distribution function (ODF), since it is considerably difficult to recuperate the EAP from a HOT model of the ADC in GDTI. In this paper, we present a novel closed-form approximation of the EAP using Hermite polynomials from a modified HOT model of the original GDTI-ADC. Since the solution is analytical, it is fast, differentiable, and the approximation converges well to the true EAP. This method also makes the effort of computing a positive ADC worthwhile, since now both the ADC and the EAP can be used and have closed forms. We demonstrate our approach with 4th-order tensors on synthetic data and in vivo human data.

## 1. Introduction

Generalized diffusion tensor imaging (GDTI) [1–3], was proposed to model the apparent diffusion coefficient (ADC) recovered by diffusion MRI (dMRI) when imaging the diffusion of water molecules in heterogeneous media like the cerebral white matter. Essentially, GDTI uses higher-order Cartesian tensors (HOTs) to model the spherical profile of the ADC. However, although the complex shape of the ADC reflects the complex geometry of the underlying tissue, it is well known that the geometry of the ADC does not correspond to the underlying fiber directions [4]. This can be understood from the  $q$ -space formalism, where it can be seen that the ADC and the diffusion signal are in the Fourier domain of the diffusion ensemble average propagator (EAP), which describes the probability of the diffusing particles. The geometry of the EAP is a direct indicator of the microstructure of the underlying tissue or fiber bundles.

But GDTI was proposed because it overcomes the limitation of diffusion tensor imaging (DTI) [5], which is inadequate for modelling the signal from regions with

multiple fiber configurations. In such regions, the HOT that is used in GDTI can model the signal and the ADC with greater accuracy than the 2nd-order diffusion tensor. Therefore, the GDTI model has been of considerable interest and has seen various developments. In particular, a number of contributions were made to estimate 4th-order HOTs in GDTI under the constraint of a positive diffusion profile since negative diffusion is nonphysical. A ternary quartic parameterization was used in [6–8], while a Riemannian approach was proposed in [9]. Other sophisticated methods were also proposed recently to estimate arbitrary even order HOTs in GDTI with the positivity constraint. In [10] the authors relied on a parameterization based on tensor decomposition into a sum of squares, and in [11], the authors used conic programming approaches to achieve this.

The GDTI model was also used to develop “biomarkers” or scalar indices such as the generalized anisotropy (GA) and the scaled entropy (SE) from HOTs modelling the ADC [3]. Additional scalar measures—in the form of invariants—were also proposed for 4th-order Cartesian tensors in [12, 13]. Overall, the GDTI model in particular and the Cartesian

tensors in general have evoked great interest and have been extensively explored in dMRI. Various tensor-based models other than GDTI have been advanced and many methods (including GDTI) using tensors have been proposed to leverage the Cartesian and the algebraic structure of HOTs. A comprehensive review can be found in [14]. These indicate the importance and usefulness of HOTs in dMRI.

However, in spite of the interest in HOTs, to describe complex shaped ADCs, the tissue microstructure can only be inferred from the shape of the EAP. But computing the EAP from the HOT model of the ADC in GDTI is not an easy task. In [2], the authors proposed a numerical fast Fourier transform scheme—to emulate diffusion spectrum imaging (DSI) [15] from GDTI—to estimate the EAP and to recover the underlying fiber directions. However, this method is computationally expensive, and although the numerical Fourier transform can compute the values of the EAP at desired points, it cannot compute a continuous and differentiable function which has great advantages. That is perhaps the reason why the GDTI approach has been overshadowed by other methods that estimate the EAP or its characteristics directly from the signal, such as orientation distribution function (ODF) from Q-ball imaging (QBI), persistent angular structure (PAS-MRI), diffusion orientation transform (DOT), and spherical deconvolution (SD) [16–20]. It is interesting to note that DOT also uses the GDTI model to analytically compute the EAP on fixed shells from the ADC. However, in DOT, the spherical harmonic (SH) basis is used to model the ADC instead of Cartesian tensors.

In this paper, we propose a modification to the original GDTI model under the  $q$ -space formalism, which allows us to compute a closed-form approximation of the EAP using Hermite polynomials. In this modified model, we still estimate HOTs from the signal, but these HOTs are used to describe the ADC over the entire  $q$ -space instead of just its spherical profile. We show that the approximated EAP from the modified GDTI model converges well to the true EAP, and that it allows us to recover the fiber directions of the microstructure correctly. Furthermore, in our modified GDTI model, when we use 4th-order tensors, we are still able to apply the constraint of a positive diffusion profile while estimating the HOT from the signal before computing the EAP approximation from the tensor. Finally since the solution is analytical, it is fast and can be implemented efficiently.

We first test this approach on a synthetically generated dataset that simulates crossing fibers. We compare the computation time of this method with a numerical discrete Fourier transform scheme to recover the EAP from the original GDTI model. We show that with our modified approach, we are able to recover the underlying fiber layout correctly and also gain considerably in computation time. This is of great relevance in visualization and in post-processing such as tractography. We also conduct experiments on *in vivo* human cerebral data to illustrate the applicability of this approach on real data. We find that the peaks of the approximate EAP clearly reveal major fiber bundles and also discern crossings in the white matter.

The rest of the paper is structured as follows. In Section 2, we present the main theory and mathematical formulation of the modified GDTI model and algorithm to estimate the EAP from the GDTI-HOT. We illustrate on 4th-order tensors. In Section 3, we describe the synthetic and *in vivo* data, describe the experiments, and present the results. In Section 4, we discuss some properties of the approximation and how the approximate EAP behaves with respect to the true EAP. We finally conclude in Section 5.

## 2. Materials and Methods

*2.1. Modified GDTI.* We recall that the signal  $E(\mathbf{q}) = S(\mathbf{q})/S_0$ , in GDTI [1], is modelled using a  $k$ th-order tensor  $\mathcal{D}^{(k)}$  (with  $k$  being even), which describes the spherical or angular profile of the ADC:

$$\begin{aligned} E(\mathbf{q})_k &= \exp\left(-4\pi^2 q^2 t \sum_{j_1=1}^3 \sum_{j_2=1}^3 \cdots \sum_{j_k=1}^3 D_{j_1 j_2 \cdots j_k}^{(k)} g_{j_1} g_{j_2} \cdots g_{j_k}\right) \\ &= \exp\left(-4\pi^2 q^2 t \sum_{m+n+p=k} D_{mnp}^{(k)} g_1^m g_2^n g_3^p\right), \end{aligned} \quad (1)$$

where  $\mathbf{q} = \gamma \delta \mathbf{G} / 2\pi$  when  $\mathbf{G}$  is the diffusion encoding gradient vector,  $t = (\Delta - \delta/3)$ , and  $g_j$  are the components of the unit gradient vector  $\mathbf{g} = \mathbf{G}/|\mathbf{G}|$ . The second equality is a reinterpretation of the first by a rearrangement of the indices that highlights the polynomial interpretation of tensors [6].  $k = 4$  gives the 4th-order diffusion tensor model. We also recall that in the  $q$ -space formalism, the diffusion signal and the EAP are related by the Fourier transform [21]:

$$P(\mathbf{r}) = \int E(\mathbf{q}) \exp(-2\pi i \mathbf{q}^T \mathbf{r}) d\mathbf{q}. \quad (2)$$

The  $q$ -space formalism entails the ‘‘pulsed’’ gradient condition that  $\delta \ll \Delta$ , implying that  $t \approx \Delta$ .

For  $k = 2$ ,  $E(\mathbf{q})_2$  is the DTI model, whose Fourier transform  $P(\mathbf{r})_2$  is well known to also be a Gaussian, which corresponds to the free diffusion propagator. We denote by  $P(\mathbf{r})_i$  the EAP computed from the Fourier transform of  $E(\mathbf{q})_i$ . However, for general  $k > 2$ , closed forms for the Cartesian Fourier transform of  $E(\mathbf{q})_k$  are hard to compute, since in Cartesian coordinates,  $E(\mathbf{q})_k$  is not separable in  $q_1, q_2$ , and  $q_3$ , the components of  $\mathbf{q}$ . In [2], where a method for recovering the EAP from GDTI is proposed,  $P(\mathbf{r})_k$  is computed numerically by evaluating  $E(\mathbf{q})_k$  more or less densely in  $q$ -space and by computing its fast Fourier transform.

In this section, we propose to modify the original GDTI model by making (1) separable in Cartesian coordinates. This is done by realizing that GDTI in fact uses two orders

$k_1$  and  $k_2$  for the radial and the angular components, respectively,

$$E(\mathbf{q})_{k_1, k_2} = \exp\left(-4\pi^2 q^{k_1} \alpha t \sum_{m+n+p=k_2} D_{mnp}^{(k_2)} g_1^m g_2^n g_3^p\right) \quad (3)$$

$$= \exp\left(-4\pi^2 q^{k_1-k_2} \alpha t \sum_{m+n+p=k_2} D_{mnp}^{(k_2)} q_1^m q_2^n q_3^p\right), \quad (4)$$

where in GDTI  $k_1 = 2$  and  $k_2 = k$ . In this formulation,  $\alpha$  is a constant with units  $m^{2-k_1}$  that makes the exponent unit free when  $k_1 \neq 2$  and  $q = |\mathbf{q}|$ . The first equality is written in the components of the unit gradient vector  $\mathbf{g}$ , while the second equality is in the components of the reciprocal space vector  $\mathbf{q}$ . To pass from the first to the second equality, the components of  $\mathbf{g}$  have to be multiplied by  $q$ , the norm of  $\mathbf{q}$ , raised to the appropriate power,  $k_2$ . To leave the equality unchanged, it is therefore necessary to also multiply by  $q^{-k_2}$ , which results in the second equality. This reformulation of GDTI allows  $E(\mathbf{q})$  to become separable in  $q_1, q_2$ , and  $q_3$  when  $k_1 = k_2 = k$ .

Although the original formulation of GDTI uses a Cartesian HOT, it was in essence written in spherical coordinates since the HOT  $\mathcal{D}^{(k)}$  was evaluated only along the unit gradient vector  $\mathbf{g}$ . Alternately, the spherical coordinates also become evident from the two separate orders  $k_1$  and  $k_2$  for the radial and the angular parts. By equating the two orders  $k_1 = k_2$ , our modification converts the signal formulation to Cartesian coordinates and recouples the radial and the angular parts. The HOT is now evaluated over the entire  $q$ -space. This reformulation allows us to compute an analytical Fourier transform of  $E(\mathbf{q})_{k,k}$  in Cartesian coordinates.

Interestingly, in spite of this reformulation, the signal in (4) still retains a monoexponential form parameterized by the diffusion HOT  $\mathcal{D}^{(k_2)}$ , like in the original formulation. In (1), the negative logarithm of the signal is  $b \cdot \text{ADC}$ , where  $b = 4\pi^2 q^2 t$ . In the modified model, if we denote  $b' = 4\pi^2 \alpha t$ , then

$$\ln(E(\mathbf{q})_{k,k}) = -b' \cdot \widehat{\text{ADC}} = -b' \sum_{m+n+p=k} D_{mnp}^{(k)} q_1^m q_2^n q_3^p. \quad (5)$$

This makes it evident that we can again estimate the HOT  $\mathcal{D}^{(k)}$  from the diffusion signal in such a way that its diffusion profile is positive, that is,  $\widehat{\text{ADC}} > 0$ . Here, the diffusion profile  $\widehat{\text{ADC}}$  is no longer a function on the sphere but rather a function on the entire  $q$ -space. However, from this equation, we also see that when  $k = 4$ , the methods that were developed to estimate a 4th-order tensor with a positive diffusion profile for the GDTI model in (1), can all be directly applied to the modified model in (4). Therefore, using the modified model in (4), it is possible to estimate a 4th-order HOT from the signal which satisfies a positive diffusion profile, before computing the EAP from this HOT. In this paper, we use the ternary quartic parameterization proposed in [8].

**2.2. Fast and Analytical EAP Approximation.** Our solution for the EAP from the modified GDTI model pivots around the following property of the Fourier transform:

$$\mathcal{F}\{x^n f(x)\} = \left(\frac{i}{2\pi}\right)^n \frac{d^n}{dt^n} \mathcal{F}\{f(x)\}(t), \quad (6)$$

where  $\mathcal{F}$  stands for the Fourier transform. If we employ  $g(x) = e^{-2\pi^2 x^2}$  for  $f(x)$ , then its Fourier transform is  $G(t) = \mathcal{F}\{g(x)\}(t) = (1/\sqrt{2\pi})e^{-t^2/2}$ . However, the derivatives of the Gaussian function  $G(t)$  generate the Hermite polynomials  $(-1)^n (d^n/dt^n) e^{-t^2/2} = He_n(t) e^{-t^2/2}$ . Therefore,

$$\mathcal{F}\{x^n e^{-2\pi^2 x^2}\}(t) = \left(\frac{-i}{2\pi}\right)^n He_n(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. \quad (7)$$

The generalization to 3D is simple since the Gaussian function is separable in the variables.

To take advantage of this property of the Fourier transform, and those of the Gaussian function, for computing a closed-form approximation of the EAP from  $E(\mathbf{q})_{k,k}$ , that is, its Fourier transform  $P(\mathbf{r})_{k,k}$ , we propose to expand  $E(\mathbf{q})_{k,k}$  as a multivariate polynomial multiplied by a 3D Gaussian function:

$$E(\mathbf{q})_{k,k} \approx \left(\sum C_{l,s,u} q_1^l q_2^s q_3^u\right) \exp(-2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)), \quad (8)$$

where  $\beta$  is a constant with units  $m^2$  to render the exponent unit free, and the new coefficients  $C_{l,s,u}$  in the polynomial expansion contain the imaging parameter  $b'$  and the coefficients of the HOT  $\mathcal{D}^{(k)}$ . If this expansion was possible, then  $E(\mathbf{q})_{k,k}$  would become separable in  $q_1, q_2$ , and  $q_3$ .

Such an expansion can be achieved from a few manipulations and a Taylor expansion:

$$\begin{aligned} E(\mathbf{q})_{k,k} &= \exp\left(\left(-4\pi^2 \alpha t \sum_{m+n+p=k} D_{mnp}^{(k)} q_1^m q_2^n q_3^p\right)\right. \\ &\quad \left.+ 2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)\right) \\ &\quad \times \exp(-2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)) \\ &= h(\mathbf{q}) \exp(-2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)), \end{aligned} \quad (9)$$

where the summation in the first equality is over  $m, n$ , and  $p$  such that  $m + n + p = k$ , as denoted in (4), and  $h(\mathbf{q}) = \exp\left(\left(-4\pi^2 \alpha t \sum_{m+n+p=k} D_{mnp}^{(k)} q_1^m q_2^n q_3^p\right) + 2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)\right)$ . Since  $h(\mathbf{q})$  is an exponential function  $e^{X(\mathbf{q})}$ , we define  $h_n(\mathbf{q})$  as the  $n$ th-order Taylor expansion of  $h(\mathbf{q})$  in the variables  $q_1, q_2$ , and  $q_3$ . Therefore,  $h_n(\mathbf{q})$  is a trivariate polynomial of degree  $n-1$  plus an error term of degree  $n$ . Ignoring the error term,  $h_n(\mathbf{q})$  has the required form  $h_n(\mathbf{q}) = \sum_{l+s+u < n} C_{l,s,u} q_1^l q_2^s q_3^u$ . Therefore, we can define the  $n$ th-order approximation of the signal:

$$\begin{aligned} E(\mathbf{q})_{k,k}^{(n)} &= h_n(\mathbf{q}) \exp(-2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)) \\ &= \left(\sum_{l+s+u < n} C_{l,s,u} q_1^l q_2^s q_3^u\right) \exp(-2\pi^2 \beta (q_1^2 + q_2^2 + q_3^2)). \end{aligned} \quad (10)$$

Since  $h_n(\mathbf{q})$  is the Taylor's expansion of an exponential function  $h(\mathbf{q})$ ,  $h_n(\mathbf{q})$  converges to  $h(\mathbf{q})$  uniformly over all  $\mathbf{R}^3$

as  $n$  is made large. Therefore,  $E(\mathbf{q})_{k,k}^{(n)}$  converges to  $E(\mathbf{q})_{k,k}$  uniformly over  $\mathbf{R}^3$  as  $n$  is made large.

As  $E(\mathbf{q})_{k,k}^{(n)}$  is separable in  $q_1, q_2$ , and  $q_3$ , it is possible to compute a closed form for its Cartesian Fourier transform, which is also separable. Using the property in (7):

$$P(\mathbf{r})_{k,k}^{(n)} = \frac{1}{(2\pi\beta)^{3/2}} \exp\left(\frac{-1}{2\beta}(r_1^2 + r_2^2 + r_3^2)\right) \times \left( \sum_{l+s+u < n} \left(\frac{-i}{2\pi}\right)^{l+s+u} C_{l,s,u} \text{He}_l(r_1) \text{He}_s(r_2) \text{He}_u(r_3) \right). \quad (11)$$

For large  $n$ , the approximation  $P(\mathbf{r})_{k,k}^{(n)}$  converges to the true EAP  $P(\mathbf{r})_{k,k}$ . In practice, we use  $n = 5, 7, 9$ .

We thus find a closed-form approximation of the EAP from the modified GDTI model of the ADC using HOTs. The solution is a polynomial multiplied by a Gaussian. Therefore, the polynomial can be interpreted as the correction to the free diffusion Gaussian EAP due to the complex heterogeneous medium.

An alternate interpretation to this method can be found from (10) and (11), which avoids modifying the GDTI model. Equations (10) and (11) resemble closely the formulation of the signal in the expansion of the cumulant generating function, and the approximation of the EAP using the Gram-Charlier series, as proposed in [22]. While in [22], in the cumulant expansion, the signal is expanded in the standard polynomial basis with the cumulants as the coefficients, in (10) the signal is in fact expanded in a subset of the standard polynomial basis. Since the Fourier transform of a monomial multiplied by a Gaussian is a Hermite polynomial multiplied by a Gaussian, in [22] too, the EAP is approximated in the Hermite polynomial basis—again with the cumulants as the coefficients using the Gram-Charlier series. Likewise in (11), the EAP is approximated in the Hermite polynomial basis.

The difference between this method and [22] lies in the fact that while [22] uses the entire polynomial (Hermite polynomial) basis to expand the signal (EAP), this method uses only a subset of these bases. Therefore, the coefficients  $C_{l,s,u}$  are no longer the cumulants. Or, in other words, if the entire polynomial basis had been used here, then  $C_{l,s,u}$  would have become the cumulants. Also, the coefficients  $C_{l,s,u}$  are not estimated directly from the signal, though they can be if (10) were used, but  $C_{l,s,u}$  are instead computed from the coefficients of the tensor  $\mathcal{D}^{(k)}$  and the Taylor expansion in (10). Therefore, changing the order  $n$  of the approximation has an effect on the approximated EAP, since it adds or subtracts terms in (11). However, as shown in the experiments, this does not affect the direction of the peaks of the approximate EAP.

We program an efficient implementation of the proposed method through symbolic computation. Using Maple and assuming (4), we expand  $E(\mathbf{q})_{k,k}$  into a Taylor series in the variables  $q_1, q_2$ , and  $q_3$  up to predefined orders  $n = 5, 7, 9$ . This expansion automatically computes for us the new coefficients  $C_{l,s,u}$  from the coefficients of the HOT  $\mathcal{D}^{(k)}$  (10).

The EAP approximation  $P(\mathbf{r})_{k,k}^{(n)}$ , is then generated by again computing the Fourier transform of  $E(\mathbf{q})_{k,k}^{(n)}$  symbolically. The expansion of the EAP is then converted to C-code using Maple, which is compiled. This routine therefore takes the imaging parameters as input, namely,  $t$  and the coefficients of  $\mathcal{D}^{(k)}$  that are estimated from the diffusion signal.  $\alpha, \beta$  are taken to be equal to 1.

### 3. Experiments and Results

Although we developed the theory for arbitrary  $k = k_1 = k_2$ , for the following experiments we consider  $k = 4$ , that is,  $E(\mathbf{q})_{4,4}$  and  $\text{Pr}_{4,4}^{(n)}$ . This is because, as we have seen, for  $E(\mathbf{q})_{4,4}$ , we can employ the estimation techniques that guarantee that the 4th-order HOT has a positive diffusion profile,  $\widehat{\text{ADC}} > 0$ . In all the following the 4th-order HOT  $\mathcal{D}^{(4)}$  is estimated from (3), with  $k_1 = k_2 = 4$  using the method described in [8]. The estimation in [8] is described for (1), which depends on the  $b$ -value, that is,  $b = 4\pi^2 q^2 t$ . We adapt this to (3) by replacing the  $b$ -value by the imaging parameters  $4\pi^2 a t$  and the ADC by the  $\widehat{\text{ADC}}$ . We test the approach first on synthetic data and then on *in vivo* human cerebral data.

**3.1. Synthetic Dataset.** To conduct controlled experiments with known ground truths, we use a multitensor approach to generate synthetic DWIs [23]. The EAP corresponding to a single fiber is taken to be an anisotropic free diffusion Gaussian distribution, parameterized by a covariance tensor  $\mathcal{D} = \text{diag}(1390, 355, 355) \times 10^{-6} \text{ mm}^2/\text{s}$  in its canonical coordinates.  $\mathcal{D}$  is rotated using rotation matrices to orient the fiber in space. We generate the signal DWIs for the single fiber by considering the  $q$ -space formalism and taking the Fourier transform of the Gaussian EAP, which results in the anisotropic Stejskal-Tanner signal equation. Multiple crossing fibers are simulated by considering an EAP, that is, the weighted sum of free diffusion Gaussians, where each Gaussian represents a fiber oriented in space. The signal DWIs for a multifiber or crossing fiber is derived easily in the same fashion as  $S(\mathbf{g}_i) = \sum_{k=1}^N w_k e^{-b\mathbf{g}_i^T \mathcal{D}_k \mathbf{g}_i}$ , such that  $\sum_k w_k = 1$ ,  $\mathcal{D}_k = \mathbf{R}_k^T \mathcal{D} \mathbf{R}_k$  with  $\mathbf{R}$  a rotation matrix,  $S(\mathbf{g}_i)$  represents the DWI along the  $i$ th gradient direction, and  $N$  are the number of fibers crossing in the voxel. We use a  $b$ -value of  $3000 \text{ s/mm}^2$  to generate the signal and corrupt it with a Rician noise with signal to noise ratio (SNR) of 30. The gradient directions are considered isotropically spread out on the sphere along 81 encoding directions. Since the dataset is generated from a fixed  $b$ -value, we consider the imaging parameter  $t = 50 \text{ ms}$ , which allows us to compute  $q$ .

**3.2. In Vivo Human Cerebral Dataset.** The *in vivo* human cerebral dataset, described in [24], was acquired with a whole-body 3T Siemens Trio scanner, with an 8-channel array head coil and maximum gradient strength of  $40 \text{ mT/m}$ . The DWIs were acquired using spin-echo echo planar imaging (EPI) (time repetition [TR] =  $12 \text{ s}$ , echo time [TE] =  $100 \text{ ms}$ ,  $128 \times 128$  image matrix, FOV =  $220 \times 220 \text{ mm}^2$ ,

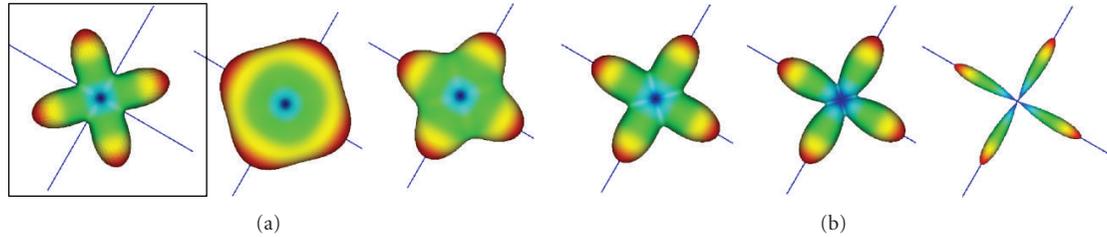


FIGURE 1: Spherical profiles of (a) the  $\widehat{\text{ADC}}$  estimated from the modified GDTI with a 4th-order tensor and (b) EAP approximation,  $P(\mathbf{r})_{4,4}^{(7)}$ , with increasing  $|\mathbf{r}|$  from  $12\ \mu\text{m}$  to  $20\ \mu\text{m}$ .

72 slices with 1.7 mm thickness (no gap) covering the whole brain). The diffusion weighting was isotropically distributed along 60 encoding directions, with a  $b$ -value of  $1000\ \text{s}/\text{mm}^2$ . Seven images without any diffusion weightings were placed at the beginning of the sequence and after each block of ten DWIs as anatomical reference for offline motion correction. Random variations in the data were reduced by averaging 3 acquisitions, resulting in an acquisition time of about 45 minutes. The SNR in the white matter of the  $S_0$  image was estimated to be approximately 37. The motion correction for the DWIs was combined with a global registration to T1 anatomy images. The gradient direction for each volume was corrected using rotation parameters. The registered images were interpolated to the new reference frame with an isotropic voxel resolution of 1.72 mm.

**3.3. Synthetic Dataset Experiment.** A first example of the synthetic data and the results of our method are shown in Figure 1. In this proof of concept experiment, we consider two fibers crossing perpendicularly in a voxel with equal weights. On the left, in Figure 1(a), is shown the positive  $\widehat{\text{ADC}}$  modelled by a 4th-order tensor which was estimated using [8]. In Figure 1(b) are shown the profiles of the analytically approximated EAP ( $n = 7$ ) for increasing norms of the vector  $\mathbf{r}$  from  $12\ \mu\text{m}$  to  $20\ \mu\text{m}$ . As expected, we observe the desired “sharpening” effect from the different profiles of the EAP as the probability of water molecules diffusing declines sharply along nonfiber directions as the radius of the probability distribution is increased. This provides a strong motivation for estimating positive higher-order diffusion tensors from the (modified) GDTI model, since now from the analytically approximated EAP it is also possible to infer underlying fiber directions.

In the main synthetic data experiment, we consider two fiber bundles crossing or overlapping in a way that makes them converge and diverge. This changes their crossing angles gradually in the region where they intersect. The voxels outside the fiber bundles are generated using an isotropic diffusion profile. We set three goals for this experiment. First, we test if our analytically approximate EAP can recover the three types of voxel models from the noisy DWIs, namely, isotropic, single fiber, and crossing fiber voxels. Second, we compare the computation time of our proposed method to the numerical Fourier transform approach. Finally, third, we also conduct tests on the effects

of the estimation order on the EAP; these are discussed in Section 4.

