NeuroImage 46 (2009) 642–651

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Fully automated classification of HARDI *in vivo* data using a support vector machine

S. Schnell [a,*], D. Saur [b], B.W. Kreher [a], J. Hennig [a], H. Burkhardt [c], V.G. Kiselev [a]

[a] *Medical Physics, Department of Diagnostic Radiology, University Medical Center Freiburg, Hugstetter Str. 55, D-79106 Freiburg, Germany*
[b] *Freiburg Brain Imaging Center, Department of Neurology, University Medical Center Freiburg, Germany*
[c] *Chair in Pattern Recognition and Image Processing, Institute of Computer Science, Albert-Ludwigs-University Freiburg, Germany*

ARTICLE INFO

Article history:
Received 20 August 2008
Revised 3 February 2009
Accepted 1 March 2009
Available online 12 March 2009

Keywords:
Spherical harmonics
Rotation invariant
Support vector machine
Classification
HARDI
Crossing fibre bundles

ABSTRACT

The purpose of this study is the classification of high angular resolution diffusion imaging (HARDI) *in vivo* data using a model-free approach. This is achieved by using a Support Vector Machine (SVM) algorithm taken from the field of supervised statistical learning. Six classes of image components are determined: grey matter, parallel neuronal fibre bundles in white matter, crossing neuronal fibre bundles in white matter, partial volume between white and grey matter, background noise and cerebrospinal fluid. The SVM requires properties derived from the data as input, the so called feature vector, which should be rotation invariant. For our application we derive such a description from the spherical harmonic decomposition of the HARDI signal. With this information the SVM is trained in order to find the function for separating the classes. The SVM is systematically tested with simulated data and then applied to six *in vivo* data sets. This new approach is data-driven and enables fully automatic HARDI data segmentation without employing a T1 MPRAGE scan and subjective expert intervention. This was demonstrated on five test *in vivo* data sets giving robust results. The segmentation results could be used as *a priori* knowledge for increasing the performance of fibre tracking as well as for other clinical and diagnostic applications of diffusion weighted imaging (DWI).

Introduction

Diffusion weighted MRI (DWI) and in particular measurements of diffusion anisotropy provides biologically relevant information about the tissue microstructure. A special focus of interest for research and clinical application of DWI is the investigation of the brain white matter (WM) structure. Such measurements allow the reconstruction of the neuronal fibre architecture in WM, the visualisation of fibre tracks and the examination of morphological connectivity between different cortical and sub-cortical regions. Data acquisition is typically performed using the so called High Angular Resolution Diffusion Imaging (HARDI) approach introduced by Tuch et al. (1999). This method consists of the application of diffusion encoding (DE) gradients in a large number of non-collinear directions. With, for instance, 64 DE gradient directions the spatially non-Gaussian diffusion behaviour of water in white matter regions with heterogeneous fibre orientations can be resolved. Therefore HARDI evolved to be the basis for many post-processing approaches for resolving the spatial structure of neuronal fibre bundles in WM. Specifically, it would be advantageous to distinguish between parallel (PF) and crossing (CF) fibre bundles.

The existing methods for inferring multiple fibre bundle populations from diffusion data can be classified into two groups (Behrens et al., 2007): model-dependent methods for the estimation of the underlying diffusion profile or model-free methods based on the inherent structure of the diffusion profile itself. The generic model-based method is Diffusion Tensor Imaging (DTI) (Basser et al., 1994), which was the first method used as a basis for the reconstruction of neuronal fibres, i