The layout of the synthetic dataset fibers and the result of the estimated  $\widehat{\text{ADC}}$  from 4th-order HOTs and the analytical EAP approximations of order 7,  $\text{Pr}_{4,4}^{(7)}$  are presented in Figure 2. It shows the two simulated fiber bundles and the three types of voxel models that constitute the synthetic dataset experiment. The effect of the EAP transformation of the  $\widehat{\text{ADC}}$  are highlighted in the zooms. Although the  $\widehat{\text{ADC}}$  profiles clearly indicate the regions with complex microstructures, that is, fiber crossings, the geometry of the  $\widehat{\text{ADC}}$  is not aligned with the fiber bundle directions. The peaks of the EAP on the other hand correctly indicate the underlying fiber directions.

To evaluate the validity of the EAP approximation, we have proposed the angular profiles of the EAP, for fixed  $|\mathbf{r}| = 20\ \mu\text{m}$  that are presented in Figure 3. In the two zooms, we take a closer look at some of the voxels in the crossing regions. In the top zoom, we see that the peaks of the angular profile of the EAP correctly detect the changing angle between the converging or diverging fiber bundles. In the bottom zoom, we see the three different types of voxels recovered by the approximate EAP, namely, the isotropic, the single fiber, and the two fibers crossings. Although the isotropic voxels also have some peaks, the peaks of the EAPs representing crossings are much more sharp, and it is easy to distinguish these two types of voxel geometries. The peaks in the isotropic voxels are caused by the signal noise and augmented order (4th) of the HOT. Nonetheless, as opposed to many other higher-order models such as PASMRI or SD, the voxels not containing fiber bundles are clearly identifiable as isotropic.

Speed is of great relevance in visualization and in processing after local estimation, such as in tractography. The closed-form of  $\text{Pr}_{4,4}^{(n)}$  makes it computationally efficient, especially since the expression for a fixed  $n$  can be hard coded and compiled. In the synthetic data experiment, we compare this approach to a numerical Fourier transform of the GDTI model. For visualization and comparison, we consider the whole slice, which is partially seen in Figure 3, with  $30 \times 30$  voxels. For the implementation of the numerical Fourier transform, we evaluate the GDTI model (1) on a  $21 \times 21 \times 21$  Cartesian grid. We evaluate the numerically computed EAP on a coarse spherical mesh with 162 vertices. The results are presented in Table 1. The computation time on our computer

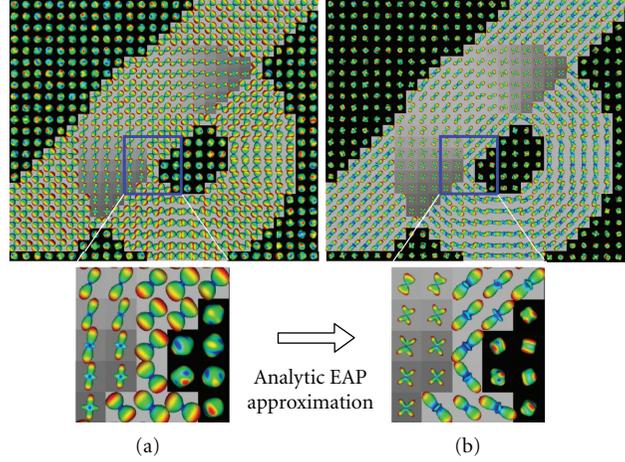


FIGURE 2: Synthetic dataset experiment. Two fiber bundles intersecting with the DWI signal corrupted by a Rician noise of  $\text{SNR} = 30$ . (a)  $\widehat{\text{ADC}}$  from 4th-order tensors. (b) Analytically approximated EAP,  $\text{Pr}_{4,4}^{(7)}$ .

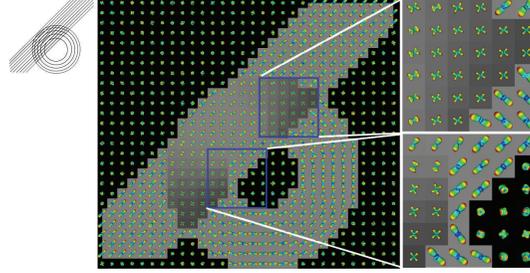


FIGURE 3: Synthetic dataset experiment. Left: fiber bundle layout. Centre:  $\text{Pr}_{4,4}^{(7)}$ . Right: zoom into the regions with crossings—top: the changing angle between the fiber bundles is detected and bottom: three types of voxels—*isotropic*, *single fiber*, and *crossing fibers*.

was 526 s. We then compute  $\text{Pr}_{4,4}^{(7)}$ , but this time on a finer spherical mesh with 2562 vertices. The computation time on the same computer was 73 sec. Despite the finer mesh,  $\text{Pr}_{4,4}^{(7)}$  is about seven times faster than the regular discrete Fourier transform. On the coarse mesh with 162 vertices, the computation time for  $\text{Pr}_{4,4}^{(7)}$  was about 10 s.

**3.4. In Vivo Dataset Experiment.** For the *in vivo* human cerebral dataset described above, we make certain assumptions about the imaging parameters since this dataset was acquired using a twice refocused Reese sequence [25], and not a standard pulsed-gradient spin-echo (PGSE) sequence. The gradient durations used in the Reese sequence were  $\delta_1 = 12.03$  ms,  $\delta_2 = 19.88$  ms,  $\delta_3 = 21.76$  ms, and  $\delta_4 = 10.15$  ms. As suggested in [26], Reese sequence parameters are sometimes adapted to the standard PGSE parameters with  $\delta = \delta_1 + \delta_2$ , and  $\Delta$  as the time between the start of  $\delta_1$  and the start of  $\delta_3$ . However, since the application times of  $\delta_i$  were unknown, we assume that  $q^2 = b$ , which implies  $4\pi^2 t = 1$ .

We choose a coronal slice from the *in vivo* dataset where three fiber bundles are known to cross. The resulting analytically approximated EAPs from 4th-order tensors in

the coronal slice are shown in Figure 4. In the plane horizontally and diagonally is the corpus callosum (CC), top to bottom is the corticospinal tract (CST), and going through the plane is the superior longitudinal fasciculus (SLF). The 4th-order HOTs were approximated from this dataset. In Figure 4, are shown the estimated order 7 approximations  $\text{Pr}_{4,4}^{(7)}$  of the EAP. The zooms highlight the crossings between the major fiber bundles. In the main zoom is the region where the three fibers, the CC, the CST, and the SLF, intersect each other. In the upper secondary zoom, the crossing between the CC and the cingulum is highlighted, which occurs due to partial voluming. In the lower secondary zoom, are seen the main voxels with three peaks, which correspond to the

## 4. Discussion

From the synthetic dataset and *in vivo* dataset experiments above, we were able to show that it is possible to recover the microstructure or fiber directions of the underlying tissue from our proposed analytical approximation of the EAP that was computed from 4th-order tensors. However, it is important to realize that the proposed approach only

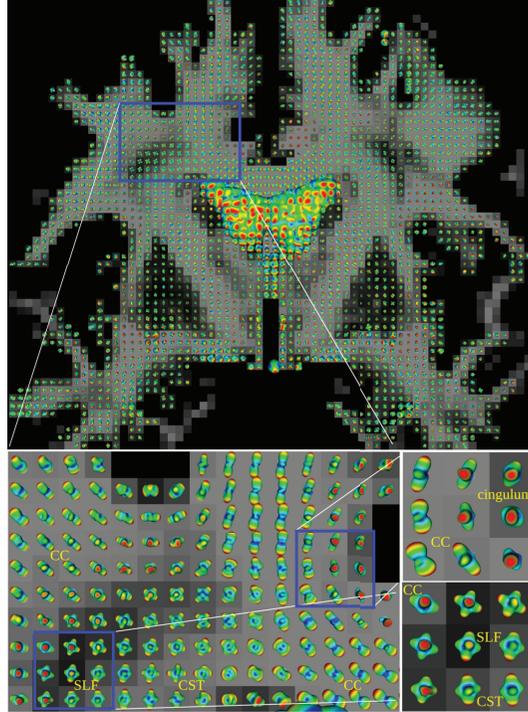


FIGURE 4: Real dataset experiment. A coronal slice with the  $\text{Pr}_{4,4}^{(7)}$ . The main zoom contains regions where three fiber bundles, namely, the CC, the CST, and the SLF intersect. The upper secondary zoom highlights the crossing between the CC and the cingulum due to partial voluming. The lower secondary zoom shows the main voxels with three peaks that correspond to the crossing between the CC, the CST, and the SLF.

TABLE 1: Computation time. A  $30 \times 30$  voxel slice of the synthetic dataset was used for computations. The numerical Fourier Transform was performed on a coarse spherical mesh with 162 vertices to compute the angular profile of the EAP. The proposed analytical approximate EAPs were computed on both the coarse mesh and a finer spherical mesh with 2562 vertices. The analytical formulation is clearly advantageous.

|      | Numerical coarse | Analytical coarse | Analytical fine |
|------|------------------|-------------------|-----------------|
| Time | 526 s = 8 m 46 s | 10 s              | 73 s            |

approximates the true EAP up to the truncation order of the Taylor series in (10). In this section, we validate the effects of this truncation or approximation on the fiber directions estimated by the approximate EAP. To do this, we again consider the synthetic dataset and look at a region where all three kinds of voxel models are present. We then compute the analytical EAP for approximation orders of  $n = 5, 7, 9$ . We compare the angular profiles of these different approximations at a fixed radius  $|\mathbf{r}|$  to the angular profiles of the fixed approximation at  $n = 7$  for varying radii.

Figure 5 shows the effect of the Taylor expansion order  $n$  on the EAP approximation. The six images are zooms into a region where the two fiber bundles converge and cross. In the top row, we present  $\text{Pr}_{4,4}^{(n)}$ , with  $n = 5, 7, 9$  evaluated for the probability radius  $|\mathbf{r}| = 16 \mu\text{m}$ . Increasing  $n$  adds more terms to the EAP approximation in (11), which adds more corrections to the approximation, making it converge better to the true EAP. As the approximation  $\text{Pr}_{4,4}^{(n)}$  is corrected,

it shows sharper peaks and narrower crossings for  $n = 9$ , than  $n = 5$ , for the same probability radius. However, this also increases the computation time. But the peaks of the lower-order approximations seem to be well aligned with the higher-order approximations. In other words, the peaks maintain their angular alignment, although they lose sharpness, and the EAP loses angular resolution, and narrow crossings become harder to discern. However, the angular resolution can be recovered, and the peaks “sharpened” in the lower-order approximations by increasing the probability radius, which saves computing time. This is shown in the bottom row, where we show  $\text{Pr}_{4,4}^{(7)}$  for the probability radius varying from  $|\mathbf{r}| = 16 \mu\text{m} - 20 \mu\text{m}$ . These experiments reveal that the effect of the Taylor expansion order  $n$  is to underestimate the EAP in the approximation. Therefore, we use the order 7 approximation  $\text{Pr}_{4,4}^{(7)}$ , as a good trade-off between convergence to the true EAP and computation time.

## 5. Conclusion

GDTI was developed to model complex ADC profiles which was an inherent shortcoming of DTI. GDTI uses HOTs of order  $k$  to model a complex ADC geometry. However, the shape of the ADC does not correspond to the underlying fiber directions. The microstructure of the tissue can be inferred from the geometry of the EAP, where in the  $q$ -space formalism, the EAP and the diffusion signal are related by the Fourier transform. But it is not easy to compute the EAP,  $P(\mathbf{r})_k$ , from the HOT model of the signal  $E(\mathbf{q})_k$  in GDTI.

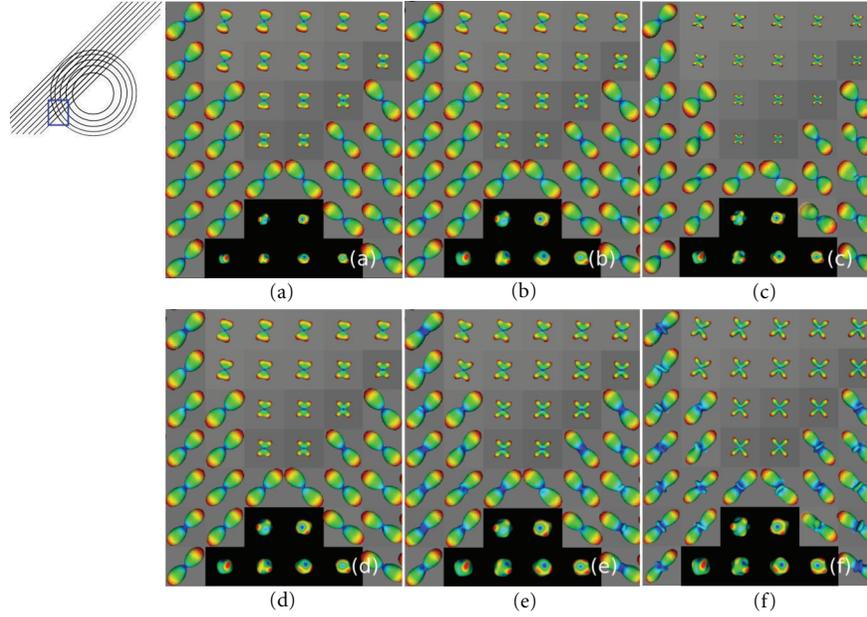


FIGURE 5: Effects of the approximating order  $n$  and the probability radius  $|\mathbf{r}|$ . In top row,  $|\mathbf{r}|$  is fixed, and we vary  $n$ . In bottom row,  $n$  is fixed, and we vary  $|\mathbf{r}|$ . Top row:  $|\mathbf{r}| = 16 \mu\text{m}$ : (a)  $\text{Pr}_{4,4}^{(5)}$ , (b)  $\text{Pr}_{4,4}^{(7)}$ , and (c)  $\text{Pr}_{4,4}^{(9)}$ . Bottom row:  $n = 7$  ( $\text{Pr}_{4,4}^{(7)}$ ): (d)  $\equiv$  (b)  $|\mathbf{r}| = 16 \mu\text{m}$ , (e)  $|\mathbf{r}| = 18 \mu\text{m}$ , and (f)  $|\mathbf{r}| = 20 \mu\text{m}$ .

We overcome this hurdle by modifying the ADC model of GDTI, which allows us to approximate  $E(\mathbf{q})$  by a multivariate polynomial approximation and by proposing a novel closed-form approximation of  $P(\mathbf{r})$  using Hermite polynomials. The solution is a polynomial times a Gaussian; therefore, the polynomial can be interpreted as the correction to the Gaussian EAP due to the inhomogeneous medium. An alternate explanation can be used to explain this method, where the signal is expanded in the polynomial basis, and the EAP is expressed in the Hermite polynomial basis, which establishes the similarity of the proposed method to [22]. Also, since the solution is analytical, it is fast, and the approximation converges well to the true EAP.

In case of an order 4 HOT, this method can be directly adapted to the methods proposed for estimating 4th-order diffusion tensors with positive diffusion profiles. Therefore, it is possible to estimate a 4th-order HOT with a positive diffusion profile using this modified model before approximating the EAP. The experiments show that estimating only the 15 coefficients of a 4th-order HOT are enough to reveal the underlying fiber bundle layout. However, this is dependent on the order of the Taylor expansion used. Although the order of the expansion does not change the angular alignment of the peaks of the approximate EAP, it does affect its angular resolution or its capability of discerning narrow crossings. Increasing of the order increases the corrections to the approximation, which improves this angular resolution. However, it also increases the computation time. The angular resolution can be recovered in lower-order approximations, by increasing the probability radius, which saves computation time. However, this overall effect

indicates that the truncation in the Taylor expansion has the effect of underestimating the true EAP in the approximation.

## Abbreviations

|         |  |
|---------|--|
| CC:     | Corpus callosum                          |
| CST:    | Corticospinal tract                      |
| DOT:    | Diffusion orientation transform          |
| DSI:    | Diffusion spectrum imaging               |
| DTI:    | Diffusion tensor imaging                 |
| DWI:    | Diffusion-weighted image                 |
| EAP:    | Ensemble average propagator              |
| GA:     | Generalized anisotropy                   |
| GDTI:   | Generalized DTI                          |
| HOT:    | Higher-(than 2) order (Cartesian) tensor |
| QBI:    | Q-ball imaging                           |
| ODF:    | Orientation distribution function        |
| PASMRI: | Persistent angular structure MRI         |
| PGSE:   | Pulsed-gradient spin echo                |
| SD:     | Spherical deconvolution                  |
| SE:     | Scaled entropy                           |
| SH:     | Spherical harmonic (basis)               |
| SLF:    | Superior longitudinal fasciculus.        |

## Acknowledgments

The authors would like to thank Dr. A. Anwender from the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, for providing them with the *in vivo* human dataset. A. Ghosh was partially supported

by the ANR Project NucleiPark and the France-Parkinson Association.

## References

- [1] E. Özarslan and T. H. Mareci, “Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging,” *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 955–965, 2003.
- [2] E. Özarslan, B. C. Vemuri, and T. H. Mareci, “Fiber orientation mapping using generalized diffusion tensor imaging,” in *Proceedings of the 2nd IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI ’10)*, pp. 1036–1039, April 2004.
- [3] E. Özarslan, B. C. Vemuri, and T. H. Mareci, “Generalized scalar measures for diffusion MRI using trace, variance, and entropy,” *Magnetic Resonance in Medicine*, vol. 53, no. 4, pp. 866–876, 2005.
- [4] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and J. Van Welden, “High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity,” *Magnetic Resonance in Medicine*, vol. 48, no. 4, pp. 577–582, 2002.
- [5] P. J. Basser, J. Mattiello, and D. Lebihan, “Estimation of the effective self-diffusion tensor from the NMR spin echo,” *Journal of Magnetic Resonance, Series B*, vol. 103, no. 3, pp. 247–254, 1994.
- [6] A. Barmpoutis, B. Jian, and B. C. Vemuri, “Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI,” in *Proceedings of the Information Processing in Medical Imaging (IPMI ’07)*, pp. 308–319, 2007.
- [7] A. Barmpoutis, M. S. Hwang, D. Howland, J. R. Forder, and B. C. Vemuri, “Regularized positive-definite fourth order tensor field estimation from DW-MRI,” *NeuroImage*, vol. 45, no. 1, pp. S153–162, 2009.
- [8] A. Ghosh, R. Deriche, and M. Moakher, “Ternary quartic approach for positive 4th order diffusion tensors revisited,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI ’09)*, pp. 618–621, July 2009.
- [9] A. Ghosh, M. Descoteaux, and R. Deriche, “Riemannian framework for estimating symmetric positive definite 4th order diffusion tensors,” in *Proceedings of the 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI ’08)*, pp. 858–865, 2008.
- [10] A. Barmpoutis and B. C. Vemuri, “Unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints,” in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI ’10)*, pp. 1385–1388, April 2010.
- [11] L. Qi, G. Yu, and E. X. Wu, “Higher order positive semidefinite diffusion tensor imaging,” *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 416–433, 2010.
- [12] A. Fuster, J. van de Sande, L. Astola, C. Poupon, J. Velterop, and B. M. ter Haar Romeny, “Fourth-order tensor invariants in high angular resolution diffusion imaging,” in *Proceedings of the 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI ’11)*, G. H. Zhang and N. Adluru, Eds., pp. 54–63, 2011.
- [13] A. Ghosh, T. Papadopoulos, and R. Deriche, “Biomarkers for HARDI: 2nd and 4th order tensor invariants,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, Barcelona, Spain, 2012.
- [14] T. Schultz, A. Fuster, A. Ghosh, L. Florack, R. Deriche, and L. H. Lim, “Higher-order tensors in diffusion imaging: a survey,” in *Visualization and Processing of Tensors and Higher Order Descriptors For Multi-Valued Data*, B. Burgeth, A. Bartroli, and C. F. Westin, Eds., Springer, New York, NY, USA, 2013.
- [15] V. Wedeen, T. G. Reese, D. Tuch et al., “Mapping fiber orientation spectra in cerebral white matter with fourier-transform diffusion mri,” in *Proceedings of the 8th Scientific Meeting and Exhibition of the International Society for the Magnetic Resonance in Medicine*, vol. 8, p. 82, 2000.
- [16] D. S. Tuch, “Q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1358–1372, 2004.
- [17] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, “Regularized, fast, and robust analytical Q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 3, pp. 497–510, 2007.
- [18] K. M. Jansons and D. C. Alexander, “Persistent angular structure: new insights from diffusion magnetic resonance imaging data,” *Inverse Problems*, vol. 19, no. 5, pp. 1031–1046, 2003.
- [19] J. D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly, “Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution,” *NeuroImage*, vol. 23, no. 3, pp. 1176–1185, 2004.
- [20] E. Özarslan, T. M. Shepherd, B. C. Vemuri, S. J. Blackband, and T. H. Mareci, “Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT),” *NeuroImage*, vol. 31, no. 3, pp. 1086–1103, 2006.
- [21] P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford University Press, Oxford, UK, 1993.
- [22] C. Liu, R. Bammer, and M. E. Moseley, “Generalized diffusion tensor imaging (GDTI): a method for characterizing and imaging diffusion anisotropy caused by non-Gaussian diffusion,” *Israel Journal of Chemistry*, vol. 43, no. 1-2, pp. 145–154, 2003.
- [23] M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander, “Deterministic and probabilistic tractography based on complex fibre orientation distributions,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 2, pp. 269–286, 2009.
- [24] A. Anwander, M. Tittgemeyer, D. Y. Von Cramon, A. D. Friederici, and T. R. Knösche, “Connectivity-based parcellation of Broca’s area,” *Cerebral Cortex*, vol. 17, no. 4, pp. 816–825, 2007.
- [25] T. G. Reese, O. Heid, R. M. Weisskoff, and V. J. Wedeen, “Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo,” *Magnetic Resonance in Medicine*, vol. 49, no. 1, pp. 177–182, 2003.
- [26] P. A. Cook, Y. Bai, N. S. Gilani et al., “Camino: open-source diffusion-MRI reconstruction and processing,” in *Proceedings of the 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, p. 2759, May 2006.

## Research Article

# The Smoothing Artifact of Spatially Constrained Canonical Correlation Analysis in Functional MRI

Dietmar Cordes,<sup>1,2,3</sup> Mingwu Jin,<sup>3,4</sup> Tim Curran,<sup>2</sup> and Rajesh Nandy<sup>5</sup>

<sup>1</sup>Department of Physics, Ryerson University, Toronto, ON, Canada M5B 2K3

<sup>2</sup>Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309, USA

<sup>3</sup>Department of Radiology, University of Colorado, Denver, CO 80045, USA

<sup>4</sup>Department of Physics, University of Texas, Arlington, TX 76019, USA

<sup>5</sup>Departments of Biostatistics and Psychology, UCLA, Los Angeles, CA 90095, USA

Correspondence should be addressed to Dietmar Cordes, dc689599@gmail.com

Received 19 March 2012; Revised 13 November 2012; Accepted 26 November 2012

Academic Editor: Carlos Alberola-Lopez

Copyright © 2012 Dietmar Cordes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A wide range of studies show the capacity of multivariate statistical methods for fMRI to improve mapping of brain activations in a noisy environment. An advanced method uses local canonical correlation analysis (CCA) to encompass a group of neighboring voxels instead of looking at the single voxel time course. The value of a suitable test statistic is used as a measure of activation. It is customary to assign the value to the center voxel; however, this is a choice of convenience and without constraints introduces artifacts, especially in regions of strong localized activation. To compensate for these deficiencies, different spatial constraints in CCA have been introduced to enforce dominance of the center voxel. However, even if the dominance condition for the center voxel is satisfied, constrained CCA can still lead to a smoothing artifact, often called the “bleeding artifact of CCA”, in fMRI activation patterns. In this paper a new method is introduced to measure and correct for the smoothing artifact for constrained CCA methods. It is shown that constrained CCA methods corrected for the smoothing artifact lead to more plausible activation patterns in fMRI as shown using data from a motor task and a memory task.

## 1. Introduction

Local canonical correlation analysis (CCA) is a multivariate statistical method in fMRI that uses the joint time course of a group of neighboring voxels, usually in a  $3 \times 3$  in-plane voxel grid, to determine the significance of activation. The value of a suitable test statistic is used as a measure of activation. Since the joint time course of the neighborhood is used, it is not immediately clear to which voxel the measure of activation should be assigned. For example, if a  $3 \times 3$  voxel neighborhood is chosen and the measure of activation is significant, without further assumptions one can only conclude that activation occurred somewhere within the  $3 \times 3$  voxel neighborhood. If the activation is assigned to all voxels of the neighborhood, loss of spatial specificity will occur. To increase spatial specificity, it has been proposed to assign the measure of activation to the center voxel of

the  $3 \times 3$  neighborhood [1, 2]. A center voxel assignment is usually justified by mathematical convenience but can also be reasoned on the fact that the fMRI BOLD response leads to patches of activation patterns that are most likely of convex shape and simple connectivity (without any holes in the interior neighborhood). However, this center voxel assignment proved to be prone to yield artifacts as activations tend to bleed to the neighboring voxels of strongly active voxels. The result is a loss of spatial specificity from this smoothing artifact.

The smoothing artifact is not only common in conventional CCA, but also in any analysis technique that involves spatial low-pass filter kernels, such as univariate (single voxel) analysis where the data have been preprocessed using Gaussian spatial smoothing. In conventional data smoothing, the smoothing artifact has been intentionally “induced” to increase the signal-to-noise ratio at the cost of

reduced specificity and occurrence of typical spatial low-pass artifacts such as blurring of edges of activation patterns.

To compensate for the smoothing artifact in conventional CCA, different assignment schemes were proposed. For example, a minimum relative weight for the center voxel was used to restrict false activations [3]. In another study using a more adaptive approach, the smoothing artifact was reduced by utilizing the spatial dependence among voxels as much as possible and assigning the significance of activation to the dominant voxel of local maxima [4]. This method was shown to be effective in eliminating the smoothing artifact in motor activation data that is known to have large contrast-to-noise ratio (CNR), however, in data where the activation is more subtle (such as hippocampal activation using an episodic memory paradigm), the method has the disadvantage of being less sensitive, according to our studies.

To reduce the smoothing artifact in CCA, it is necessary to constrain the spatial weights properly and impose the condition that the center voxel always has the largest weight. Constrained CCA (cCCA) with positivity constraints have been proposed for fMRI. Friman et al. [5] as well as Ragnehed et al. [6] use nonnegative spatial weights with maximum weight of the center voxel in order to ensure spatial low-pass filter properties of cCCA. This has the additional benefit of constraining CCA to eliminate spurious correlations occurring in conventional CCA where spatial filters can have positive and negative coefficients.

To our knowledge, the smoothing artifact in cCCA has never been studied. Recently, we provided a mathematical framework for cCCA and computed ROC properties of cCCA with different linear constraints and a nonlinear constraint for activation patterns of motor data and episodic memory data [7, 8]. In this paper we expand our previous research and investigate in detail the smoothing artifact that is associated with each spatial constraint in cCCA. Furthermore, we provide a novel approach of how to correct the measure of activation for the smoothing artifact. Results for motor activation data and episodic memory activation data are presented. Parts of this paper have been published in abstract form (one page) at a recent conference [9].

## 2. Theory

**2.1. Constrained CCA (cCCA).** In the following we briefly review CCA and cCCA, and explicitly consider the constraints introduced recently [8]. Mathematically, CCA is a generalization of the General Linear Model (GLM) by allowing the incorporation of spatial basis functions according to

$$\begin{aligned} & (\alpha_1 f_1(\xi) + \dots + \alpha_s f_s(\xi)) \otimes Y(\xi, t) \\ & = \beta_1 x_1(t) + \dots + \beta_r x_r(t) + \varepsilon(t), \end{aligned} \quad (1)$$

where the data are given by  $Y(\xi, t)$ ,  $\xi$  is the vector representing the spatial coordinates  $x$ ,  $y$ , and  $z$ , and  $t$  is time. The functions  $f_i(\xi)$ ,  $i = 1, \dots, s$  represent the spatial basis functions modeling the activation pattern in a neighborhood. The functions  $x_j(t)$ ,  $j = 1, \dots, r$  are the temporal basis functions modeling the signal observed (which is the result of a convolution of the hemodynamic response function and the

stimulus function). The coefficients  $\alpha_i$  and  $\beta_j$  are the spatial and temporal weights, respectively, that are being determined and optimized by the data for each individual neighborhood using an optimization routine. The symbol  $\otimes$  denotes spatial convolution and  $\varepsilon(t)$  is a Gaussian-distributed random error term. If the number of spatial basis functions is reduced to a single function, (1) becomes

$$f_1(\xi) \otimes Y(\xi, t) = \beta_1 x_1(t) + \dots + \beta_r x_r(t) + \varepsilon(t). \quad (2)$$

When  $f_1(\xi)$  is a simple Gaussian function, we obtain the conventional GLM used frequently in fMRI.

Equation (1) can be represented conveniently in matrix form. In the following we assume that the functions  $f_i(\xi)$ ,  $i = 1, \dots, s$  are spatial Dirac delta functions defining a local neighborhood within a  $3 \times 3$  pixel neighborhood ( $s \leq 9$ ). Let  $\mathbf{Y}$  be the matrix representing  $s$  voxel time courses with dimension  $t \times s$  and  $\mathbf{X}$  the conventional design matrix of size  $t \times r$  for the  $r$  temporal regressors. Furthermore, let  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  be two unknown vectors of size  $s \times 1$  and  $r \times 1$ , respectively. In CCA, we look for the linear combinations of voxel time courses  $\mathbf{Y}\boldsymbol{\alpha}$  and temporal regressors  $\mathbf{X}\boldsymbol{\beta}$  such that the correlation between both quantities is maximum. This leads to an eigenvalue problem with  $\min(s, r)$  solutions from which the solution with the largest eigenvalue (i.e., maximum canonical correlation) is being chosen. Without constraints on the  $\alpha_i$ , the specificity of the activation pattern obtained by CCA is low and could result in artifacts (see e.g., [8]). To put constraints on the spatial weights  $\boldsymbol{\alpha}$  in order to restrict the space of unreasonable solutions for fMRI, we consider the following four scenarios for the components  $\alpha_i$  of  $\boldsymbol{\alpha}$ , where  $\alpha_1$  is the weight for the center voxel and the other  $\alpha_i$ 's represent the weights for the  $s$  neighborhood voxels.

Constraint 1 (Simple Constraint). One has

$$\alpha_1 > 0, \quad \alpha_i \geq 0 \quad \forall i \geq 2. \quad (3)$$

Constraint 2 (Sum Constraint). One has

$$\alpha_1 \geq \sum_{i=2}^s \alpha_i > 0, \quad \alpha_i \geq 0 \quad \forall i \geq 2. \quad (4)$$

Constraint 3 (Average Constraint). One has

$$\alpha_1 \geq \frac{1}{s-1} \sum_{i=2}^s \alpha_i > 0, \quad \alpha_i \geq 0 \quad \forall i \geq 2. \quad (5)$$

Constraint 4 (Maximum Constraint). One has

$$\alpha_1 \geq \max(\alpha_i) > 0, \quad \alpha_i \geq 0 \quad \forall i \geq 2. \quad (6)$$

Note that the neighborhood size  $s$  is not a fixed quantity, but is determined from the data by cCCA and can differ for each center voxel.

**2.2. Smoothing Artifact.** The smoothing artifact in CCA is defined as the probability of incorrectly declaring the center voxel of a configuration of size  $s$  ( $s \leq 9$  for a  $3 \times 3$  neighborhood) to be active. In the following, we outline how to compute the posterior probability to detect the smoothing artifact in real data using a Bayesian framework. The posterior probability,  $P$ , that a center voxel is not active when it was in fact declared active, is given by

$$P = p(\text{center voxel is not active} \mid \omega > \omega_0, \text{cnr}, M, \text{CNR}, s), \quad (7)$$

where  $\omega > \omega_0$  indicates that the center voxel was declared active (statistic  $\omega > \text{threshold } \omega_0$  with  $\omega \in [0, \infty)$ ),  $\text{cnr}$  is the univariate contrast-to-noise ratio of the center voxel,  $M$  labels the method of data analysis,  $\text{CNR}$  is the contrast-to-noise ratio of the entire configuration defining the neighborhood within a  $3 \times 3$  pixel region, and  $s$  is the size of the configuration (i.e., number of declared active voxels  $\leq 9$  within the neighborhood). For abbreviation, we define the set of parameters,  $\theta$ , to be

$$\theta = \{\text{cnr}, M, \text{CNR}, s\}. \quad (8)$$

Then, according to Bayes' theorem for conditional probabilities, (7) can be written as

$$P = p(\omega > \omega_0 \mid \text{center voxel is not active}, \theta) \times p(\text{center voxel is not active} \mid \theta) (p(\omega > \omega_0 \mid \theta))^{-1}, \quad (9)$$

which is of the form

$$P = \frac{P_1 P_3}{P_2}, \quad (10)$$

where

$$P_1 = p(\omega > \omega_0 \mid \text{center voxel is not active}, \theta), \quad (11)$$

$$P_2 = p(\omega > \omega_0 \mid \theta), \quad (12)$$

$$P_3 = p(\text{center voxel is not active} \mid \theta). \quad (13)$$

The term  $P_1$  is called the bleeding artifact because it represents the probability that an inactive voxel is declared as active. We determine  $P_1$  as a function of the size,  $s$ , of the configuration only and not as a function of the geometrical shape of the configuration. Note that the dependence on  $s$  is an approximation, because in reality there are  $2^8 = 256$  possible configurations that can contain 0 to 8 active voxels (corresponding to  $s \in \{1, \dots, 9\}$  since  $s$  labels the neighborhood size within a  $3 \times 3$  pixel grid, which always includes the center voxel, independent if the center voxel is active or not). Each configuration of size  $s$  has, depending on its distance of all voxel members to the center voxel, a slightly different value for  $P_1$ . For example, configurations with  $s = 7$  leads to 3 different classes based on a distance measure, that is, class 1 = {center voxel, 4 corner voxels, and 2 midedge voxels}, class 2 = {center voxel, 3 corner voxels,

and 3 midedge voxels}, class 3 = {center voxel, 2 corner voxels, and 4 midedge voxels}.

According to our simulations,  $P_1$  is strongly dependent on  $s$  but not on a particular configuration of  $s$ . Only a weak dependence based on different class memberships exist, which we neglect for the purpose of this research. To estimate  $P_1$ , it is thus reasonable to group all configurations for a particular  $s$  together and compute an average value of  $P_1$  over all possible configurations with size  $s$ .

**2.2.1. Estimation of  $P_1$**  (See (11)). The term  $P_1$  can be estimated from simulations using a mixture of resampled resting-state data and activation data at given  $\theta$  using kernel density estimation [10]. Resampled resting-state data are considered null data with respect to any task fMRI function since the temporal structure is destroyed by resampling using the wavelet transform. This resampling, however, does not destroy the autocorrelations inherent in resting-state data. Furthermore, the resampling does not affect the spatial correlations within the data because the permutations of the wavelet coefficients are kept the same for each voxel time series in a particular simulation; however, different simulations use different permutations [11, 12].

The simulated data are superpositions of time series from a  $3 \times 3$  pixel neighborhood of null data and activation data. Since the entire neighborhoods are used from resting-state data, realistic spatial correlations of the simulated data are obtained. In particular, for a configuration of  $s$  active voxels in the  $3 \times 3$  neighborhood, the simulated voxel time courses,  $y_i(t)$ , are obtained by

$$y_i(t) = \begin{cases} y_i^{(0)}(t), & \text{for } i = 1, \\ \beta x(t) + y_i^{(0)}(t), & \text{for } i \in \{2, \dots, s\}, \end{cases} \quad (14)$$

where  $i = 1$  refers to the center voxel and all other  $i$  to the surrounding voxels of the configuration of size  $s$  within the  $3 \times 3$  neighborhood. All  $y_i^{(0)}(t)$  correspond to resampled resting-state time courses and represent spatially and temporally correlated null (noise) data. Note that the center voxel is always inactive by design to compute  $P_1$ . Thus,  $P_1$  is a strong function of  $\text{CNR}$  of the configuration but not of the value  $\text{cnr}$  (which is the contrast-to-noise ratio of the inactive center voxel), and the dependence of  $P_1$  on  $\text{cnr}$  can be neglected. The activation is determined by the hemodynamic response function,  $x(t)$ , of interest multiplied by factor  $\beta$  so that the configuration has a given  $\text{CNR}$ . In order to compute the  $\text{CNR}$  we use the general definition

$$\text{CNR} = \left( \frac{\sum \lambda_i}{\sum \zeta_i} \right)^{1/2}, \quad (15)$$

where  $\lambda_i$  and  $\zeta_i$  are the eigenvalues of the covariance matrix of the activation signal and noise, respectively [13]. Note that (15) can be used for a single voxel time series or an entire neighborhood of arbitrary size. To determine the activation signal and noise of a configuration using cCCA, we convert the cCCA problem into a multivariate multiple regression problem of the form

$$\mathbf{Y}\boldsymbol{\alpha} = \mathbf{X}\mathbf{B}\boldsymbol{\alpha} + \mathbf{E}\boldsymbol{\alpha}, \quad (16)$$

where  $\mathbf{Y}$  are the data (size  $t \times s$ ),  $\boldsymbol{\alpha}$  is the optimum spatial weight vector (size  $s \times 1$ ),  $\mathbf{X}$  is the design matrix (size  $t \times r$ ),  $\mathbf{B}$  is the matrix of regression weights (size  $r \times s$ ), and  $\mathbf{E}$  is a residual error matrix (size  $t \times s$ ). For a given contrast vector  $\mathbf{c}$ , we reparameterize the design matrix  $\mathbf{X}$  and obtain a transformed design matrix  $\tilde{\mathbf{X}}$  such that

$$\tilde{\mathbf{X}} = [\mathbf{X}_{\text{eff}} \mathbf{X}_{\perp}], \quad (17)$$

where

$$\mathbf{X}_{\text{eff}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{-1} \quad (18)$$

is the first regressor of the new design matrix  $\tilde{\mathbf{X}}$  that is associated with a parameter estimate equivalent to the original contrast  $\mathbf{c}'\mathbf{B}\boldsymbol{\alpha}$  [8, 14]. The matrix  $\mathbf{X}_{\perp}$  is perpendicular to  $\mathbf{X}_{\text{eff}}$  and plays no role in the estimation of  $\mathbf{c}'\boldsymbol{\beta}$ . Then, the signal  $\mathbf{S}(t)$  is obtained by

$$\mathbf{S} = \mathbf{X}_{\text{eff}}\mathbf{B}, \quad (19)$$

and the noise  $\mathbf{N}(t)$  is obtained by

$$\mathbf{N} = (\mathbf{Y} - \mathbf{X}_{\text{eff}}\mathbf{B})\boldsymbol{\alpha}. \quad (20)$$

**2.2.2. Estimation of  $P_2$  (See (12)).** This term can be estimated directly from the real data. In this case, for each  $M$  and  $s > 1$ ,  $P_2$  is a 2D function of cnr and CNR, but depends strongly only on CNR so that the dependence on cnr can be neglected. Note that for  $s = 1$ ,  $\text{cnr} = \text{CNR}$ , and in this case  $P_2$  is a 1D function of cnr only. It is possible to determine first the joint probability density  $p(\omega, \text{CNR} \mid s, M)$  using 2D kernel density estimation with a 2D Gaussian kernel, which then can be integrated numerically to obtain  $P_2$  according to

$$P_2(\omega_0, \text{CNR}, s, M) = \frac{\int_{\omega_0}^{\infty} p(\omega, \text{CNR} \mid s, M) d\omega}{\int_0^{\infty} p(\omega, \text{CNR} \mid s, M) d\omega}. \quad (21)$$

Note that  $P_2(\omega_0, \text{CNR}, s, M)$  for fixed  $\{\omega_0, s, M\}$  has a sigmoidal shape approaching the value 1 for  $\text{CNR} > 0.6$ . Thus, voxels that are declared active at a family-wise error rate (FWE)  $< 0.05$  have necessarily a large CNR for which  $P_2(\omega_0, \text{CNR}, s, M) \rightarrow 1$  (see Section 4).

**2.2.3. Estimation of  $P_3$  (See (13)).** The term  $P_3$  is less difficult to determine because it is independent of the value of the statistic  $\omega$  and depends strongly on the univariate cnr of the center voxel (configuration with size  $s = 1$ , and  $M = 1$ ), that is,

$$\begin{aligned} P_3 &= p(\text{center voxel is not active} \mid \theta) \\ &\approx p(\text{center voxel is not active} \mid \text{cnr}, s = 1, M = 1) \\ &= \frac{p(\text{center voxel is not active, cnr} \mid s = 1, M = 1)}{p(\text{cnr} \mid s = 1, M = 1)}, \end{aligned} \quad (22)$$

where  $M = 1$  labels the univariate single voxel analysis method without smoothing. Then,  $P_3$  is only a function of

cnr and can be estimated from linear mixture modeling of the real data assuming that the data consists only of active and inactive voxels with unknown fractions. With this assumption, we define the cnr distribution of the data as  $h(\text{cnr})$ , consisting of the mixtures  $f(\text{cnr})$  and  $G_{\mu, \sigma}(\text{cnr})$  using

$$h(x) = a \frac{1}{d} f\left(\frac{\text{cnr}}{d}\right) + (1 - a)G_{\mu, \sigma}(\text{cnr}). \quad (23)$$

The distribution  $f(\text{cnr})$  is estimated from resampled resting-state data and the scaled distribution  $(1/d)f(\text{cnr}/d)$  reflects the null distribution in activation data. The fact that  $f(\text{cnr})$  is scaled by constant  $d$  is rooted in the observation that in activation data more neural activity exists and maybe by spatial correlations or other hemodynamic means the distribution of the signal corresponding to inactive voxels is shifted to slightly larger values of cnr. The second term on the right in (23),  $G_{\mu, \sigma}(\text{cnr})$ , represents the cnr distribution of active voxels modeled by a Gaussian distribution with mean  $\mu$  and variance  $\sigma$ . All the parameters  $a, d, \mu$ , and  $\sigma$  are obtained from least squares fitting using activation data. Then,

$$P_3(\text{cnr}) = \frac{a(1/d)f(\text{cnr}/d)}{h(\text{cnr})}. \quad (24)$$

**2.2.4. Final Result of Estimation of  $P$  (See (9) and (10)).** Overall, the posterior probability that a center voxel is not active is given by

$$\begin{aligned} P &= p(\text{center voxel is not active} \mid \omega > \omega_0, \theta) \\ &\approx \frac{p(\omega > \omega_0 \mid \text{center voxel is not active}, \theta)P_3(\text{cnr})}{p(\omega > \omega_0 \mid \theta)} \quad (25) \\ &\approx P_1(\omega_0, \text{CNR}, M, s)P_3(\text{cnr}), \end{aligned}$$

since  $p(\omega > \omega_0 \mid \theta) \rightarrow 1$  for voxels declared highly active (i.e., FWE  $< 0.05$ ). In the following, we call the function  $P_1(\omega_0, \text{CNR}, M, s)$  the smoothing artifact function. To correct for the smoothing artifact we propose the rule:

$$\text{Voxel is assigned to be inactive if } P > 0.5 \quad (26)$$

and assign zero to the measure of activation if this statement is true. If this statement is not true, the measure of activation is unchanged.

### 3. Materials and Methods

fMRI was performed for 6 normal subjects with IRB approval (according to institutional requirements) in a 3.0T GE HDx MRI scanner equipped with an 8-channel head coil and parallel imaging acquisition using EPI with imaging parameters: ASSET = 2, ramp sampling, TR/TE = 2 sec/30 ms, FA = 70 deg, FOV = 22 cm  $\times$  22 cm, thickness/gap = 4 mm/1 mm, 25 slices, and resolution 96  $\times$  96. Three fMRI data sets were obtained for each subject. In the following we briefly describe the paradigms and refer the reader for more detail to our previous article [7].

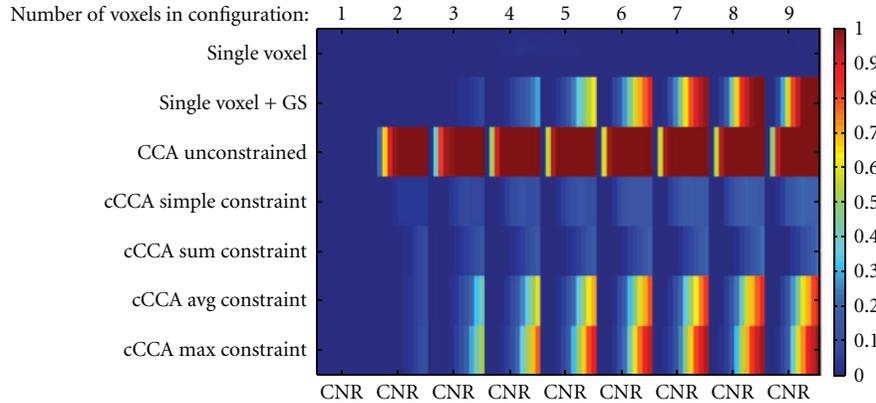


FIGURE 1: Smoothing artifact of the center voxel for different analysis methods as a function of the configuration size and CNR for simulated motor activation data. Note that the figure is composed of 9 separate images belonging to configuration sizes 1 to 9, and each image for a particular configuration size is a function of the CNR ranging from 0 to 1 in steps of 0.1 in the horizontal direction. The center voxel for each configuration is by design nonactive and the other members (neighboring voxels belonging to the particular configuration) are active with given CNR. In general the smoothing artifact increases with increasing CNR except for single-voxel analysis, where no artifact exists (as expected). Regarding the computation of the false positive fraction, the statistical thresholds were chosen corresponding to a family-wise error rate (FWE)  $< 0.05$  for re-sampled resting-state data.

The first data set was collected during resting-state where the subject tried to relax and refrain from executing any overt task with eyes closed. The second data set was collected while the subject was performing an episodic memory task with oblique coronal slices collected perpendicular to the long axis of the hippocampus. Specifically, this task consisted of memorization of novel faces paired with occupations and contained 6 periods of encoding, distraction, and recognition tasks as well as short instructions where words on the screen reminded subjects of the task ahead. The third data set was obtained by performing an event-related motor task involving bilateral finger tapping while the subject was looking at a screen.

**3.1. Data Analysis.** All fMRI data were realigned using Statistical Parametric Mapping (SPM5, <http://www.fil.ion.ucl.ac.uk/spm/>) and maximum motion components were found to be less than 0.6 mm in all directions. In a preprocessing step, all voxel time series were corrected for different slice timings and high-pass filtered by regression using a discrete cosine basis with cut-off frequency 1/120 Hz [15]. No temporal low-pass filtering was carried out. All voxels with intensity larger than 10% of the mean intensity were used in the analysis. This threshold effectively eliminated all nonbrain voxels leading to an average of about 4500 voxels per slice. All activation maps were thresholded using a FWE  $< 0.05$  determined by using nonparametric methods [7, 16] with wavelet resampled resting-state data [11, 12].

**3.2. Basis Functions for CCA.** All voxel time courses and temporal regressors were mean subtracted (over time) and variance normalized. As local spatial basis functions we use Dirac delta functions in each  $3 \times 3$  in-slice neighborhood.

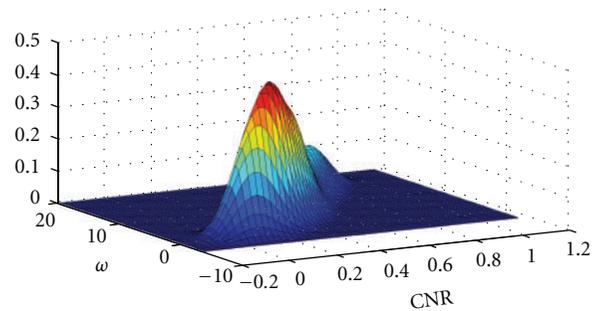


FIGURE 2: Example of the calculated joint probability density  $p(\omega, \text{CNR} | s, M)$  for motor activation data for a neighborhood size  $s = 5$  and cCCA with the maximum constraint. The function was determined using 2D kernel density estimation with a 2D Gaussian kernel. The variables  $\omega$  and CNR specify the statistic and the contrast-to-noise ratio of the 5-voxel neighborhood, respectively. The threshold to obtain FWE  $< 0.05$  is  $\omega_0 = 6.7$ .

For the temporal modeling, we specified design matrices as in SPM5 containing all conditions of the paradigms. In particular, for the memory paradigm we modeled instruction (I), encoding (E), recognition (R), and control (C) by temporal reference functions whereas for the motor paradigm, fixation (F), and motor task (M) were modeled according to the paradigm timings. All reference functions were convolved as usual with the standard SPM5 two-gamma hemodynamic response function. For the motor task we computed activation maps for the contrast M-F, and for the memory task we used the contrast E-C. We used reparameterization of the design matrix  $\mathbf{X}$  (see (18)) to incorporate the contrast of interest and optimized the spatial coefficients for each spatial constraint using the methods proposed in our previous publication [8].

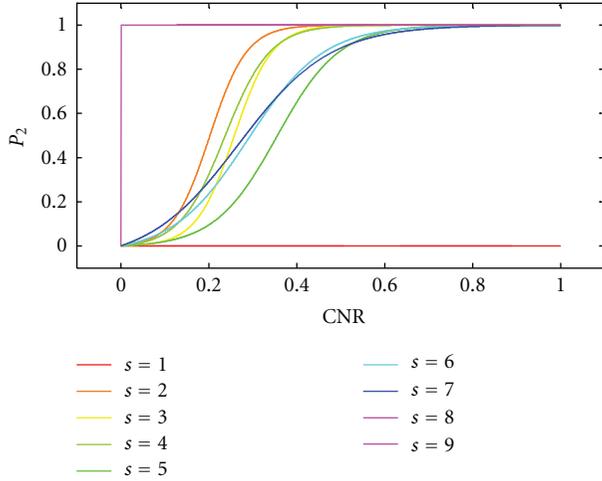


FIGURE 3: Estimation of  $P_2(\omega_0, \text{CNR}, s, M) = \int_{\omega_0}^{\infty} p(\omega, \text{CNR} | s, M) d\omega / \int_0^{\infty} p(\omega, \text{CNR} | s, M) d\omega$  from motor activation data using numerical integration for all configurations of size  $s \in \{1, \dots, 9\}$  for cCCA with the maximum constraint. The joint probability density  $p(\omega, \text{CNR} | s, M)$  in the integrands are determined with 2D kernel density estimation using a 2D Gaussian kernel (for an example see Figure 2). Note that for  $\text{CNR} > 0.6$ ,  $P_2$  approaches the value 1 rapidly for all  $s \neq 1$ . This relationship is also true for data obtained from the memory experiment. Note that configurations with  $s = 1$  have no significance in contributing to the smoothing artifact because the smoothing artifact is by definition equal to zero for  $s = 1$ .

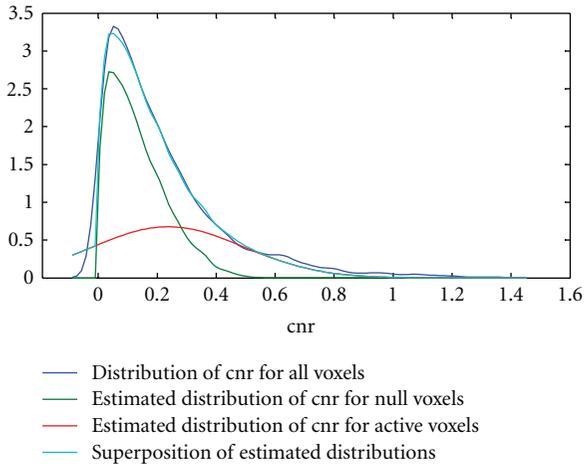


FIGURE 4: Density estimation of active and inactive voxels in motor activation data. The top curve (dark blue) shows the distribution of the  $\text{cnr}$  using kernel density estimation techniques. The green and red curves are the estimated distributions of the  $\text{cnr}$  for the inactive and active voxels, respectively. Note that the distribution of the inactive voxels (green curve) was derived from wavelet resampled resting-state data using a dilation variable, whereas the distribution of the active voxels (red curve) was derived from a Gaussian distribution. The mixture of the estimated distributions for null and active voxels is given by the light blue curve showing very small differences to the raw curve (dark blue).

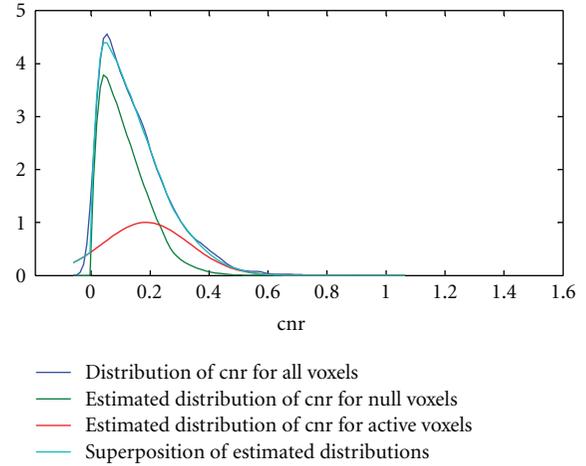


FIGURE 5: Density estimation of active and inactive voxels in memory activation data. The top curve (dark blue) shows the distribution of the  $\text{cnr}$  using kernel density estimation techniques. The green and red curves are the estimated distributions of the  $\text{cnr}$  for the inactive and active voxels, respectively. Note that the distribution of the inactive voxels (green curve) was derived from wavelet resampled resting-state data using a dilation variable, whereas the distribution of the active voxels (red curve) was derived from a Gaussian distribution. The mixture of the estimated distributions for null and active voxels is given by the light blue curve showing very small differences to the raw curve (dark blue).

## 4. Results and Discussion

**4.1. The Smoothing Artifact Function.** Using simulated data, the smoothing artifact function  $P_1$  (see (11)) was determined for the motor paradigm with contrast M-F and memory paradigm with contrast E-C, respectively. Simulations were carried out for all methods  $M$  (single voxel analysis, single voxel analysis with Gaussian spatial smoothing, unconstrained CCA, CCA with the simple constraint, CCA with the sum constraint, CCA with the average constraint, and CCA with the maximum constraint),  $\text{CNR}$  in the range  $[0, 1]$  in steps of 0.1, and configuration sizes 1 to 9. All possible 256 configurations in a  $3 \times 3$  neighborhood with inactive center voxel were simulated 1000 times and then regrouped according to the sizes  $s = \{1, \dots, 9\}$ . Figure 1 shows the smoothing artifact function for the motor paradigm for a typical subject. An almost identical figure was obtained for the memory paradigm. The threshold  $\omega_0$  corresponds to  $\text{FWE} = 0.05$ . Please note that this figure is a composition of nine different images where each image belongs to a configuration of a particular size (1 to 9) and each abscissa is the  $\text{CNR}$  ranging from 0 to 1 in steps of 0.1. The vertical axis labels the different analysis methods applied and the color determines the value of  $P_1$ , ranging from 0 to 1. Bluish color indicates that the smoothing artifact is negligible whereas red color indicates that the smoothing artifact is significant. It is obvious that single-voxel analysis without Gaussian smoothing does not show any smoothing artifact and single-voxel analysis with spatial smoothing leads to a significant smoothing artifact, the larger the  $\text{CNR}$  and the larger

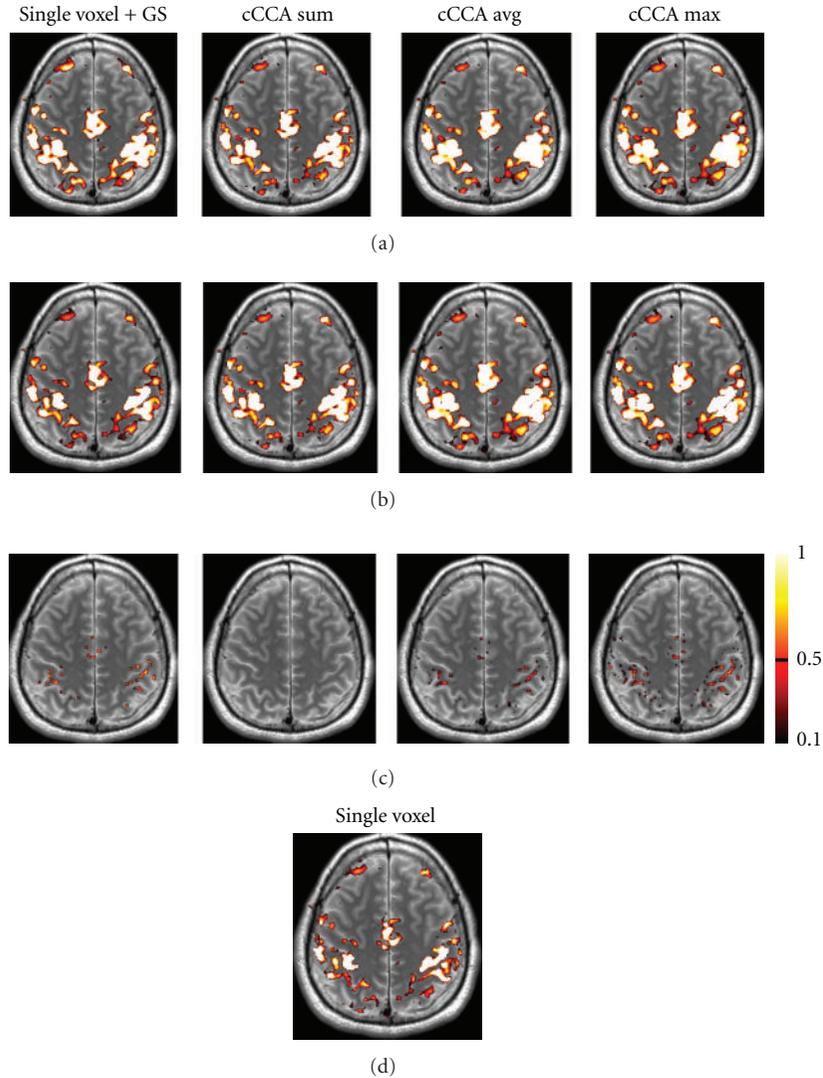


FIGURE 6: Motor activation maps for contrast “motor” minus “fixation” using different data analysis methods (single voxel with Gaussian smoothing (GS), CCA with the sum constraint (cCCA sum), CCA with the average constraint (cCCA avg), CCA with the maximum constraint (cCCA max)). In (a), original activations maps are shown at FWE  $< 0.05$ . In (b), activation maps corrected for the smoothing artifact are shown. Corrections are done for  $P > 0.5$ . In (c), voxels affected by the smoothing artifact are shown. The color scale on the right refers to the magnitude of the smoothing artifact in (c). In (d) we show for comparison the activation map for single voxel analysis without Gaussian smoothing.

the neighborhood is. It is also obvious that unconstrained CCA has the largest smoothing artifact and this artifact is already large for configuration sizes of  $s = 2$  and  $\text{CNR} = 0.2$ . However, choosing the simple nonnegativity constraint for cCCA almost completely eliminates the smoothing artifact ( $P_1 < 0.35$ ). Similarly, cCCA with the sum constraint has a smoothing artifact that is even lower ( $P_1 < 0.3$ ) and should be considered the method of choice if a high specificity is desirable. The cCCA methods with the more complicated constraints (avg constraint and max constraint) show a significant smoothing artifact for configuration sizes of  $s \geq 3$ , as long as the CNR is large ( $\text{CNR} > 0.6$ ). These two cCCA methods have very high sensitivity but can lead to false activations when the configuration size is large.

4.2. *Density Estimation of  $P_2$*  (See (21)). The function  $P_2(\omega_0, \text{CNR}, s, M) = \int_{\omega_0}^{\infty} p(\omega, \text{CNR} | s, M) d\omega / \int_0^{\infty} p(\omega, \text{CNR} | s, M) d\omega$  was calculated in MATLAB (<http://www.mathworks.com/>) by 2D kernel density estimation of  $p(\omega, \text{CNR} | s, M)$  using an optimum bandwidth estimator according to Sheather and Jones [17]. In general,  $p(\omega, \text{CNR} | s, M)$  has a bimodal distribution for configuration sizes  $s \in \{2, 3, 4, 5, 6, 7\}$ . For lower  $s$ , the larger mode of the density occurs at lower values of  $\{\omega, \text{CNR}\}$ , whereas for larger values of  $s$ , the larger mode occurs at higher values of  $\{\omega, \text{CNR}\}$ . For  $s \in \{8, 9\}$ , the density becomes unimodal with mode located at large values of  $\{\omega, \text{CNR}\}$ . Also note, that  $\omega$  is strongly correlated with  $\text{CNR}$ , which is expected. An example of  $p(\omega, \text{CNR} | s, M)$  is given in Figure 2 for

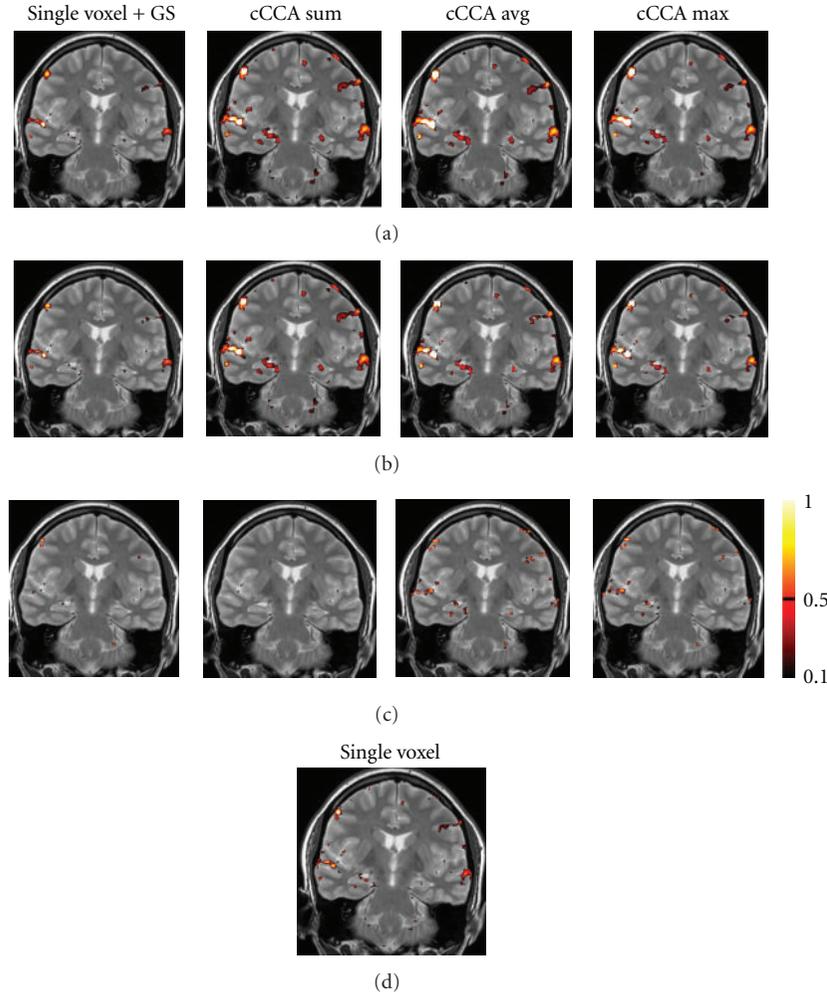


FIGURE 7: Memory activation maps for contrast “encoding” minus “control” using different data analysis methods (single voxel with Gaussian smoothing (GS), CCA with the sum constraint (cCCA sum), CCA with the average constraint (cCCA avg), and CCA with the maximum constraint (cCCA max)). In (a), original activations maps are shown at FWE  $< 0.05$ . In (b), activation maps corrected for the smoothing artifact are shown. Corrections are done for  $P > 0.5$ . In (c), voxels affected by the smoothing artifact are shown. The color scale on the right refers to the magnitude of the smoothing artifact in (c). In (d) we show for comparison the activation map for single-voxel analysis without Gaussian smoothing.

$s = 5$  and cCCA with the maximum constraint. The shape of  $P_2(\omega_0, \text{CNR}, s, M)$  obtained by numerical integration of  $\int_{\omega_0}^{\infty} p(\omega, \text{CNR} | s, M) d\omega / \int_0^{\infty} p(\omega, \text{CNR} | s, M) d\omega$  and density smoothing is shown in Figure 3 for all  $s$  and  $0 \leq \text{CNR} \leq 1$ . Note the S-shaped form obtained for  $P_2(\omega_0, \text{CNR}, s, M)$  for all integrations of bimodal distributions involving  $p(\omega, \text{CNR} | s, M)$ , whereas for  $s = 1$  the function  $P_2$  is zero for  $\text{CNR} \leq 1$  and for  $s \in \{8, 9\}$   $P_2$  has the value 1 for  $0 < \text{CNR} \leq 1$ . The function  $P_2$  for size  $s = 1$  plays no role in determining the posterior probability  $P$  because the smoothing artifact is zero by definition, since a single-voxel-neighborhood cannot have any bleeding of signal strength.

**4.3. Density Estimation of the Null  $\text{cnr}$  Distribution in Activation Data.** In Figure 4 we computed the null  $\text{cnr}$  density function  $(1/d)f(\text{cnr}/d)$  using real motor activation data of a typical subject and obtained a dilation parameter

$d = 1.26$ , indicating that the null distribution of the  $\text{cnr}$  obtained from resampled resting-state data is slightly inflated in activation data. The overall fit of the density functions  $(1/d)f(\text{cnr}/d)$  and  $G_{\mu, \sigma}(\text{cnr})$  is good leading to a small residual mean squared error =  $0.014 \pm 0.114$  (compare the light blue curve and the dark blue curve in Figure 4). A very similar curve was obtained for the memory paradigm using data from a different subject. Here the dilation parameter was found to be  $d = 1.27$  and mean squared error =  $0.011 \pm 0.105$  (Figure 5). Overall, the density fits obtained were similar.

**4.4. Correcting the Smoothing Artifact in Motor and Memory Data.** In Figures 6 and 7 we show examples of the severity of the smoothing artifact for activation data thresholded at the  $P < 0.05$  level, corrected for multiple comparison (i.e., FWE  $< 0.05$ ). The number of voxels affected by the smoothing artifact can be considerable for single voxel with Gaussian

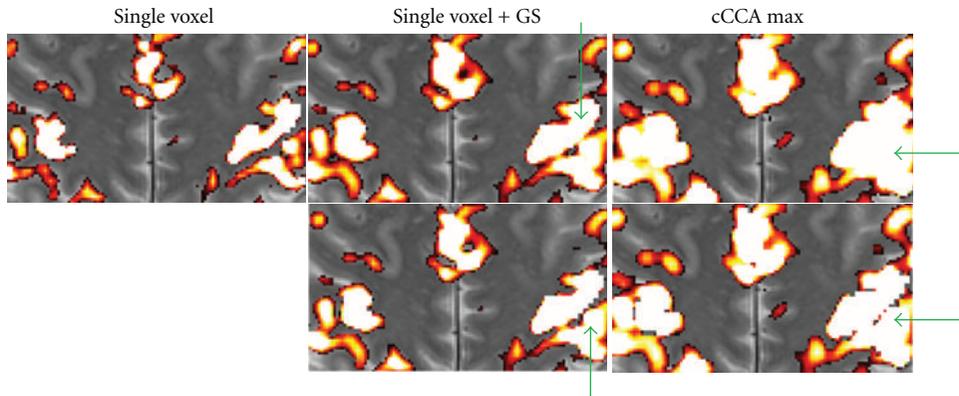


FIGURE 8: Magnified motor activation maps from Figure 6 for selected analysis methods (single voxel without Gaussian smoothing, single voxel with Gaussian smoothing (GS), and CCA with the maximum constraint (cCCA max)). The top row shows the original images at FWE  $< 0.05$  without any correction for the smoothing artifact. The bottom row shows the original images corrected for the smoothing artifact. The green arrows point to major differences of the activation patterns. Note that correction for the smoothing artifact leads to a splitting of activation pattern in the left (radiological convention) motor cortex (see bottom row images with green arrows pointing to the ROI).

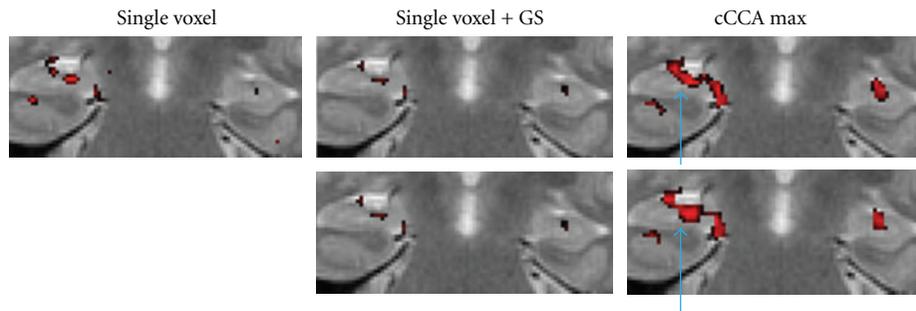


FIGURE 9: Magnified memory activation maps from Figure 7 for selected analysis methods (single voxel without Gaussian smoothing, single voxel with Gaussian smoothing (GS), and CCA with the maximum constraint (cCCA max)). The top row shows the original images at FWE  $< 0.05$  without any correction for the smoothing artifact. The bottom row shows the original images corrected for the smoothing artifact. The blue arrows point to major differences of the activation patterns. Note that hippocampal activation after correction for the smoothing artifact appears to have three separate foci in the right temporal lobe (radiological convention) as shown in the bottom image on the right (blue arrow).

smoothing and cCCA with the average constraint as well as cCCA with the max constraint, as seen in motor data (Figure 6). For cCCA with the sum constraint, however, there is no correction for the smoothing artifact necessary because the sum constraint produces a sufficiently dominant weight for the center voxel so that inactive voxels cannot obtain a dominant weight in the neighborhood of active voxels. We did not find any voxel with a smoothing artifact  $> 0.1$  confirming that cCCA with the sum constraint has largest specificity of the proposed cCCA methods. The activation patterns that are corrected for the smoothing artifact show small changes compared to the uncorrected ones, however, these changes can provide important information of the activation profile. For example, in Figures 8 and 9 we show a magnified region of the left motor cortex and the right hippocampus, respectively, for selected analysis methods (single voxel with and without Gaussian smoothing, cCCA with the maximum constraint). Here we see that correction for the smoothing artifact leads to a separation of the right motor cortex (see green arrows in Figure 8). This result

is consistent with the activation pattern from single voxel analysis without Gaussian smoothing. We believe that for the motor activation data, single-voxel analysis is already accurate due to the high cnr of the BOLD response for motor activation. Regarding the hippocampal activation, we see that the correction for the smoothing artifact leads to a clear separation of hippocampal activation into three focal regions (see blue arrow in Figure 9). It is conceivable that the corrected activation maps are more accurate representations of true hippocampal activations in this high-resolution study because it is known that the hippocampus is composed of the CA fields (CA1, CA2, CA3, and CA4), the dentate gyrus and subiculum, and each of these subregions has a specific function in memory. The obtained corrections of the activation pattern are more probable than a continuous elongated activation pattern obtained with cCCA without correction for the smoothing artifact.

We chose to correct the smoothing artifact when  $P > 0.5$ . This condition is still a conservative correction for activation maps. To obtain better specificity but at a cost of losing

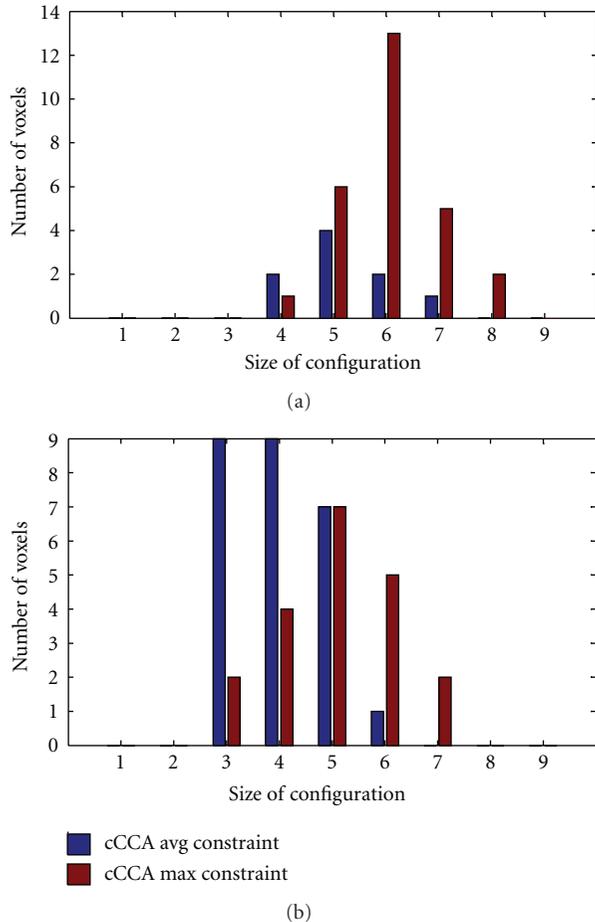


FIGURE 10: Number of voxels affected by the smoothing artifact for two different constrained CCA methods (CCA with the average constraint (cCCA avg) and CCA with the maximum constraint (cCCA max)) as a function of the neighborhood size of the configuration. (a) is for motor data, (b) for memory data (obtained from Figures 6 and 7, resp.). Note that CCA with the max constraint leads to configurations of larger neighborhoods that need to be corrected for the smoothing artifact compared to CCA with the average constraint. The result for CCA with the sum constraint is not shown because the smoothing artifact associated with this constraint is negligible (compare Tables 1 and 2).

TABLE 1: Number of voxels affected by the smoothing artifact for different data analysis methods as a function of  $P >$  threshold in motor activation data for contrast “motor” minus “fixation”.

| Threshold | SV + GS | cCCA sum | cCCA avg | cCCA max |
|-----------|---------|----------|----------|----------|
| 0.5       | 34      | 0        | 9        | 27       |
| 0.4       | 38      | 0        | 14       | 40       |
| 0.3       | 40      | 0        | 24       | 51       |
| 0.2       | 42      | 0        | 48       | 104      |
| 0.1       | 43      | 0        | 98       | 174      |

sensitivity, it may be worthwhile to lower the threshold. Tables 1 and 2 show the number of voxels affected by the smoothing artifact for thresholds 0.1 to 0.5. Note that lowering the threshold for  $P$  to 0.2 leads to a dramatic

TABLE 2: Number of voxels affected by the smoothing artifact for different data analysis methods as a function of  $P >$  threshold in memory activation data for contrast “encoding” minus “control”.

| Threshold | SV + GS | cCCA sum | cCCA avg | cCCA max |
|-----------|---------|----------|----------|----------|
| 0.5       | 5       | 0        | 26       | 20       |
| 0.4       | 5       | 0        | 32       | 20       |
| 0.3       | 6       | 0        | 37       | 24       |
| 0.2       | 7       | 0        | 45       | 33       |
| 0.1       | 10      | 0        | 55       | 44       |

increase in the number of voxels. Thus,  $P > 0.2$  should be avoided. The choice  $P > 0.3$  is probably a good compromise of achieving better specificity and still maintaining high sensitivity for the examples shown here. However, the decision to use a lower threshold than 0.5 will primarily dependent on the particular application of the research. We preferred  $P > 0.5$  which lead to a relatively small number of voxels that needed to be corrected. With this threshold the sensitivity of the methods is still very large and mostly voxel configurations of sizes 4 to 8 in motor data and 3 to 7 in memory data were affected by the smoothing artifact (Figure 10). Note that cCCA with the max constraint leads to larger configuration sizes (mean value  $s = 5.6$ ) that are affected by the smoothing artifact than cCCA with the average constraint (mean value  $s = 4.3$ ). This fact is expected due to the increased freedom of the spatial constraints in cCCA with the maximum constraint leading, on average, to larger configuration sizes which are more probable to induce a smoothing artifact than the other constrained cCCA methods.

## 5. Conclusions

We summarize the ideas introduced in this study and results obtained as follows.

- (1) We investigated the smoothing artifact in CCA and proposed a new technique to reduce this artifact in fMRI data analysis.
- (2) Using data from a motor activation paradigm and an episodic memory paradigm, we showed examples of activation maps obtained with constrained CCA methods, the corresponding magnitude of the smoothing artifact, and activation maps corrected for the smoothing artifact.
- (3) For all data studied, we found no appreciable smoothing artifact for cCCA with the sum constraint.
- (4) The best overall performance was obtained by cCCA with the maximum constraint corrected for the smoothing artifact. We recommend this technique for fMRI data analysis to obtain high sensitivity and good specificity.

## Acknowledgment

This research is supported by the NIH/NIA (Grant no. 1R21AG026635).

## References

- [1] O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using canonical correlation analysis," *Magnetic Resonance in Medicine*, vol. 45, pp. 323–330, 2001.
- [2] R. R. Nandy and D. Cordes, "Novel nonparametric approach to canonical correlation analysis with applications to low CNR functional MRI data," *Magnetic Resonance in Medicine*, vol. 50, no. 2, pp. 354–365, 2003.
- [3] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Detection of neural activity in fMRI using maximum correlation modeling," *NeuroImage*, vol. 15, no. 2, pp. 386–395, 2002.
- [4] R. Nandy and D. Cordes, "Improving the spatial specificity of canonical correlation analysis in fMRI," *Magnetic Resonance in Medicine*, vol. 52, no. 4, pp. 947–952, 2004.
- [5] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Adaptive analysis of fMRI data," *NeuroImage*, vol. 19, no. 3, pp. 837–845, 2003.
- [6] M. Ragnehed, M. Engström, H. Knutsson, B. Söderfeldt, and P. Lundberg, "Restricted canonical correlation analysis in functional MRI-validation and a novel thresholding technique," *Journal of Magnetic Resonance Imaging*, vol. 29, no. 1, pp. 146–154, 2009.
- [7] M. Jin, R. Nandy, T. Curran, and D. Cordes, "Extending local canonical correlation analysis to handle general linear contrasts for fMRI data," *International Journal of Biomedical Imaging*, vol. 2012, Article ID 574971, 14 pages, 2012.
- [8] D. Cordes, M. Jin, T. Curran, and R. Nandy, "Optimizing the performance of local canonical correlation analysis in fmri using spatial constraints," *Human Brain Mapping*, vol. 33, no. 11, pp. 2611–2626, 2012.
- [9] D. Cordes, M. Jin, T. Curran, and R. Nandy, "The bleeding artifact of spatially constrained canonical correlation analysis in functional MRI," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM '11)*, p. 1613, Montreal, Canada, 2011.
- [10] B. W. Silvermann, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, NY, USA, 1986.
- [11] M. Breakspear, M. J. Brammer, E. T. Bullmore, P. Das, and L. M. Williams, "Spatiotemporal wavelet resampling for functional neuroimaging data," *Human Brain Mapping*, vol. 23, no. 1, pp. 1–25, 2004.
- [12] E. Bullmore, C. Long, J. Suckling et al., "Colored noise and computational inference in neurophysiological (fMRI) time series analysis: re-sampling methods in time and wavelet domains," *Human Brain Mapping*, vol. 12, pp. 61–78, 2001.
- [13] D. Cordes and R. Nandy, "Independent component analysis in the presence of noise in fMRI," *Magnetic Resonance Imaging*, vol. 25, no. 9, pp. 1237–1248, 2007.
- [14] S. Smith, M. Jenkinson, C. Beckmann, K. Miller, and M. Woolrich, "Meaningful design and contrast estimability in FMRI," *NeuroImage*, vol. 34, no. 1, pp. 127–136, 2007.
- [15] R. S. J. Frackowiak, Ed., *Human Brain Function*, Elsevier Science, San Diego, Calif, USA, 2nd edition, 2004.
- [16] R. Nandy and D. Cordes, "A semi-parametric approach to estimate the family-wise error rate in fMRI using resting-state data," *NeuroImage*, vol. 34, no. 4, pp. 1562–1576, 2007.
- [17] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society, Series B*, vol. 53, no. 3, pp. 683–690, 1991.

## Research Article

# A New GLLD Operator for Mass Detection in Digital Mammograms

**N. Gargouri,<sup>1</sup> A. Dammak Masmoudi,<sup>1</sup> D. Sellami Masmoudi,<sup>1</sup> and R. Abid<sup>2</sup>**

<sup>1</sup> Computer Imaging and Electronic System Group, CEM Laboratory, Department of Electrical Engineering, Sfax Engineering School, University of Sfax, P.O. Box 1169, 3038 Sfax, Tunisia

<sup>2</sup> El Farabi Radiology Center, 14 Janvier Avenue, 3000 Sfax, Tunisia

Correspondence should be addressed to N. Gargouri, norhene.gargouri@live.fr

Received 19 July 2012; Revised 12 November 2012; Accepted 21 November 2012

Academic Editor: Juan Ruiz-Alzola

Copyright © 2012 N. Gargouri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During the last decade, several works have dealt with computer automatic diagnosis (CAD) of masses in digital mammograms. Generally, the main difficulty remains the detection of masses. This work proposes an efficient methodology for mass detection based on a new local feature extraction. Local binary pattern (LBP) operator and its variants proposed by Ojala are a powerful tool for textures classification. However, it has been proved that such operators are not able to model at their own texture masses. We propose in this paper a new local pattern model named gray level and local difference (GLLD) where we take into consideration absolute gray level values as well as local difference as local binary features. Artificial neural networks (ANNs), support vector machine (SVM), and k-nearest neighbors (kNNs) are, then, used for classifying masses from nonmasses, illustrating better performance of ANN classifier. We have used 1000 regions of interest (ROIs) obtained from the Digital Database for Screening Mammography (DDSM). The area under the curve of the corresponding approach has been found to be  $A_z = 0.95$  for the mass detection step. A comparative study with previous approaches proves that our approach offers the best performances.

## 1. Introduction

Breast cancer is the major public health problem in the world. It constitutes the most common cancer among the female population [1]. A study developed by American cancer society estimates that between one in eight and one in twelve will be diagnosed with breast cancer in their life time [2]. The European community estimates that breast cancer corresponds to 19% of cancer death. Moreover, it represents 24% of cancer cases [3]. In a Tunisian country, breast cancer is 16,5% of cancer death [4]. Mostly, 25% of all cases of breast cancer deaths occur if women were diagnosed between the age of 40 and 49. Although breast cancer incidence has increased over the last decade, breast cancer mortality has declined among women of all ages [5], thanks to the development of both breast cancer treatment and mammography screening.

Among the different imaging modalities used for the detection of breast cancer, mammography remains the most used one to reveal breast abnormalities. Vacek et al. [14]

demonstrate that the ratio of breast tumor detection in Vermont (USA), when applying screening mammography, increased from 2% to 36% between 1995 and 1999.

Nowadays, the digital mammography gives the opportunity of increasing the use of the CAD systems in order to help the expert radiologists in the interpretation and diagnosis of mammograms [15].

However, the rapid improvement of full digital mammography has been accompanied by natural increase of such systems. The CAD is a set of tools developed to help radiologists in the detection and interpretation of mammographic images [16].

Back in 2001, Freer and Ulissey [16] have proposed an algorithm using substantial dataset containing 12,860 cases and have concluded that the application of CAD in the analysis of screening mammograms may increase the malignancies detection at an early stage. The main disadvantage of existent CAD systems is the lack of general algorithms producing good results for all cases and images. Masses and microcalcifications are common lesions found in

TABLE 1: Previously developed approaches on mass detection based on feature extraction and on learning. In this table, we specify for each approach the feature extraction technique, the classifier, the ratio which indicates the number of real masses/number of normal ROIs, and the obtained results. In the feature extraction methods, ICA, PCA, and 2DPCA correspond, respectively, to independent component analysis, principal component analysis and two-dimensional PCA. In the classification stage, ANN, NN, and SVM correspond, respectively, to the artificial neural network, nearest neighbors, and support vector machines. Generally, the evaluation of the works is given in terms of  $A_z$  where  $A_z$  represents the area under the ROC curve, except for both works of Christoyianni et al. and Leonardo et al. giving the correct classification true positive and true negative in percentage.

| Author                   | Year | Classifier based |            |       |            |     |     |       | Classifier | Ratios   | Results      |
|--------------------------|------|------------------|------------|-------|------------|-----|-----|-------|------------|----------|--------------|
|                          |      | Texture          | Morphology | Shape | Gray level | ICA | PCA | 2DPCA |            |          |              |
| Qian et al. [6]          | 2001 | ✓                |            | ✓     |            |     |     |       | ANN        | 200/600  | $A_z = 0.86$ |
| Christoyianni et al. [7] | 2002 | ✓                |            |       | ✓          |     | ✓   |       | ANN        | 119/119  | 88.23%       |
| Oliver et al. [8]        | 2006 |                  |            |       |            |     | ✓   |       | C4.5 + NN  | 196/392  | $A_z = 0.83$ |
| Oliver et al. [9]        | 2007 |                  |            |       |            |     |     | ✓     | NN         | 256/1536 | $A_z = 0.86$ |
| Varela et al. [10]       | 2007 |                  | ✓          |       | ✓          |     |     |       | ANN        | 60/60    | $A_z = 0.90$ |
| Leonardo et al. [11]     | 2009 | ✓                |            | ✓     |            |     |     |       | SVM        | 250/1177 | 92.63%       |

mammographic images. We will focus in this paper on mass-related lesions. In order to develop an improved computer-aided clinical decision classifying the tumor and identifying the stage of the cancer, we must ensure whether it is an area which contains a mass or not. The mass detection is therefore a valuable step in diagnosis. Our work focuses on classification of the tissue of the breast as mass or mass free. So, mass detection system is able to assist health professionals in finding out mass abnormalities in mammograms.

Several algorithms are typically based on only one view. However, some recent approaches have used multiple views [17], but this has three main drawbacks.

- (i) Image views must be properly registered to allow a correct comparison of regions.
- (ii) Gray level values must also be correctly registered.
- (iii) There are some cases where comparison is not possible, because no correspondance between pixels can be done (e.g., the case of patients which have suffered from a previous breast surgery).

It is important to note that algorithms typically working with one view can always be applied to multiple views. Textural information has already been used to solve this problem and has been introduced in several works [6, 10]. Oliver et al. have proposed an algorithm for mammographic mass detection based on LBP [18]. Results have indicated that the use of LBP and its extensions has been promising in different comparative studies and has been applied in different texture analysis tasks [6, 19]. However, LBP descriptors are not able to model mammogram texture because they are mapping only the differences of pixel gray level values. So, we will work here on a new approach taking into consideration the texture whole information, the local difference and local gray values as features, namely, gray level and local difference features (GLLD). Accordingly, we intend to investigate the efficiency of the GLLD based approach as a method of feature extraction. We perform our experiments on a set of 1000 ROIs obtained from the DDSM database.

This paper, using a single view, proposes a new CAD methodology in order to achieve better performances in

terms of false negative and false positive using GLLD operator. The remaining of this paper is organized as follows. Section 2 shows some related works on mass detection in mammogram images. In Section 3, we present a brief review of LBP operator and an analysis of our GLLD based approach. In Section 4, a brief description of a set of classification methods is given, namely, support vector machine (SVM), k-nearest neighbors (kNNs), and artificial neural network. Section 5 is reserved to the validation of the GLLD proposed technique simulation, and results and discussion are conducted. In the last section, we summarize the paper contribution and end our work by some concluding remarks and future work.

## 2. Background

Several image processing techniques have been formulated as tools that can assist early automatic mass detection [20, 21]. Algorithms for mammographic mass detection using a single image view are based on a characteristic classifier scheme: for a given database consisting of known cases, the decision making system learns how to distinguish between the two kinds of ROIs (mass and nonmass ROI). Thereafter, once the given system has been trained, a new ROI can be rightly classified. Among all these detection algorithms, we can differentiate between two strategies. The first one includes the algorithms which extract features usually related to their texture from the ROI and then trains a classifier. Approches lying in such strategy are summarized in Table 1. The second strategy turns this problem into a template matching one. Each new ROI is compared to all the remaining ROI images obtained from the database in order to be finally classified as mass or nonmass. Table 2 shows different approaches applying this strategy.

On the one hand, Qian et al. [6] have analyzed the implementation of an adaptive CAD to develop a fully automatic procedure for mass segmentation and classification which consists in training a novel Kalman-filtering neural network to classify features extracted from wavelet decomposition [6]. On the other hand, Christoyianni et al. [7] have extracted features based on independent component analysis (ICA),

TABLE 2: Previously developed approaches on mass detection based on template matching. In this table, the ratio which indicates the number of real masses/number of normal ROIs.

| Author               | Year | Gray level | Template-matching based |         |                     | Similarity | Ratios       | Results |
|----------------------|------|------------|-------------------------|---------|---------------------|------------|--------------|---------|
|                      |      |            | Shape                   | Entropy |                     |            |              |         |
| Chang et al. [12]    | 2001 |            | ✓                       | ✓       | Likelihood function | 300/300    | $A_z = 0.83$ |         |
| Tourassi et al. [13] | 2007 | ✓          |                         | ✓       | Mutual function     | 901/919    | $A_z = 0.81$ |         |

gray level and texture, in order to train the ANN as a classifier. Furthermore, they have applied the principal component analysis (PCA) for the preprocessing step to overcome the problem of complexity and of increasing dimensionality. Oliver et al. [8] have proposed a different strategy. The latter is based on the translation of eigenfaces approach for face detection/classification problems to the mass detection one. They introduced the concept of spanning the ROI subspace of an original image space. As result of such transformation, they have obtained a vector which describes the contribution of each eigenrau for the representation of the corresponding image. They have used these vectors in the construction of the models for the step of training. In [9], Oliver et al. have extended their proposed method based on PCA by using the two-dimensional PCA (2DPCA) technique. Varela et al. [10] have proposed a methodology based on extracting gray level as well as morphological features and classifying, using ANN, the new ROI. Leonardo et al. [11] have proposed an algorithm for the detection of masses in mammographic images. The technique is based on the use of textural and shape measures for K-means clustering algorithm and the SVM, aiming at detecting masses in mammographic images.

As shown in Table 2, the proposed approach of Chang et al. [12] and Tourassi et al. [13] has been based on a template matching-based approach. They have proposed for classification purposes to undertake a comparison of the new ROI with the remaining ROIs in the database composed of ROIs depicting masses. The difference between these works appears in the similarity measure function. As indicated in Table 2, there are only limited publications trying to detect masses using template-matching based methods. From the two tables, we can conclude that one of the main dissimilarity among these recent works is the ratio between the ROIs depicting abnormality and the total number of cropped images. It is important to note that when the number of normal ROIs increases, the number of ROIs wrongly classified is likely to increase. One should remember that the purpose of this work is the classification of mammographic masses and normal breast tissue. All the developed methods allow the tradeoff between the reduction of false positive fraction and the increase of false negative fraction. Such trade off can be ensured when using the receiver operating characteristics (ROC) [22] in the performance evaluation step. This is the case of most of the approaches in Tables 1 and 2. The ROC curve is a graphical curve representing the true positive rate (sensitivity) versus the false positive rate (100 specificity), extensively used in classifier performance evaluation. Points representing ROC curve correspond to sensitivity/specificity pair representing a particular decision threshold. The AUC (known as  $A_z$ ) is an information about

the overall performance of the approach. Furthermore, the latter is a metric which can be used to compare different features, and it allows the reduction of the ROC curve to a single value summarizing expected performance. A reasonable test should have

$$0.5^\circ \leq \text{AUC}^\circ \leq 1. \quad (1)$$

Most of the approaches in the first strategy have the drawback that a large number of features need to be calculated but only the most discriminant will be selected [7, 8]. Besides, for the second strategy, the used similarity function measure for classifying needs to be recomputed for each element. In our paper, to overcome such limitations, the LBP operator has been investigated with the idea of performing gray scale invariant texture analysis. The latter has proved to be relevant in many applications. However, it shows some limitations when applied to mammographic image. For instance, it gives the same results with two different absolute gray levels. Knowing that the gray level information is of great importance in mammography, our approach will add to LBP absolute gray level information rather than gray level difference. We will focus in our approach on making use of small size feature vector as well as possible. In the following section, we will introduce our mass detection methodology based on the GLLD for the extraction of texture features obtained from the ROIs.

### 3. Local LBP Approach and Improvements

Texture classification is nowadays a challenging problem. It is an active topic in computer vision research. Early methods of texture classification are based on statistical analysis of images with different textures. The most representative ones are the cooccurrence matrix method [23] and filtering for texture-classification methods [24]. At an early stage, exploratory models were developed to investigate rotation invariance in texture classification, such as hidden Markov model [25] and Gaussian Markov random field [26]. Varma and Zisserman [27] have proposed to learn from a training set a rotation invariant texton and to classify the obtained texture according to its texton distribution. Varma and Zisserman [28] have later proposed to use the image patch in order to represent features directly. Some recently proposed works have been developed for scale as well as affine invariant texture classification. Later, Ojala et al. [29] have proposed the LBP histogram application in order to achieve a rotation invariant texture classification. It is worth noting that the LBP is efficient in describing local image pattern and its performance in computer vision

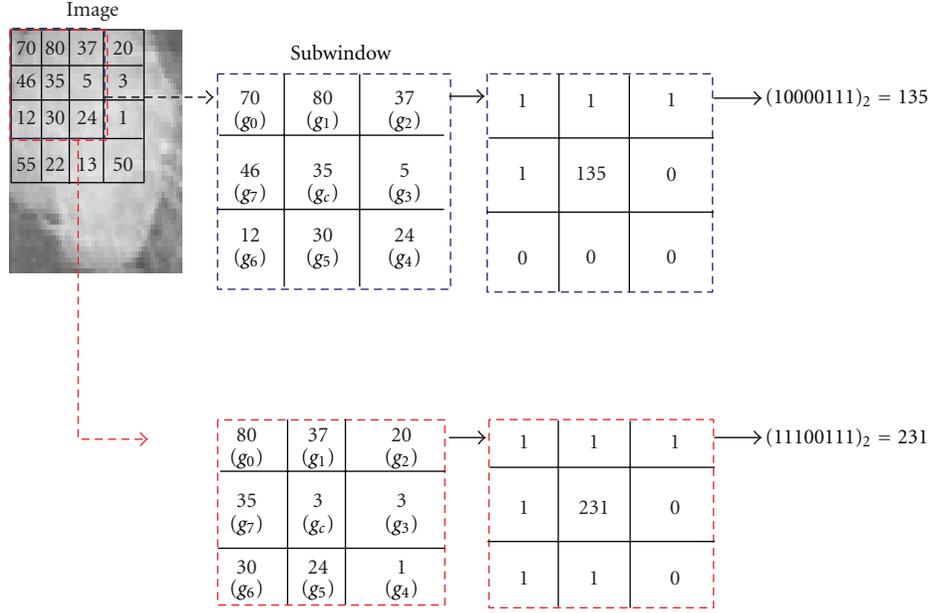


FIGURE 1: Example of basic LBP operator.

and pattern recognition is promising. However, it still needs to be improved for mammography texture modeling. In order to generate texton, Ojala et al. [30] have applied the Absolute Gray Level Difference (AGLD) between each pixel and its neighbors. After that, the obtained histogram has been used to represent the image texture. Then, Ojala et al. [29] have proposed the LBP using the sign of the difference for the representation of local patterns. In [31], Ojala et al. suggested to use signed gray level Difference (SGLD) and its multidimensional distribution for the description of texture and considered LBP as a simplification of SGLD. With such variants of LBP, there still remain questions that need answering, such as what information is lost in the considered code? How to represent the missing information to obtain better texture modeling? Here, we propose a new feature extractor to improve the system performance, based on GLLD features.

**3.1. Brief Review of LBP Formulation.** The LBP operator used eight neighboring pixels when considering the center gray value as threshold. This operator generates “1” if the considered neighbor value is greater or equal to that of the center. Otherwise, it generates “0.”

Accordingly and referring to Figure 1, LBP [29] code may be computed as follows:

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \quad (2)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

where  $g_c$  corresponds to the gray value of the central pixel,  $g_p$  corresponds to the value of its neighbors, and

( $p = 0, 1, \dots, P - 1$ ) and  $P, R$  correspond to the number of neighbors and to the radius of the neighborhood, respectively. The binary code is then represented with an 8-bit number.  $g_p$  coordinates are  $((R \cos(2\pi p/P)$  and  $R \sin(2\pi p/P))$ . If neighbors are not in the image grids, their gray values may be estimated by interpolation. After identifying LBP pattern of each pixel  $(i, j)$ , we associate LBP histogram to the whole image, with a given image size  $(N1 * N2)$  as

$$H_{\text{LBP}}(k) = \sum_{i=1}^{N1} \sum_{j=1}^{N2} f(\text{LBP}_{P,R}(i, j), k), \quad k \in [0, K], \quad (3)$$

$$f(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise,} \end{cases}$$

where  $K$  corresponds to the maximum gray level value.

Let  $U$  be a function corresponding to the value of an LBP pattern, it is defined as the number of transitions (i.e., change from 0 to 1 or 1 to 0) in the following pattern:

$$U(\text{LBP}_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|. \quad (4)$$

Patterns corresponding to limited transitions or discontinuities are with  $U \leq 2$ , in a binary presentation. Otherwise, these patterns are noted as uniform LBP patterns [29]. The mapping from the original  $\text{LBP}_{P,R}$  to  $\text{LBP}_{P,R}^{u2}$ , knowing that the superscript  $u2$  refers to uniform patterns, may be implemented using a look-up table containing  $2^P$  elements.

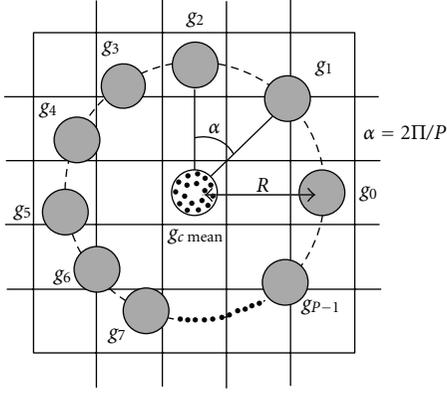


FIGURE 2: The central pixel  $g_c$  and its  $P$  circularly symmetric neighbor with radius  $R$ .

A local rotation invariant pattern is defined as follows [32]:

$$\text{LBP}_{P,R}^{\text{riu2}} = \begin{cases} \sum_{p=0}^{P-1} |s(g_p - g_c)| & \text{if } U(\text{LBP}_{P,R}) \leq 2 \\ P + 1 & \text{otherwise.} \end{cases} \quad (5)$$

The mapping from  $\text{LBP}_{P,R}$  to  $\text{LBP}_{P,R}^{\text{riu2}}$ , knowing that the superscript *riu2* corresponds to rotation invariant uniform patterns, may be implemented using a look up table.

**3.2. GLLD Feature-Based Approach.** The main limitation using LBP code is that it may give the same results with two completely different gray levels when the differences with the neighbors are the same.

Knowing that for mammographic images, the gray level information is directly related to the breast tissue density, gray level and local difference are two important features of the texture which must be used together in order to have more accurate results.

In our approach, we propose to calculate the average for each  $3 \times 3$  neighborhood and to attribute it to the central pixel. The new value of the central pixel is noted as  $g_{c \text{ mean}}$ .

Given the new value of the central pixel  $g_{c \text{ mean}}$  and its  $P$  circularly symmetric neighbor (see Figure 2), the subtraction of the value of  $g_{c \text{ mean}}$  is presented as follows:

$$T = t(g_{c \text{ mean}}, g_0 - g_{c \text{ mean}}, \dots, g_{P-1} - g_{c \text{ mean}}). \quad (6)$$

Thus, the difference between  $g_{c \text{ mean}}$  and  $g_p$  may be represented as  $\text{diff}_p = g_p - g_{c \text{ mean}}$  and the local difference may be represented with a vector noted  $\text{diff}_p$  knowing that  $\text{diff}_p = [\text{diff}_0, \dots, \text{diff}_{P-1}]$ ,  $\text{diff}_p$  describes the local image structure around the  $g_{c \text{ mean}}$ . Because of its robustness and efficiency, the obtained vector  $\text{diff}_p$  is decomposed of sign and modulus components in order to achieve much better performance in texture classification. In our proposal,  $s_p$  corresponds to the sign of the differences, and it is obtained by thresholding with respect to the value of  $g_{c \text{ mean}}$  as expressed in (8). However,  $m_p$  corresponds to the absolute value of  $\text{diff}_p$  as expressed in (9).

We obtain, also, two vectors, the sign vector  $[s_0, \dots, s_{P-1}]$  and the modulus vector  $[m_0, \dots, m_{P-1}]$ , with

$$\text{diff}_p = m_p \cdot s_p, \quad (7)$$

knowing that

$$s_p = \begin{cases} 1, & \text{diff}_p \geq 0 \\ -1, & \text{diff}_p < 0, \end{cases} \quad (8)$$

$$m_p = |\text{diff}_p|. \quad (9)$$

Figure 3 shows an illustration example of the proposed method. Aiming at recognizing robustly and efficiently the texture patterns, we should extract both absolute and relative features from pixel gray levels.

The modulus component provides discriminant information to the sign component; the intensity value of the central pixel corresponding to the mean value of its neighbors may also give us useful information [33, 34]. It will also be seen that by coding the sign, the modulus, and the central gray level features into rotation invariant binary codes and fusing them, results may provide much better performance in mammogram texture classification than using each one by itself. This fusion provides useful information about local gray level which is so important in the stage of mass detection in mammographic images.

**3.2.1. SGLLD, MGLLD, and CGLLD Operators.** In this subsection, we present the gray level and local difference (GLLD) different processing steps to explore the proposed three features, which are illustrated in Figure 4.

We start by extracting different ROIs from mammographic images. After that, in the selected ROI, each central gray level corresponds to the mean of its neighbors and its local difference. The latter is decomposed into sign and modulus components as expressed in (7). Given a pixel in the image, the sign coding component is noted as (SGLLD) and is computed by comparing it with the values of its neighbors as follows:

$$\text{SGLLD}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_{c \text{ mean}}) 2^p, \quad (10)$$

where  $s(x)$  is defined by

$$s(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0, \end{cases} \quad (11)$$

where  $g_{c \text{ mean}}$  is the average value of the central pixel and its neighbors. Inspired by the method of coding (SGLLD), the coding of the Magnitude component is noted as (MGLLD) and is defined as follows:

$$\text{MGLLD}_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^p, \quad (12)$$

$$t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c, \end{cases} \quad (13)$$

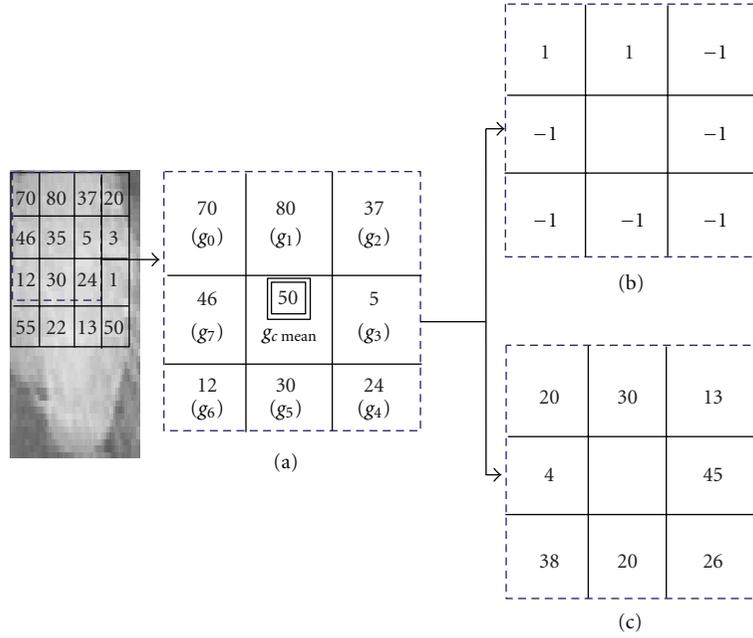


FIGURE 3: (a) A  $3 \times 3$  block with central pixel corresponding to the mean value of its neighbors. (b) The sign components. (c) The magnitude components.

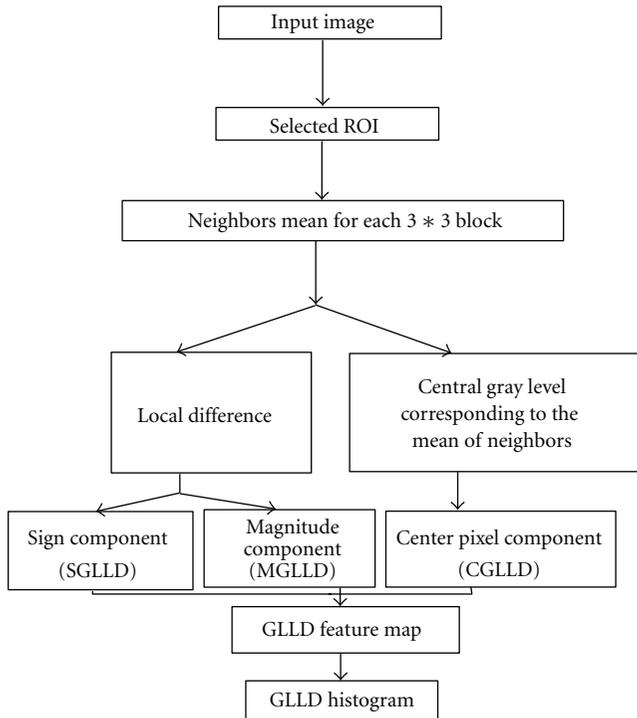


FIGURE 4: Different processing steps of the proposed GLLD based approach.

where  $c$  corresponds to a global gray level threshold which is determined adaptively. We set it as the average value from the whole image.

The new value of the central pixel, which expresses the gray level of the image, represents also a discriminant information. So, to make it consistent with the two previous operators SGLLD and MGLLD, we code it as

$$\text{CGLLD}_{P,R} = t(g_c \text{ mean}, c_I), \quad (14)$$

where  $t$  is already defined in (13),  $c_I$  corresponds to the threshold and is set as the mean gray level of the whole input image. CGLLD is defined to extract the image local gray level. Figure 5 illustrates the image results after the application of the three operators and their fusion.

For the three obtained codes, the rotation invariant version is defined to achieve rotation invariant classification. Each code carries specific texture information, that is why we concatenate them to build the GLLD feature, which corresponds to a vector. So, the three obtained histograms were concatenated to one histogram (cf. Figure 6).

The procedure consists in using the GLLD in order to build local descriptor of the obtained ROIs knowing that the concatenation leads to global description and the obtained global and local GLLD texture descriptor are, then, used as features for mass detection.

The following images (Figure 7) illustrate the obtained histogram for different ROIs. In this figure, we have considered three trivial examples as well as three challenging examples which were misclassified by a radiologist. All of the six ROI examples were correctly classified by our GLLD texture features. In a further section, we will focus on validating our approach statistically on the DDSM database.

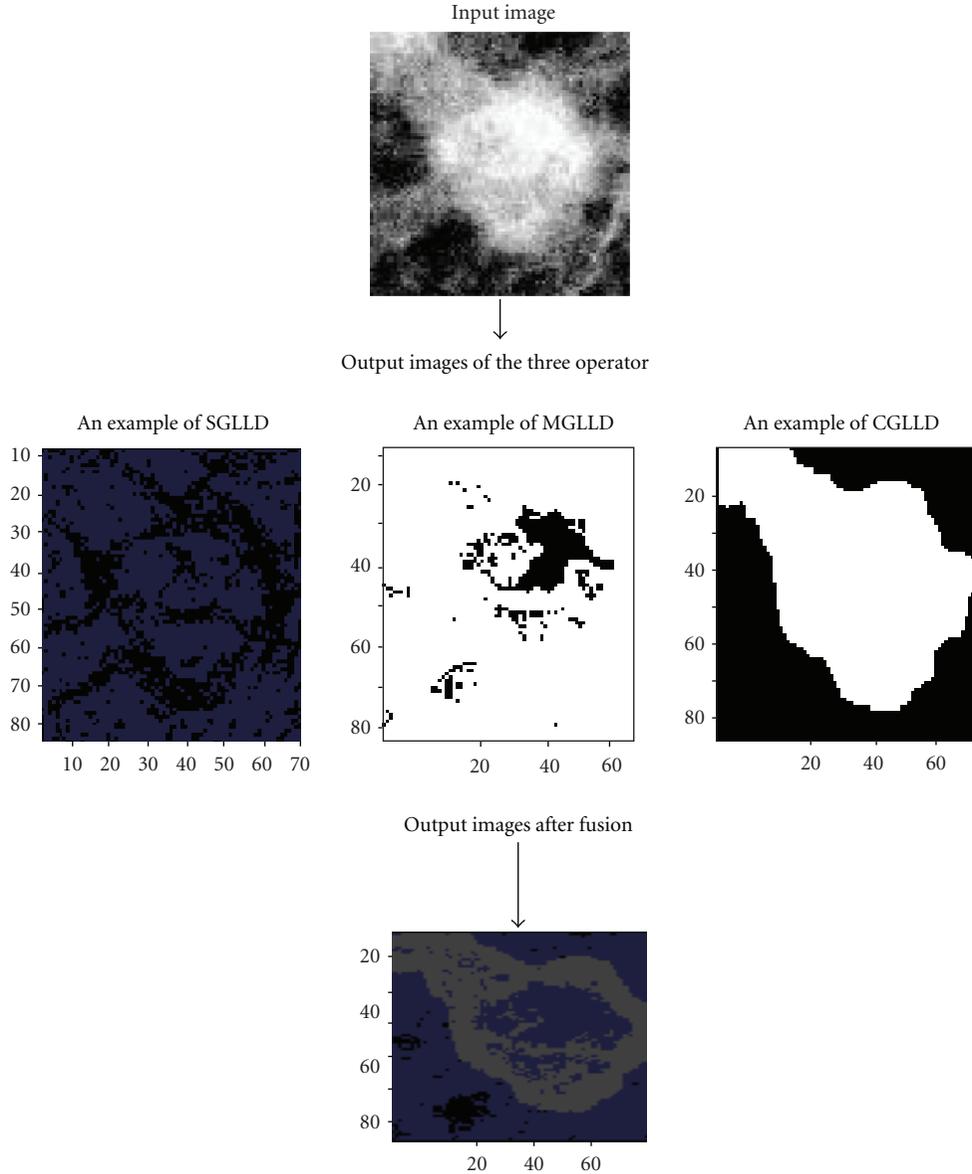


FIGURE 5: The image results after the application of the three operator and their fusion.

## 4. Classification

The last step of our proposal is mass classification. For the sake of generality and for doing a best choice of the classifier, an investigation of three classifiers will be undertaken, namely, support vector machine (SVM), k-nearest neighbors (kNNs) and artificial neural network. The following subsection give a brief review of such classifiers.

**4.1. Support Vector Machine.** The SVM is a largely used classification technique introduced by Vapanick [35]. It learns how to discriminate between positive and negativ(in our case mass and non mass), by finding a hyperplane as a decision surface separating the classes. The hyperplane is defined by support vectors. The SVM uses an optimization method identifying the support vectors  $s_i$ , the weights  $a_i$ , and

the bias  $b$  which are used for the classification of the vectors  $x$  according to the following equation:

$$C(x) = \sum_i a_i \varphi(s_i, x) + b, \quad (15)$$

where  $\varphi$  corresponds to a kernel function.  $\varphi$  refers to a dot product in the case of a linear kernel. Then, if  $c \geq 0$ ,  $x$  is classified as a member belonging to the first class. Otherwise, it is classified as a member belonging to the second class.

**4.2. k-Nearest Neighbors.** kNN classifier is a well-known method in a large number of applications. Since kNN is memory based, no models need to be trained. For a given instance  $x$ , the kNN first finds the  $k$  closest training points with respect to a particular distance metric. Then, it uses

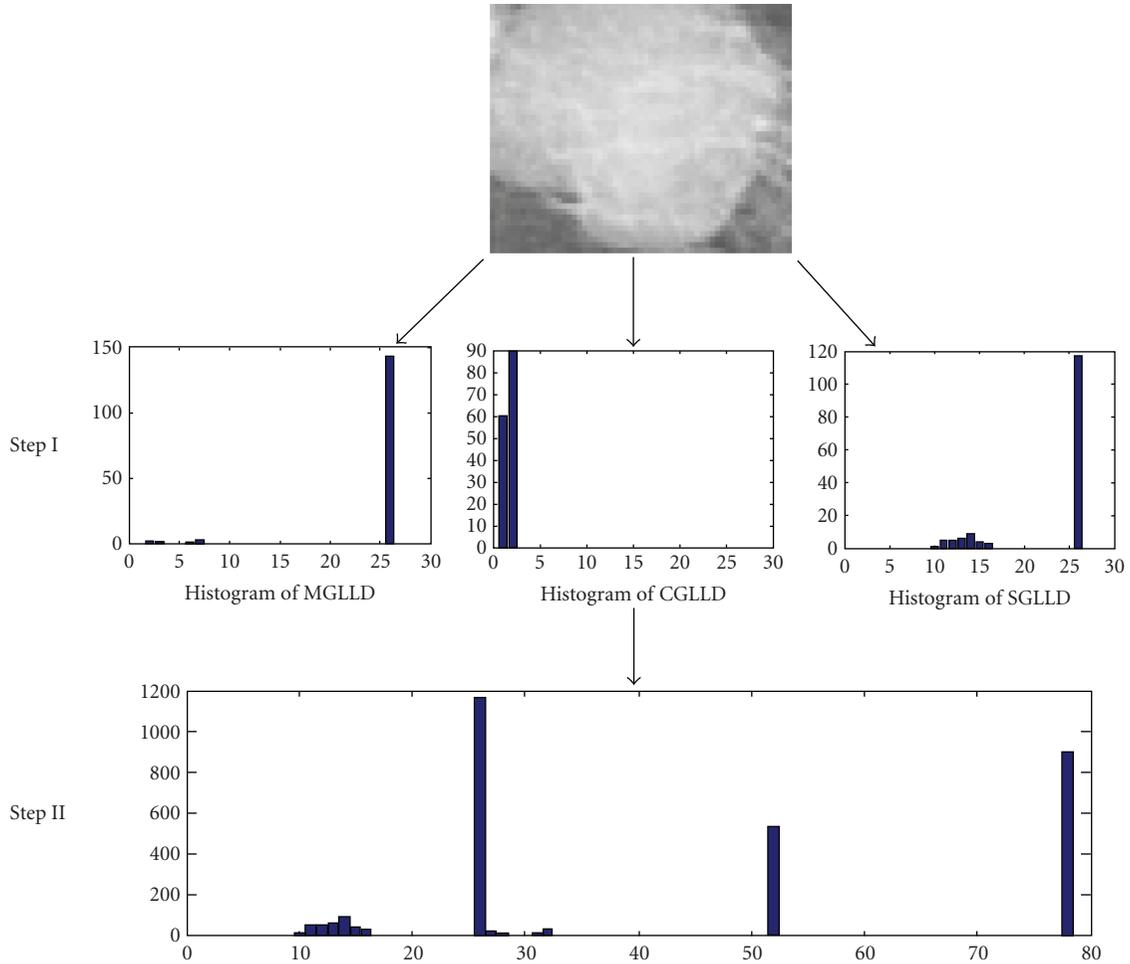


FIGURE 6: The extraction of the different local features from an ROI sample. Step (I): the texture features can be computed by building the histogram over the corresponding ROI. Step (II): the histogram from the three operators is concatenated to build the texture features of the selected ROI.

its labels in order to classify the instance  $x$  by majority vote [36]. In this study, we use the Euclidian distance to determine the nearest neighbors of the query element, and  $k$  is used as a training parameter. For each element  $x$ , the output score corresponds to the ratio of the winning class elements among the total number of neighborhood in the corresponding dataset.

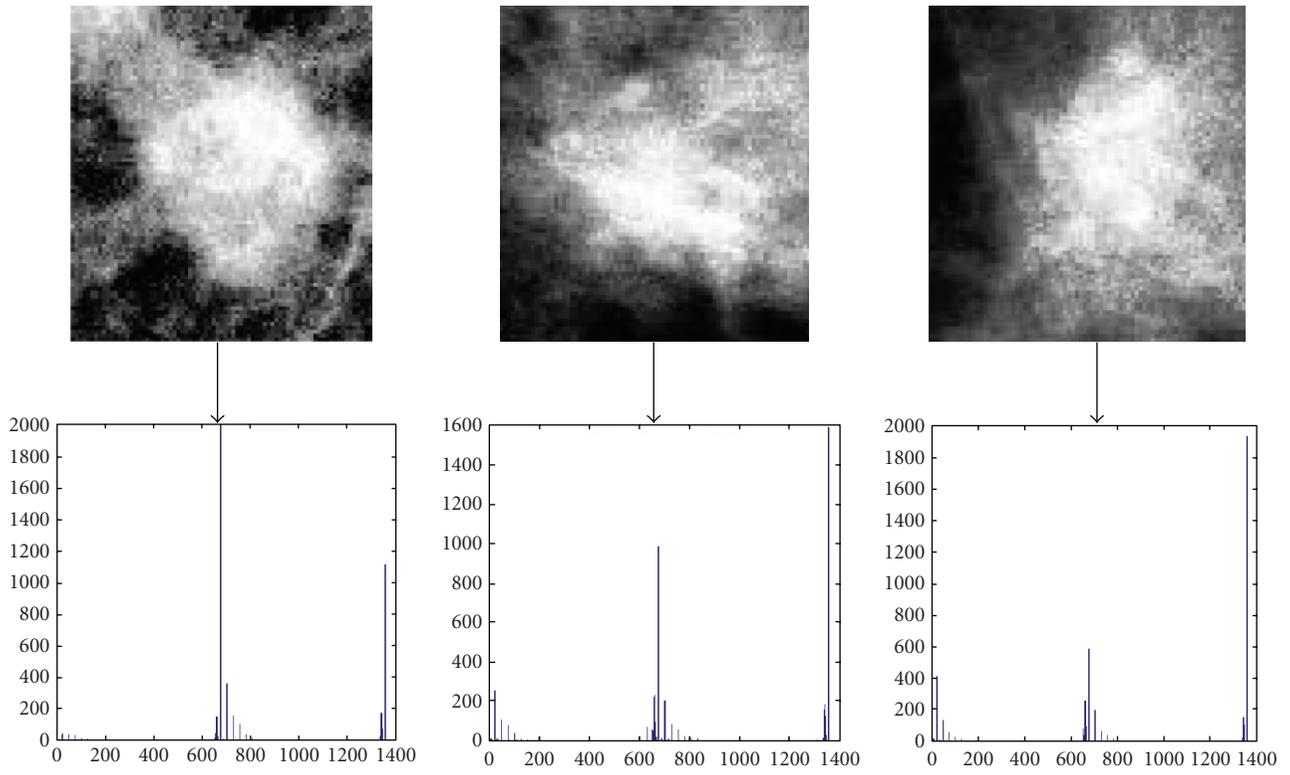
**4.3. Artificial Neural Network.** ANN has been widely used in many applications where the expert knowledge is not clearly defined [37]. The idea of the ANN has been inspired from the biological nervous system and has been successfully applied in medical imaging. This technique is based on the adjustment of weights between the neurons for any input-output function approximation. Therefore ANN, has been widely used in digital mammography to mimic this computational power and the perception capabilities of human brain.

Two basic types of ANN, the multilayer perceptron (MLP) as well as the radial basis function network (RBF), are frequently used in recent works.

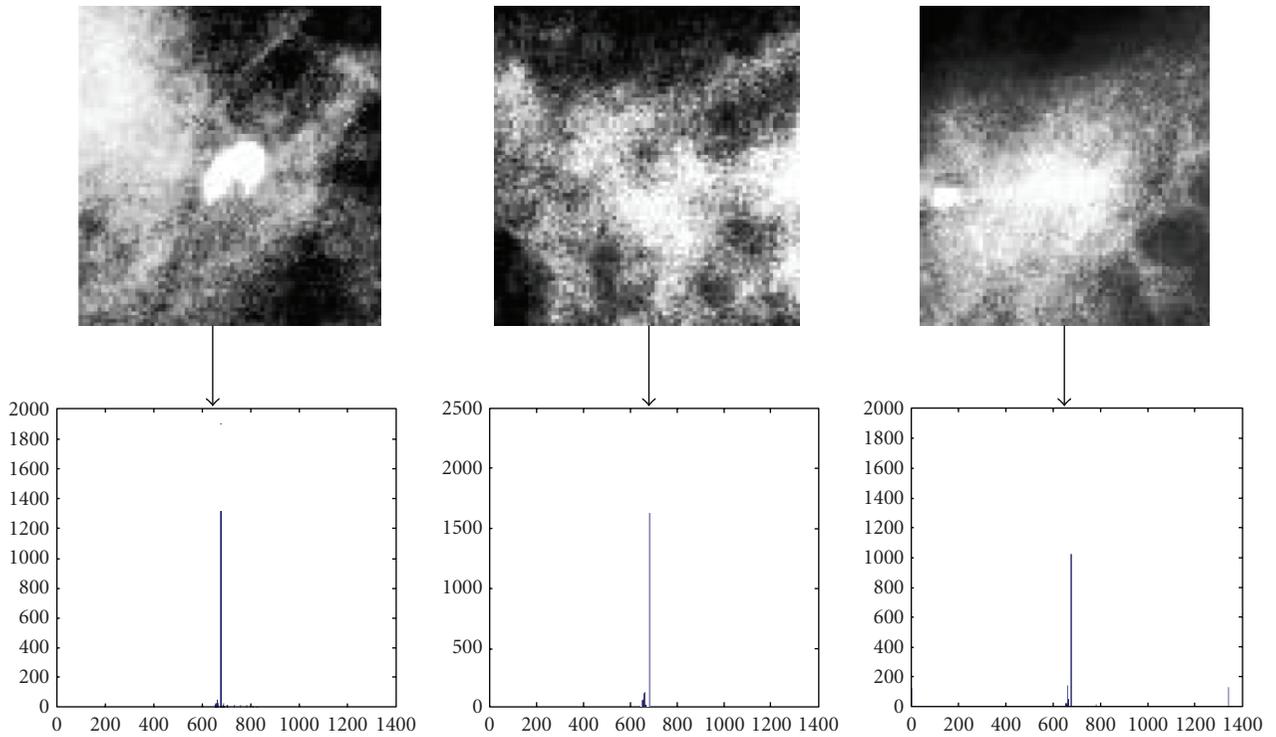
On the one hand, multilayer perceptrons (MLPs) are feedforward ANN models typically trained with static backpropagation. The MLPs find their way into many applications which require static pattern classification. Their principal advantage is the ease of use and the approximation capability of any input/output map.

On the other hand, radial basis function (RBF) networks are nonlinear hybrid networks containing a single hidden layer of processing elements. This layer uses gaussian transfer functions and the sigmoidal functions used by MLPs [38]. This type of ANN is in generally used when the number of samples is so small ( $<100$ ). So, the limitation of the RBF neural network is that it is very sensitive to the dimensionality and has more and more difficulties if the number of units is large.

Based on this assumption, and knowing that the GLLD feature size is of 1352, we intend to investigate the MLP to exploit the results using the ANN [39]. Details of the used MLP network parameter are presented in Table 3. Let us consider  $x = (x_1, x_2, \dots, x_d)^T$  the input vector,  $\omega = (\omega_1, \omega_2, \dots, \omega_d)^T$ , the weight vector, and  $g(x) = (1 + e^{-x})^{-1}$



(a) Three crops of true masses, and their corresponding GLLD histograms



(b) Three crops of texture classified by the radiologist as false positif and their corresponding GLLD histograms

FIGURE 7: GLLD feature distributions extracted and concatenated to constitute the final histogram.

TABLE 3: Details of MLP network parameter.

| Number | Functions used for MLP    | Used parameters  |
|--------|---------------------------|------------------|
| 4      | Activation                | Sigmoid function |
| 5      | Hidden Layer              | 1                |
|        | Number of hidden units    | 20               |
| 6      | Input neurons             | 1352             |
| 7      | Output neuron             | 1                |
| 8      | Maximum mean square error | 0.001            |
| 9      | Number of iterations      | 2000             |

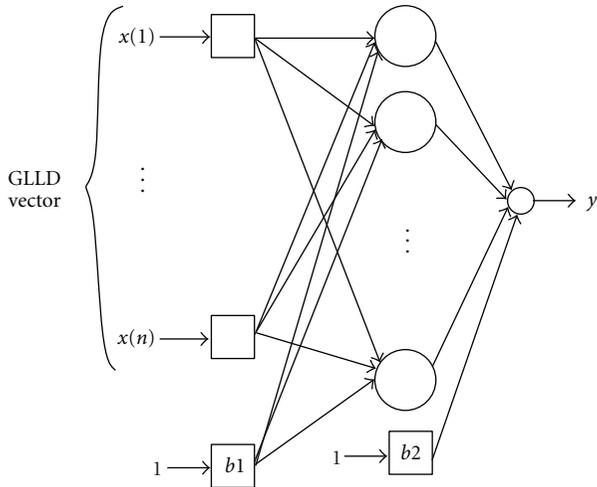


FIGURE 8: MLP classifier architecture.

the activation function which corresponds to a sigmoid function, and the network output is thus defined as follows:

$$y = g(\omega^T x_b) = g\left(\sum_{i=1}^d \omega_i x_i - b\right). \quad (16)$$

For each ROI sample, GLLD features are computed and used in the classification step as inputs of the neural network. Figure 8 illustrates the applied neural network.

The evaluation of the effectiveness of the training is based on the measure of the network relative error as follows:

$$E = \frac{\sum_{i=1}^n (y - T)}{\text{number of applied samples}}, \quad (17)$$

where  $y$  is the ROI corresponding to masses or nonmasses resulting from ANN and  $T$  corresponds to the target. The use of artificial network may lead to low error rates. After the training step, generalization errors may be evaluated for various features and network conditions. Figure 9 maps the different steps of the proposed method.

In the following section, the obtained results for ANN, SVM, and kNN classifiers will be illustrated for comparison purposes.

## 5. Experimental Results

This section is composed of the following parts. First of all, the database used in the evaluation is presented. Afterwards,

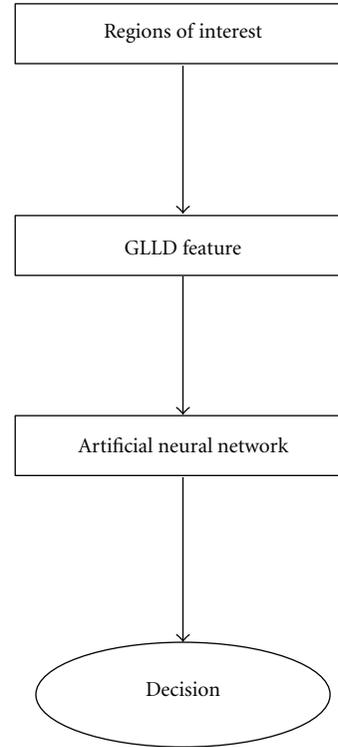


FIGURE 9: Implementation of the proposed method.

we illustrate the results for different rotation invariant rows under setting. We then do an investigation of the feature relevance, by using each of the proposed features (SGLLD, MGLLD, CGLLD) separately as input vector to the classifier. In further step, we made the classifier input a concatenated vector made up with different feature vectors. The obtained feature vector allow as to compare different methods of classification. Then, we made our experiments for different ROI image sizes. Finally, a comparative study of our proposal to those in the state-of-the-art will be done for a fair evaluation.

**5.1. Mammogram Dataset.** Our approach has been evaluated based on publicly available database taken from the DDSM database [40].

DDSM contains 2620 individuals, available in 43 volumes. A volume corresponds to the collection of different cases. A case is the collection of all information to the mammography exam of one patient. Each case in the DDSM database contains two images, of each breast, that is, in each case the mammograms include a craniocaudal and mediolateral oblique view (CC and MLO, resp.). The DDSM database provides the metadata (date of study, breast density, assessment categories, etc.) of each abnormality using the breast imaging reporting and data System (BI-RADS) lexicon, it provides, also, the corresponding chain codes of the suspicious regions. With these chain codes, the outlines of the abnormalities may be identified. The DDSM provide delineations of mass regions. However, precision of such delineations is not adequate for validation in our approach,

TABLE 4: Classification rates when using different number of rotation invariant rows under settings of  $(P, R) = (8, 1)$ ,  $(P, R) = (16, 2)$ , and  $(P, R) = (24, 3)$ .

| $P, R$ | 8, 1 | 16, 2 | 24, 3 |
|--------|------|-------|-------|
| $A_z$  | 0.93 | 0.94  | 0.95  |

TABLE 5:  $A_z$  comparison of the different methods of classification (SVM, kNN, ANN) when utilizing GLLD as a feature extraction technique.

|                                    | $A_z$ |     |      |
|------------------------------------|-------|-----|------|
|                                    | kNN   | SVM | ANN  |
| $\text{GLLD}_{24,3}^{\text{riu}2}$ | 0.89  | 0.9 | 0.95 |

since it was done on downsampled images of DDSM database (by factor of 8) [41], see Section 2. Therefore, we were based in the extraction of ROI's on manual segmentation entertained by two expert radiologists of more than 30 years of clinical experience from the Farabi imaging. We should also notice that all the considered masses in DDSM are biopsically proven ones.

**5.2. Influence of Rotation Invariant Rows under Settings.** The study was based on 1000 ROI extracted from mammograms from DDSM database. These ROIs were randomly selected and separated into two sets: 500 samples for training and 500 samples for tests. In the training set as well as in the setting set, we used 250 samples corresponding to masses and 250 samples corresponding to nonmasses. The evaluation of our mass detection algorithm is performed by applying a leave-one-out methodology, where the input ROI is classified by using the appropriate classification method and the procedure is reapplied for all the remaining ROIs used as input.

From the results presented in Table 4, we can conclude that with  $(P, R) = (24, 3)$ , the area under curve for the GLLD is increased from 0.93 to 0.95. Rows under settings choice affect very slightly  $A_z$  performance. In the next experiments,  $\text{GLLD}_{24,3}^{\text{riu}2}$  will be used.

**5.3. Investigation of the Method of Classification.** From the comparative study, as shown in Table 5, we note that the ANN provides the best results. This can be attributed to its higher performance as function approximator.

**5.4. Investigation of the Relevance of the Features.** As illustrated in Figure 10 and Table 6 the CAD system achieves better performance ( $A_z = 0.93$ ) when using the sign component than the modulus component. However, their fusion may provide much better results in texture classification than using either sign or modulus ( $A_z = 0.95$ ). The AUC of the  $\text{GLLD}_{24,3}^{\text{riu}2}$  after the fusion of the three operators  $\text{SGLLD}_{24,3}^{\text{riu}2}$ ,  $\text{MGLLD}_{24,3}^{\text{riu}2}$ , and  $\text{CGLLD}_{24,3}^{\text{riu}2}$  using the ANN as classifier is about ( $A_z = 0.95$ ) for the experimental set.

As can be noted from Table 6, the  $\text{GLLD}_{24,3}^{\text{riu}2}$  feature provides useful information about local gray level which is the

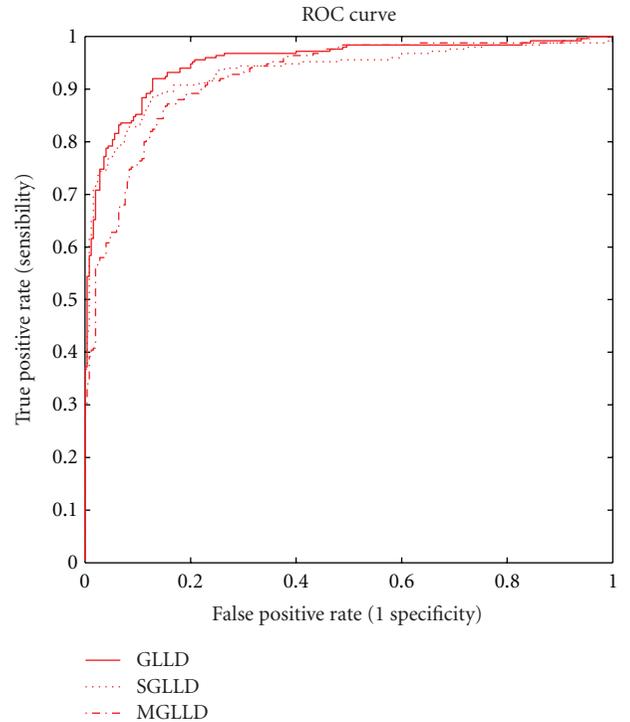


FIGURE 10: ROC curve corresponding to a subset of 1000 ROIs images from the DDSM database.

most significant one for mass detection in mammographic images.

**5.5. Results Varying the ROI Image Sizes.** Based on the size of the lesion, we use six different group of ROI images, which is an important aspect for the correct classification of the masses. These classes correspond to the following specified mass sizes intervals [9]: size 1:  $<10 \text{ mm}^2$ , size 2:  $(10-60) \text{ mm}^2$ , size 3:  $(60-120) \text{ mm}^2$ , size 4:  $(120-190) \text{ mm}^2$ , size 5:  $(190, 270) \text{ mm}^2$ , and size 6:  $>270 \text{ mm}^2$ . However, the used numbers of masses in each class size were, respectively, 28, 32, 37, 57, 69, and 33 masses. Table 7 illustrates the  $A_z$  values for each class of ROI image sizes and the obtained mean  $A_z$  values. We include in this table a quantitative comparison with the work of Oliver et al. [8, 9] where the same sizes are considered. Oliver et al. in [8, 9] have used our database of ROI ratio (1/3), the same specified mass size intervals, and the same number of masses in each class size.

The results presented in Table 7 have shown that the  $\text{GLLD}_{24,3}^{\text{riu}2}$  features are effective for mass detection at different ROI image sizes, and the latter is an important aspect for correct classification of the masses. Our method proves its performance in the most difficult case, which correspond to the smaller masses. Note that for this proposed ratio, better results are obtained for all the size intervals.

**5.6. Comparison with Some Consequent Approaches on Mass Detection in the State-of-the-Art.** Table 8 shows the different  $A_z$  performance values for different approaches in the

TABLE 6:  $A_z$  for different existing local pattern-based features and the GLLD proposed one.

|       | LBP + ANN | SGLLD $_{24,3}^{riu2}$ + ANN | MGLLD $_{24,3}^{riu2}$ + ANN | GLLD $_{24,3}^{riu2}$ + ANN |
|-------|-----------|------------------------------|------------------------------|-----------------------------|
| $A_z$ | 0.89      | 0.93                         | 0.92                         | 0.96                        |

TABLE 7: Obtained  $A_z$  values (ratio 1/3) of the classification of masses according to the ROI image sizes. The final column illustrates the mean  $A_z$  value. Size 1 to size 6 correspond to the different ROIs image sizes, from smaller to bigger one.

| Method                      | $A_z$  |        |        |        |        |        | Mean |
|-----------------------------|--------|--------|--------|--------|--------|--------|------|
|                             | Size 1 | Size 2 | Size 3 | Size 4 | Size 5 | Size 6 |      |
| Oliver et al. [8]           | 0.53   | 0.7    | 0.7    | 0.68   | 0.72   | 0.83   | 0.7  |
| Oliver et al. [9]           | 0.81   | 0.83   | 0.87   | 0.84   | 0.89   | 0.93   | 0.86 |
| GLLD $_{24,3}^{riu2}$ + ANN | 0.98   | 0.99   | 0.97   | 0.92   | 0.9    | 0.93   | 0.94 |

TABLE 8: Presented  $A_z$  values for different methods in the state-of-the-art aiming at mass detection and that of the proposed one.

| Method                      | Number of used ROIs | Ratio | $A_z$ |
|-----------------------------|---------------------|-------|-------|
| Qian et al. [6]             | 800                 | 1/3   | 0.86  |
| Chang et al. [12]           | 600                 | 1/1   | 0.83  |
| Varela et al. [10]          | 120                 | 1/1   | 0.90  |
| Oliver et al. [9]           | 1792                | 1/2   | 0.83  |
| Tourassi et al. [13]        | 1820                | 1/1   | 0.81  |
| GLLD $_{24,3}^{riu2}$ + ANN | 1100                | 1/1   | 0.95  |
| Human observers             | 1100                | 1/1   | 0.87  |

state-of-the-art presented in Section 2. Such which were represented comparison demonstrates the effectiveness of the proposed GLLD operator in mass detection. For instance, [10, 12, 13], which used ratio (1/1) the same as in our cases, obtained  $A_z$  values 0.83, 0.90, and 0.81, respectively, as our obtained  $A_z = 0.95$ .

## 6. Conclusion

CAD systems have been used and gained greater utility in recent years, as a second virtual reader for the medical images, contributing to increase an early detecting of breast cancer. This work presents a new method for mammographic mass detection based on textural features. Our proposal combines gray level as well as local differences. The combined descriptors are, respectively, SGLLD $_{24,3}^{riu2}$ , MGLLD $_{24,3}^{riu2}$ , and CGLLD $_{24,3}^{riu2}$  providing a final texture feature descriptor named GLLD, which will be used to classify the ROIs to masses and mass free. The ANN classifier gives better performances in term of classification owing to its higher function approximation. Different image sizes were considered for better improving detection rates. Finally, a comparative study with previous works was done for fairer evaluation. Such comparison illustrates that our proposed method leads to the best performance  $A_z = 0.95$ . The specialists who I have collaborated with found that the proposed CAD improved the sensitivity of mammography screening. In fact, CAD system is useful in situations where there is a high interobserver variability, lack of trained observers, or

impossibility to perform the double reading with two or more radiologists as stated in the BIRADS categories. Future work will be focused in the classification of masses into the four Breast Imaging-Reporting and Data System (BI-RADS) categories.

## Acknowledgments

The authors would like to thank DR. Abid Riadh, Dr. Turki Hassen, DR. Kassatr Fathi, and DR. Haddouk Bechir, Radiologists at El Farabi Imaging center, Sfax, Tunisia, and at the Faculty of Medecine of Sfax for their helpful discussions and advises. Sincerely thanks are addressed to MVG and VGG for sharing their source codes of LBP.

## References

- [1] Eurostat, "Health statistics—Atlas on mortality in the European Union," European Commission, 2002.
- [2] American Cancer Society, "Cancer Facts and Figures 2012," American Cancer Society, Atlanta, Ga, USA, 2012.
- [3] J. Esteve, A. Kricke, J. Ferlay, and D. Parkin, "Facts and figures of cancer in the European community," International Agency for Research on Cancer, Lyon, France, 1993.
- [4] N. Mourali, "Haro sur la première cause de mortalité dans le monde," Livret Santé. 2010.
- [5] E. A. Sickles, "Breast cancer screening outcomes in women ages 40–49: clinical experience with service screening using modern mammography," *Journal of the National Cancer Institute. Monographs*, no. 22, pp. 99–104, 1997.
- [6] W. Qian, X. Sun, D. Song, and R. A. lark, "Digital mammography—wavelet transform and Kalman-filtering neural network in mass segmentation and detection," *Academic Radiology*, vol. 8, no. 11, pp. 1074–1082, 2001.
- [7] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, no. 5, pp. 309–319, 2002.
- [8] A. Oliver, J. Martí, R. Martí, A. Bosch, and J. Freixenet, "A new approach to the classification of mammographic masses and normal breast tissue," in *Proceedings of the IAPR International Conference on Pattern Recognition*, vol. 4, pp. 707–710, 2006.
- [9] A. Oliver, X. Lladó, J. Martí, and R. Martí, "False positive reduction in breast mass detection using two-dimensional PCA," *Lecture Notes on Computer Science*, pp. 154–161, 2007.

- [10] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto, and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms," *Computers in Biology and Medicine*, no. 2, pp. 214–226, 2007.
- [11] O. M. Leonardo, B. J. Geraldo, O. C. S. Arist, C. P. Anselmo, and G. Marcelo, "Detection of masses in digital mammograms using K-means and support vector machine," *ELCVIA*, vol. 8, no. 2, pp. 39–50, 2009.
- [12] Y. H. Chang, L. A. Hardesty, C. M. Hakim et al., "Knowledge-based computer-aided detection of masses on digitized mammograms: a preliminary assessment," *Medical Physics*, vol. 28, no. 4, pp. 455–461, 2001.
- [13] G. D. Tourassi, B. Harrawood, S. Singh, J. Y. Lo, and C. E. Floyd, "Evaluation of information theoretic similarity measures for content-based retrieval and detection of masses in mammograms," *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 140–150, 2007.
- [14] P. M. Vacek, B. M. Geller, D. L. Weaver, and R. S. Foster, "Increased mammography use and its impact on earlier breast cancer detection in Vermont, 1975–1999," *Cancer*, vol. 94, no. 8, pp. 2160–2168, 2002.
- [15] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, no. 2, pp. 211–215, 1967.
- [16] T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center," *Radiology*, vol. 220, no. 3, pp. 781–786, 2001.
- [17] S. van Engeland and N. Karssemeijer, "Combining twomammographic projections in a computer aided mass detection method," *Medical Physics*, vol. 34, pp. 898–905, 2007.
- [18] A. Oliver, X. Lladó, J. Freixenet, and J. Martí, "False positive reduction in mammographic mass detection using local binary patterns," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 286–293, 2007.
- [19] J. Zhang and T. Tan, "Brief review of invariant texture analysis methods," *Pattern Recognition*, vol. 35, no. 3, pp. 735–747, 2002.
- [20] R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, no. 3-4, pp. 312–348, 2007.
- [21] A. Rojas and A. Nandi, "Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection," *Computerized Medical Imaging and Graphics*, vol. 32, no. 4, pp. 304–315, 2008.
- [22] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [23] R. M. Haralik, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems*, vol. 3, pp. 610–621, 1973.
- [24] T. Randen and J. H. Husy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 291–310, 1999.
- [25] J. L. Chen and A. Kundu, "Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden Markov model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 208–214, 1994.
- [26] H. Deng and D. A. Clausi, "Gaussian VZ-MRF rotation-invariant features for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 951–955, 2004.
- [27] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.
- [28] M. Varma and A. Zisserman, "Texture classification: are filter banks necessary," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 691–698, 2003.
- [29] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [30] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.
- [31] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikäinen, "Texture discrimination with multidimensional distributions of signed gray-level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, 2001.
- [32] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram Fourier features," in *Proceedings of the Scandinavian Conference on Image Analysis*, vol. 5575 of *Lecture Notes in Computer Science*, pp. 61–70, Springer, Berlin, Germany, 2009.
- [33] M. Varma and A. Zisserman, "Texture classification: are filter banks necessary?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 691–698, June 2003.
- [34] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.
- [35] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2009.
- [37] D. E. Rumelhart and G. E. Hinton, in *Learning Internal Representations By Error Propagation, Parallel Distributed Processing*, vol. 1, pp. 318–362, MIT Press, Cambridge, Mass, USA, 1986.
- [38] J. Padmavati, "A comparative study on breast cancer prediction using RBF and MLP," *International Journal of Scientific and Engineering Research*, vol. 2, no. 1, 2011.
- [39] M. Oravec and J. Pavlovicová, "Face Recognition methods based on feedforward neural networks, principal component analysis and self-organizing map," *Radio Engineering*, vol. 16, no. 1, pp. 51–57, 2007.
- [40] D. K. M. Heath and K. W. Bowyer, "Current status of the digital database for screening mammography," in *Proceedings of the 4th International Workshop on Digital Mammography*, 1998.
- [41] E. Song, S. Xu, X. Xu et al., "Hybrid segmentation of mass in mammograms using template matching and dynamic programming," *Academic Radiology*, vol. 17, no. 11, pp. 457–460, 2010.

## Research Article

# DCE-MRI and DWI Integration for Breast Lesions Assessment and Heterogeneity Quantification

C. Andrés Méndez,<sup>1</sup> Francesca Pizzorni Ferrarese,<sup>1</sup>  
Paul Summers,<sup>2</sup> Giuseppe Petralia,<sup>2</sup> and Gloria Menegaz<sup>1</sup>

<sup>1</sup>Dipartimento di Informatica, Università degli Studi di Verona, Strada le Grazie 15, CA'Vignal, 37134 Verona, Italy

<sup>2</sup>Divisione di Radiologia, Istituto Europeo di Oncologia, Via Ripamonti 435, 20141 Milano, Italy

Correspondence should be addressed to C. Andrés Méndez, andres.mendezg@gmail.com

Received 20 July 2012; Revised 9 October 2012; Accepted 23 October 2012

Academic Editor: Carl-Fredrik Westin

Copyright © 2012 C. Andrés Méndez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to better predict and follow treatment responses in cancer patients, there is growing interest in noninvasively characterizing tumor heterogeneity based on MR images possessing different contrast and quantitative information. This requires mechanisms for integrating such data and reducing the data dimensionality to levels amenable to interpretation by human readers. Here we propose a two-step pipeline for integrating diffusion and perfusion MRI that we demonstrate in the quantification of breast lesion heterogeneity. First, the images acquired with the two modalities are aligned using an intermodal registration. Dissimilarity-based clustering is then performed exploiting the information coming from both modalities. To this end an ad hoc distance metric is developed and tested for tuning the weighting for the two modalities. The distributions of the diffusion parameter values in subregions identified by the algorithm are extracted and compared through nonparametric testing for posterior evaluation of the tissue heterogeneity. Results show that the joint exploitation of the information brought by DCE and DWI leads to consistent results accounting for both perfusion and microstructural information yielding a greater refinement of the segmentation than the separate processing of the two modalities, consistent with that drawn manually by a radiologist with access to the same data.

## 1. Introduction

Responses to cancer treatment are increasingly differentiated based not only on tumor type, but also on genetic and histochemical biomarkers. Exemplifying the progress in this respect is breast cancer. Biopsy-derived histological biomarkers offer high biological specificity and play an important role in determining the choice of chemotherapeutic agent. As different parts of a tumor often show different histological signatures or have evolved to different stages of tumor progression that may impact on their response to a given therapy, it is important to obtain a complete coverage of the tumor. Biopsies, however, are difficult to localize within the breast, are subject to sampling errors, and can seldom be repeated. Thus, there is growing clinical interest

in the possible role of imaging to describe anatomical and physiological heterogeneity of tumors [1, 2].

Magnetic resonance imaging (MRI) methods such as dynamic contrast enhanced (DCE) and diffusion weighted (DW) MRI methods are amongst those of interest as they provide noninvasive digital biomarkers with good spatial coverage and repeatability [3]. DCE-MRI uses serial acquisition of images during and after the injection of intravenous contrast agent and has been shown to reflect tumor vascularity [4, 5]. DWI, on the other hand, generates images that are sensitized to water displacement at the diffusion scale and can be used to calculate a quantitative index reflecting the apparent freedom of diffusion (apparent diffusion coefficient (ADC)). Preclinical and clinical data show that ADC reflects regional cellularity [6–8].

DCE-MRI has a high sensitivity for breast cancer detection (89–100%), while DWI has shown utility in predicting suitable therapies and monitoring response [9]. A recognized weakness of DCE and DW-MRI is their lack of specificity between tumor types as overlap between the findings of benign and malignant lesions results in variable specificity (37–86%) [9]. This is not entirely surprising given that across cancer types the common features tend to include such processes as cell proliferation, angiogenesis, and necrosis. The ability of DCE- and DW-MRI to provide a spatial depiction of these anatomical and physiological conditions within a tumor makes them natural tools for probing tumor heterogeneity. The reporting of MRI has long relied on visual assessment of several scans having different contrasts, but in relation to breast cancer, few studies have exploited this inherently multiparametric data in a unified manner [10–12]. Moreover, the most recent works mainly address the problem of comparing and retrospectively integrating the contributions from the different modalities, without exploiting the conjunct information. Nevertheless, these works have highlighted the potential of combining DCE-MRI and DWI to differentiate the core of the tumor from peritumoral tissues and normal tissues and thus provide an indication of lesion heterogeneity [13].

In this work, we propose the multimodal integration of the information provided by DCE-MRI and DWI of breast cancer lesions for evaluating their heterogeneity, that is, to divide the lesion into zones that share certain similarity when using combined information coming from different imaging domains. The ultimate intention of this protocol is to allow a more extensive, reproducible characterization of heterogeneity in tumors that have been previously identified by a clinician.

In all previous reports on breast lesion segmentation the representation of DCE curves and ADC maps has been that of features in a vector space defined by the image values [14–17]. In this work a different approach is followed exploiting dissimilarity-based representations (DBR) [18]. The concept of dissimilarity-based representation consists of focusing on the contrast, or distance, between objects and of measuring it by a suitable criterion. The term *object* refers, in the present context, to the information represented by each particular voxel. This information need not be of a single type and in this case consists of both signal intensities (i.e., the time-intensity enhancement curve for DCE-MRI) and the ADC parameter value (derived from DW-MRI). A key concept in DBR is that of a *proximity relation* between two objects, which does not need to be explicitly represented in a feature space. Objects are characterized through pairwise dissimilarities; instead of using an absolute characterization of the objects by a set of features, problem-centric knowledge is used to define a measure that estimates the dissimilarity between objects. Here, both DCE and DWI contribute to such a measure leading to a novel multimodal approach to tissue characterization.

This paper is organized as follows. Section 2 describes the pipeline including the clustering and registration processing steps. Section 3 presents the results, which are then discussed in Section 4, and Section 5 derives conclusions.

## 2. Materials and Methods

This section provides an overview of the pipeline shown in Figure 1 and details the methodological choices with respect to both clustering and registration. The DCE-MRI data are first visually inspected to identify a time point where the lesion has the higher contrast with respect to the surrounding tissue. Multimodal registration is carried out between DW-MRI and DCE-MRI images, allowing a spatial mapping of both volumes. Dissimilarity-based clustering is then performed integrating information from both acquisition modalities. Statistical analysis, consisting of nonparametric tests, were applied on the ADC distributions defined by the obtained clusters. An assessment of the results was carried out by clinical experts, and, for the sake of completeness, an evaluation of the tightness and separation of the clusters was also performed.

**2.1. Multimodal Registration.** In order to perform voxelwise dissimilarity-based clustering that incorporates both DCE-MRI and DWI data, it is necessary to first spatially align the two datasets. The problem of registering between DCE-MRI and DWI becomes an increasingly difficult task in a highly compressible and elastic tissues like the breast, with its inhomogeneous anisotropic soft tissue, inherent nonrigid behavior, and lack of solid landmarks to guide the registration as fixed references. A standard registration protocol was used. Due to the highly distinct contrast and intensity characteristics of the two modalities as well as the low resolution of the DWI volumes, the registration process was divided into two steps, each following a standard multiresolution strategy. In the first step, rigid and affine transformations were performed successively in order to align and match the features of the fixed (DCE-MRI) and moving (DWI) images following a 5-level Gaussian scale space. In the second step a multiresolution cubic B-spline transformation with a regularization penalty was performed to elastically refine the alignment. Lesion-specific masks based on regions delineated by clinical experts were used in order to assign a greater weight to the voxels in the lesion area [19]. Normalized mutual information (NMI) was used as registration metric. In order to regularize the deformation, we used a bending energy penalty which is based on the spatial derivatives of the transformation [20]. The methodology used for registration was implemented in Elastix [19], and all the steps have been widely validated in literature [20, 21].

The registration protocol was applied to the  $b_0$  images from the DWI dataset and their transformation to the DCE-MRI space validated for each subject through visual inspection by an expert. The resulting transformation was applied to the remaining  $b$ -values, and the ADC was estimated on the transformed DWI images.

**2.2. Dissimilarity-Based Clustering Methodology.** The next step in the processing methodology is the construction of a dissimilarity matrix. This matrix consists of a set of row vectors, one for each voxel. These vectors represent the voxels

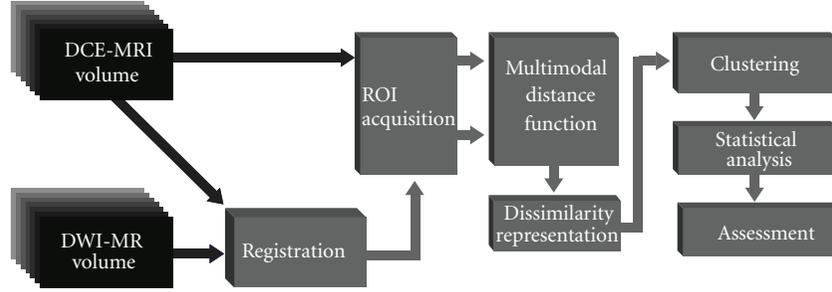


FIGURE 1: Perfusion/diffusion analysis and integration pipeline.

in a vector space constructed by the dissimilarities to each other voxel. Usually, such a space can be safely treated as an Euclidean space equipped with the standard inner product definition.

Let  $X = \{x_1, \dots, x_n\}$  be a voxel-based dataset. Given a dissimilarity function, a data-dependent mapping  $D$  is defined as  $D(\cdot, R) : X \rightarrow \mathbb{D}^n$  linking  $X$  to the so-called dissimilarity space [22]. The complete dissimilarity representation yields a square matrix consisting of the dissimilarities between all pairs of objects, such that every object is described by an  $n$ -dimensional dissimilarity vector  $D(x, X) = [d(x, x_1) \dots d(x, x_n)]^T$ .

A distance function  $D_{DCE}$  based on the adaptive dissimilarity index first proposed in [23] has been exploited in a previous work [24] for calculating the pairwise proximity between DCE-MRI perfusion curves. There are two main approaches to quantifiably compare two time series: one makes use of the distances between the absolute values of their elements while the other focuses on the similarity of their behavior along time. Unlike conventional time-series distance functions, which focus only on the closeness of the values observed at corresponding points in time, ignoring the interdependence relationship between elements that characterize the time-series behavior, the proposed distance function takes into account the proximity with respect to values as well as the temporal correlation for the proximity with respect to behavior. For two voxel-derived perfusion curves  $S_1 = (u_1, \dots, u_p)$  and  $S_2 = (v_1, \dots, v_p)$ , closeness with respect to behavior is defined as the combination of their monotonicity, that is, if both curves increase or decrease simultaneously, and the closeness of their growth rate over a determined period [23]. Both criteria are quantified by the temporal correlation present in the first term of the distance function  $D_{DCE}$ , (1). The complete distance function  $D_{DCE}$  for DCE-MRI derived perfusion curves is defined as follows:

$$D_{DCE}(S_1, S_2) = \frac{2}{1 + \exp(\text{Cort}(S_1, S_2))} dH(S_1, S_2), \quad (1)$$

where  $S_1 = (u_1, \dots, u_p)$  and  $S_2 = (v_1, \dots, v_p)$  are two voxel-derived perfusion curves sampled at time instants  $(t_1, \dots, t_p)$  [23, 25]. Cort is the temporal correlation (2), and  $dH$  is the

Hausdorff distance, defined in (3), which is used to measure the distance between both voxelwise perfusion curves:

$$\text{Cort}(S_1, S_2) = \frac{\sum_{i=1}^{p-1} (u_{(i+1)} - u_i)(v_{(i+1)} - v_i)}{\sqrt{\sum_{i=1}^{p-1} (u_{(i+1)} - u_i)^2} \sqrt{\sum_{i=1}^{p-1} (v_{(i+1)} - v_i)^2}}, \quad (2)$$

$$dH(S_1, S_2) = \max \left\{ \max_{u \in S_1} \min_{v \in S_2} \|u - v\|, \max_{v \in S_2} \min_{u \in S_1} \|v - u\| \right\}. \quad (3)$$

The integration of the diffusion information into the dissimilarity function is accomplished through the addition of an ADC-dependent term  $D_{ADC}$  (4). This term is defined as a sigmoid function which makes use of the normalized difference between the ADCs ( $ADC_{S_1}$  and  $ADC_{S_2}$ ) of the voxels under consideration, which ranges from 0 to 1:

$$D_{ADC}(S_1, S_2) = \frac{1}{1 + \exp(-k_{ADC}(\|(ADC_{S_1} - ADC_{S_2}) / \max\{ADC_{ROI}\}\| - 0.5))}. \quad (4)$$

The tuning parameter  $k_{ADC}$  weights the contribution of  $D_{ADC}$  to the complete dissimilarity measure  $D$  by modulating the shape of the sigmoid function. When the value of the normalized difference between ADCs is low, denoting similar ADC values between voxels, the dissimilarity function  $D_{ADC}$  approaches zero. On the contrary, when the value of the normalized difference between ADCs is high, denoting a large dissimilarity between ADC values between voxels,  $D_{ADC}$  approaches one, making the overall dissimilarity measure approaches the value of  $D_{DCE}$ . The impact of the different values of  $k_{ADC}$  is illustrated in Figure 2.

The complete dissimilarity function  $D$  is then the product of  $D_{ADC}$  and  $D_{DCE}$ :

$$D = D_{ADC} \cdot D_{DCE}. \quad (5)$$

This global measure enables the monitoring of the performance as a function of the relative weight given to the ADC, as well as of different values of  $k_{ADC}$ .

**2.3. Performance Assessment.** In each of the patients, a ROI was delineated by an expert around the lesion in the motion-corrected DCE-MRI volumes. Since unsupervised classification is sensitive to the general structure and distribution

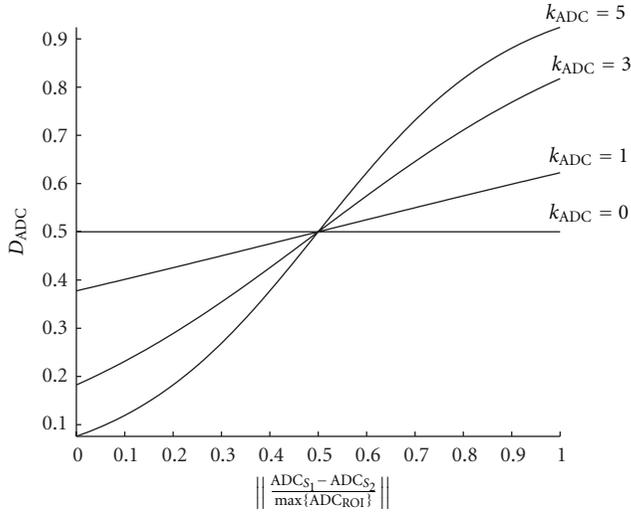


FIGURE 2: Effects of varying the tuning parameter  $k_{ADC}$  from (4).

of the data, the ROI was drawn just exceeding the area of the enhancing lesion, allowing for a clear delineation of the heterogeneity of the lesion inside the ROI. The time-intensity curves normalized to the baseline at  $t = 0$  and the corresponding ADC values from the voxels inside the ROI were treated as independent objects on a voxel by voxel basis. Using  $D$  from (5), a dissimilarity matrix was derived on a slicewise basis from the pairwise dissimilarities of the elements in the corresponding ROI. In such a space, each element was represented by a row vector whose dimensionality was defined by the cardinality of the ROI.

Once the dissimilarity space was constructed, the  $K$ -means algorithm [26] was used to group the voxels in the ROI into clusters. The initial centroids were calculated automatically following a preliminary clustering step with a random 10% subsample, as a strategy to improve the algorithm initialization avoiding a misplacement of the initial seeds.  $K$ -means minimize the sum over all clusters of the within-cluster sums of point-to-cluster-centroid distances using, in this case, the squared Euclidean distance.

For selecting the  $K$  number of clusters the standard clinical assessment protocol has been taken into consideration. It considers only three classes (persistent, plateau, and wash-out). An additional has been included for the surrounding tissue considering that the ROI exceeds the estimated limits of the enhancing lesion.

In order to perform a comparison with established methods the clustering procedure was also performed following a morphologic feature-based approach. This method relies on descriptors derived from the voxelwise time-intensity curves, comprising mainly specific characteristics of the shape of such curve. The features extracted from the DCE-MRI voxelwise time-intensity curves are baseline, maximum signal difference, time to peak, area under curve, maximum enhancement, wash-in rate, maximum slope of increase, wash-out rate, and the intercept of the line fitting the tail

of the time-intensity curve with the axis  $t = 0$ . The use and definition of these morphologic features to describe the contrast agent intake can be found in the related literature [14, 17, 27]. Further, the clustering procedure was repeated incorporating the ADC of each voxel as an additional feature to the morphologic descriptor vectors calculated previously. The ADC and the morphologic features were standardized by subtracting their mean and dividing by their standard deviation. The results of these two procedures were compared with our method in order to assess the clustering and data representation outcome.

**2.4. Patient Population.** Data were acquired from 21 patients (age  $50 \pm 13.8$  years). All the patients had been diagnosed to have primary ductal carcinoma.

DWI was acquired with a single-shot spin-echo (SE) echo planar imaging (EPI) sequence in three orthogonal diffusion encoding directions ( $x$ ,  $y$ , and  $z$ ) using 4  $b$  values (0, 250, 500 and 1000  $s/mm^2$ ) with parallel imaging (acceleration factor 2). Subjects were breathing freely, with no gating applied. The dataset consisted of 30 transverse slices (slice thickness 5 mm, no slice gap) and TR/TE 4800/71 ms, matrix  $90 \times 150$  over the field of view (FOV)  $184.5 \times 307.5$  mm.

DCE-MRI was performed using a 3D T1-weighted FLASH sequence (TR/TE 7.4/4.7 ms) with a flip angle of  $25^\circ$  and an acquisition matrix of  $384 \times 384 \times 128$  and field of view (FOV)  $340 \times 340 \times 166$  mm. Each 120-slice set was collected in 90 s at 8 time points for approximately 12 min of scanning. A catheter placed within an antecubital vein delivered 0.1 mmol/kg of the contrast agent, gadopentetate dimeglumine, (Magnevist, Wayne, NJ, USA) over 20 s (followed by saline flush) after the acquisition of one baseline dynamic scan. The DCE-MRI time series was motion corrected using the scanner manufacturer's in-line procedure.

### 3. Results

The regions resulting from dissimilarity-based clustering were rendered as colored overlays on the morphological images on each slice. The results from a representative patient are displayed in Figure 3. After clustering was performed on the normalized curves, the resulting clusters were assessed by the radiologists to validate the segmentation of both the central tumoral and surrounding regions. Figure 3(b) shows examples of the clusters obtained, while Figures 3(c) and 3(d) represent the plots of the average time-intensity perfusion curves calculated on the raw and normalized data, respectively. The plots show the impact that the normalization step has in highlighting the intercluster differences. The central region exhibits a characteristic pattern in the DCE-MRI of a high early enhancement followed by a rapid washout, indicative of angiogenesis (Figure 3(d), red line). Typically, surrounding this central region lays a cluster featuring a pattern of rapid enhancement followed by a signal plateau (Figure 3(d), orange line). The outermost cluster surrounding these two central regions features a slow enhancement behavior (Figure 3(d), yellow line).

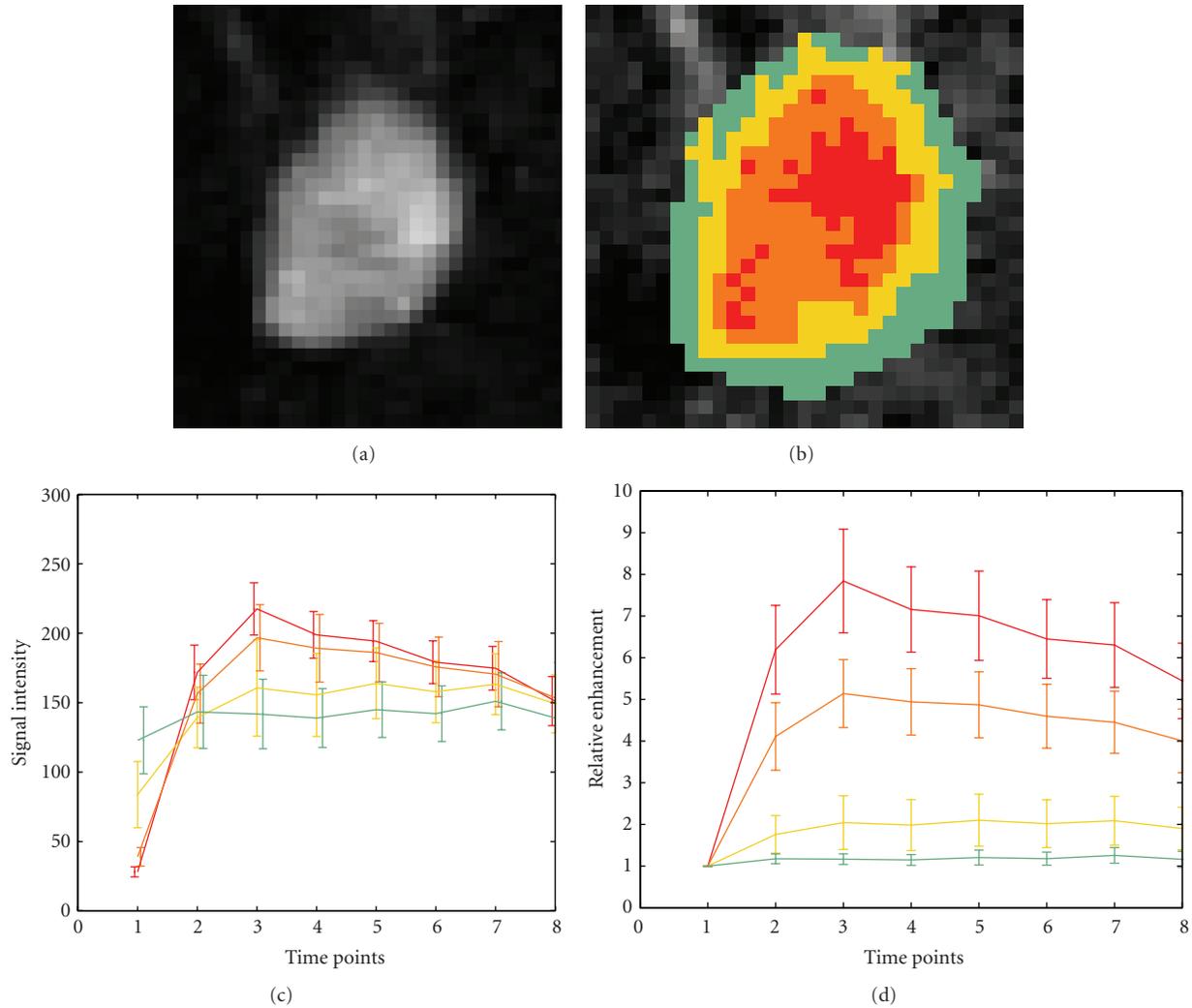


FIGURE 3: DCE-MRI image (a) and overlaid lesion clustering (b), comparison between the average raw (c), and normalized curves (d) calculated for each cluster.

The voxels corresponding to the each cluster were extracted from the spatially registered 3D ADC maps in order to perform statistical analysis. The analysis was carried out in the whole 3D ROI, that is, taking into account the ADC values corresponding to all the clustered slices as a single volume. Normality tests (Jarque-Bera) revealed that the ADC values for the different clusters analyzed were not normally distributed. Accordingly, a nonparametric test (Wilcoxon-signed-rank test) was used ( $P = 0.05$ ) to evaluate whether the tumor's subregions corresponded to regions in the ADC maps with statistically different PDFs. In this way we found that the distributions of the ADC values in the DCE-MRI defined regions were statistically different, in each one of the two conditions, in 19 out of 21 patients.

The radiologist reviewed the overlays in comparison to the DCE seen as a dynamic loop, the DWI images, and the ADC maps derived from them, as well as T2 STIR images. Criteria for the review were whether or not any of the subregions obtained by the method corresponded to a zone

of necrosis based on the complete set of images and whether one or more regions that would be classified as either benign or malignant have been subdivided.

Figure 4 illustrates a typical case setting  $k_{ADC}$  to 1, 3, and 5. From the obtained results it was highlighted by the experts the usefulness of varying the parameter  $k_{ADC}$  to emphasize different characteristics of the lesion. A high  $k_{ADC}$  allows the discrimination between the core tumor and the surrounding regions by giving a higher weight to the difference between ADCs. This is mainly due to the fact that there is a progressive increase in ADC from the core of the tumor to peritumor tissues to normal tissues that leads to the possibility to use the ADC for locoregional staging [28]. Lowering  $k_{ADC}$  allows the subdivision of the core based on DCE-MRI dissimilarity and the evaluation of the heterogeneity of the tumor thanks to the balanced contribution of DCE and DWI in the distance function  $D$ .

For the sake of cluster comparison and validation among different methods, the silhouette analysis was used in all

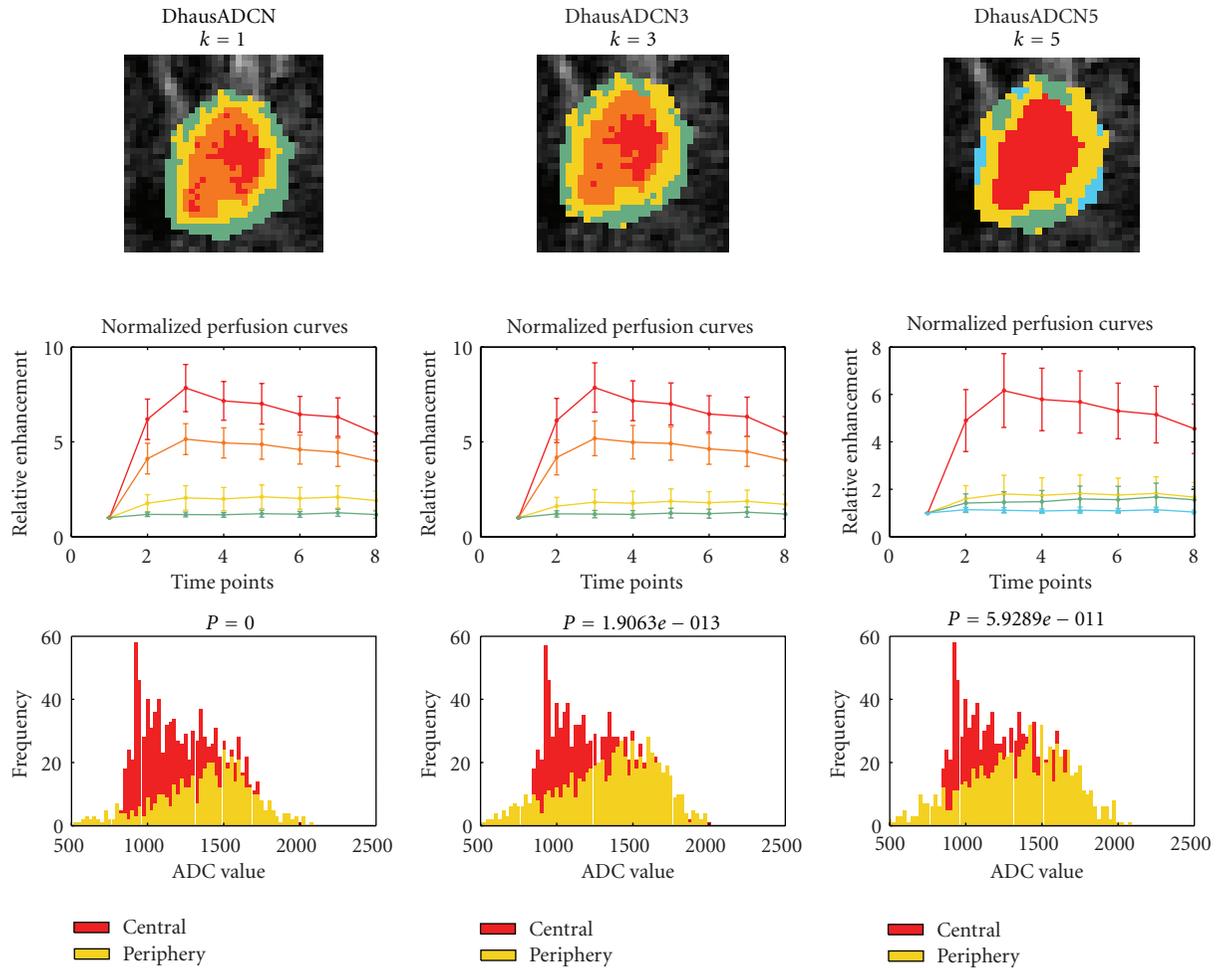


FIGURE 4: Clustering results using different values for the tuning parameter  $k_{ADC}$  (1, 3, and 5).

the clustering results. The silhouette analysis measures how close each point in one cluster is to points in the same cluster and how far away it is to points in the neighboring clusters. This is performed by quantitatively comparing the clusters by their tightness and separation and its average width provides an evaluation of cluster validity [29]. The silhouette analysis highlighted an improved performance of 31% for the clustering performed using  $k_{ADC} = 1$  with respect to the established approach that employs morphologic features derived from the DCE-MRI time-intensity curves and the ADC as an additional feature (Table 1).

#### 4. Discussion

As a general strategy, we have demonstrated a dissimilarity clustering based on multidimensional data derived from diffusion and perfusion MRI. Extension of the algorithm to additional data is straightforward, though the computational demand rises, and the similarity metric will likely need to incorporate further context-specific knowledge. As examination of tumor heterogeneity is carried out on a tumor by tumor basis, the data space can be restricted to

areas containing lesions already located, but not necessarily segmented. For the specific use of DCE and DW-MRI, the lower resolution of the DWI data presents an issue of partial volume effects that affects the clustering of small lesions, but this issue is not specific to any one characterization strategy.

The two free parameters of the protocol: number of clusters ( $K$ ), and relative weighting of the diffusion data ( $k_{ADC}$ ), warrant discussion as the present work provides only a starting approximation to their choice, and the values may well be pathology dependent. For an unsupervised classification as used herein, the number of clusters should follow the actual structure and separation of the data into natural groups.

For breast tumors such as ductal carcinoma, the reporting of DCE-MRI data is currently based on a three-way division, while DWI is binary between normal and abnormal. The three DCE curve types (a rise and fall, a rise to a plateau, and a steady rise) have established clinical utility in predicting tumor malignancy [30]. This is not to say however that only three subgroups are possible, nor that these subgroupings are predictive of treatment response, which is the motive for examining tumor heterogeneity. In fact, works such as [14] have demonstrated that as the

TABLE 1: Silhouette analysis scores describing cluster compactness and separation for the whole ROI and for each relevant region for the kinetic features and the multimodal lesion assessment (MMLA) methods (the higher the better).

| Method                     | Mean        | Central 1 | Central 2 | Periferic |
|----------------------------|-------------|-----------|-----------|-----------|
| Morphologic features       | <b>0.51</b> | 0.53      | 0.51      | 0.49      |
| Morphologic features + ADC | <b>0.47</b> | 0.49      | 0.48      | 0.44      |
| MMLA                       | <b>0.62</b> | 0.57      | 0.65      | 0.61      |
| MMLA + ADC, $k_{ADC} = 1$  | <b>0.62</b> | 0.57      | 0.64      | 0.62      |
| MMLA + ADC, $k_{ADC} = 3$  | <b>0.58</b> | 0.54      | 0.59      | 0.60      |
| MMLA + ADC, $k_{ADC} = 5$  | <b>0.57</b> | 0.56      | 0.58      | 0.58      |

temporal resolution increases, a higher number of curve archetypes can be naturally identified and can be used for classification of voxelwise perfusion curves.

We consider it noteworthy, therefore, that when  $K$  was reduced to just three or four groups, these were identifiable with the 3 enhancement patterns (or these three and non-enhancement) used in clinical practice for the assessment of the breast cancer. As well, the confines of the groups with DCE-MRI time-course patterns consistent with malignant and benign tumors coincided very closely with the tumor margin drawn by a radiologist. Increasing the  $K$  value showed the expected progressive splitting of these groups as  $K$  increased, with  $k_{ADC}$  providing a distinction in the way this splitting proceeded based on the relative weight given to the diffusion data. The benefits of increasing the number of clusters are evident for understanding the heterogeneity of the lesion and the distribution of voxels that share certain similarities; however, the increase of the number of clusters should go hand in hand with cluster and data analysis techniques in order to avoid false or meaningless divisions. The overall protocol would also benefit from an integrating methodology such as cluster ensembles, in order to combine the multiple base clusterings done with different  $k_{ADC}$  values into a unified consolidated clustering, reaching with this a consensus solution.

The primary criteria for noninvasive assessment of tumors based on DCE MRI involve three enhancement patterns (four including necrosis/nonenhancement). In the clinical data used for this study this assessment criteria have limited the validation to the visual interpretation of enhancement patterns based on the conventional interpretation of DCE curves, with a reader-dependent incorporation of ADC information. Ultimately, the envisaged application is in anticipating and evaluating treatment response. If tumor heterogeneity in terms of both perfusion and diffusion is to be encompassed, the conventional 3-way categorization may not be adequate or appropriate and indeed for other organs this rating is less common. We are now looking into robust methods for further validation of the processing pipeline that would enable a clinical exploitation of the multimodal analysis. Access to ground truth beyond radiological and biopsy evaluation is needed and likely requires voxelwise comparison of with histology of resections, a process that requires modifications to the surgical procedure that were not justified for this first demonstration of the method. Even were histology image data available, a significant task remains

in the spatially correlation of individual MRI voxels with the histological results in order to get the requisite voxel-scale validation.

## 5. Conclusions

In this paper, we presented a general methodology for heterogeneity quantification that integrates information from diffusion (an indicator of cellularity) and perfusion (reflecting blood volume, flow, and vascular permeability) MRI images and illustrated its use in application to ductal carcinoma. The demonstration illustrated that multimodal clustering leads to improved selectivity and yields a greater refinement of the segmentation of tissues within the lesion than the separate processing of the two modalities.

By demonstrating that statistically consistent subgroups can be defined within tumors based on a combination of DCE-MRI and DWI-MRI data, we have indicated a means for objectively segmenting tumors that can be used for larger studies to examine clinical impact. Moreover, the appearance of statistically distinct perfusion regions within the tumor at moderate and low ADC weightings that in turn have statistically distinct ADC distributions suggests there is a useable distinction present that is not capitalized upon in present clinical practice.

## References

- [1] V. Sharifi-Salamatian, B. Pesquet-Popescu, J. Simony-Lafontaine, and J. P. Rigaut, "Index for spatial heterogeneity in breast cancer," *Journal of Microscopy*, vol. 216, no. 2, pp. 110–122, 2004.
- [2] X. Yang and M. V. Knopp, "Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review," *Journal of Biomedicine and Biotechnology*, vol. 2011, Article ID 732848, 12 pages, 2011.
- [3] M. C. Lloyd, P. Allam-Nandyala, C. N. Purohit et al., "Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it?" *Journal of Pathology Informatics*, vol. 1, article 29, 2010.
- [4] M. V. Knopp, F. L. Giesel, H. Marcos, H. Von Tengg-Kobligk, and P. Choyke, "Dynamic contrast-enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples," *Topics in Magnetic Resonance Imaging*, vol. 12, no. 4, pp. 301–308, 2001.
- [5] M. Y. Su, Y. C. Cheung, J. P. Fruehauf et al., "Correlation of dynamic contrast enhancement MRI parameters with

- microvessel density and VEGF for assessment of angiogenesis in breast cancer,” *Journal of Magnetic Resonance Imaging*, vol. 18, no. 4, pp. 467–477, 2003.
- [6] P. Gibbs, G. P. Liney, M. D. Pickles, B. Zelhof, G. Rodrigues, and L. W. Turnbull, “Correlation of ADC and T2 measurements with cell density in prostate cancer at 3.0 Tesla,” *Investigative Radiology*, vol. 44, no. 9, pp. 572–576, 2009.
- [7] Y. Matsumoto, M. Kuroda, R. Matsuya et al., “In vitro experimental study of the relationship between the apparent diffusion coefficient and changes in cellularity and cell morphology,” *Oncology Reports*, vol. 22, no. 3, pp. 641–648, 2009.
- [8] M. D. Jenkinson, D. G. Du Plessis, T. S. Smith, A. R. Brodbelt, K. A. Joyce, and C. Walker, “Cellularity and apparent diffusion coefficient in oligodendroglial tumours characterized by genotype,” *Journal of Neuro-Oncology*, vol. 96, no. 3, pp. 385–392, 2010.
- [9] S. C. Partridge, W. B. DeMartini, B. F. Kurland, P. R. Eby, S. W. White, and C. D. Lehman, “Quantitative diffusion-weighted imaging as an adjunct to conventional breast MRI for improved positive predictive value,” *American Journal of Roentgenology*, vol. 193, no. 6, pp. 1716–1722, 2009.
- [10] T. E. Yankeelov, M. Lepage, A. Chakravarthy et al., “Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results,” *Magnetic Resonance Imaging*, vol. 25, no. 1, pp. 1–13, 2007.
- [11] L. R. Jensen, B. Garzon, M. G. Heldahl et al., “Diffusion-weighted and dynamic contrast-enhanced MRI in evaluation of early treatment effects during neoadjuvant chemotherapy in breast cancer patients,” *Journal of Magnetic Resonance Imaging*, vol. 34, no. 5, pp. 1099–1109, 2011.
- [12] H. Yabuuchi, Y. Matsuo, T. Okafuji et al., “Enhanced mass on contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images,” *Journal of Magnetic Resonance Imaging*, vol. 28, no. 5, pp. 1157–1165, 2008.
- [13] Z. Yili, H. Xiaoyan, D. Hongwen et al., “The value of diffusion-weighted imaging in assessing the ADC changes of tissues adjacent to breast carcinoma,” *BMC Cancer*, vol. 9, article 18, 2009.
- [14] C. Lavini, M. C. de Jonge, M. G. H. van de Sande, P. P. Tak, A. J. Nederveen, and M. Maas, “Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system,” *Magnetic Resonance Imaging*, vol. 25, no. 5, pp. 604–612, 2007.
- [15] C. K. Kuhl, P. Mielcareck, S. Klaschik et al., “Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions?” *Radiology*, vol. 211, no. 1, pp. 101–110, 1999.
- [16] H. Degani, V. Gusic, D. Weinstein, S. Fields, and S. Strano, “Mapping pathophysiological features of breast tumors by MRI at high spatial resolution,” *Nature Medicine*, vol. 3, no. 7, pp. 780–782, 1997.
- [17] Y. Gal, A. Mehnert, A. Bradley, D. Kennedy, and S. Crozier, “Feature and classifier selection for automatic classification of lesions in dynamic contrast-enhanced MRI of the breast,” in *Proceedings of the Digital Image Computing: Techniques and Applications (DICTA ’09)*, pp. 132–139, December 2009.
- [18] E. Pekalska, P. Paclik, and R. P. W. Duin, “A generalized kernel approach to dissimilarity-based classification,” *Journal of Machine Learning Research*, vol. 2, article 11, 2002.
- [19] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “Elastix: a toolbox for intensity-based medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.
- [20] D. Rueckert, “Nonrigid registration using free-form deformations: application to breast mr images,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [21] Y. Guo, R. Sivaramakrishna, C. C. Lu, J. S. Suri, and S. Laxminarayan, “Breast image registration techniques: a survey,” *Medical and Biological Engineering and Computing*, vol. 44, no. 1-2, pp. 15–26, 2006.
- [22] E. Pekalska and R. P. W. Duin, *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, vol. 64, Word Scientific, 2005.
- [23] A. D. Chouakria and P. N. Nagabhushan, “Adaptive dissimilarity index for measuring time series proximity,” *Advances in Data Analysis and Classification*, vol. 1, no. 1, pp. 5–21, 2007.
- [24] C. A. Mendez, F. Pizzorni Ferrarese, P. Summers et al., “Multimodal MRI-based tissue classification in breast ductal carcinoma,” in *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging (ISBI ’12)*, pp. 142–145, 2012.
- [25] A. Douzal-Chouakria and C. Amblard, “Classification trees for time series,” *Pattern Recognition*, vol. 45, no. 3, pp. 1076–1091, 2012.
- [26] R. Xu and D. Wunsch, “Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [27] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, “Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI,” *Medical Physics*, vol. 33, no. 8, pp. 2878–2887, 2006.
- [28] J. Barcelo, J. C. Vilanova, and A. Luna, “DWI of the breast,” in *Diffusion MRI Outside the Brain*, pp. 203–230, 2012.
- [29] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [30] American College of Radiology, *Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas)*, 2003.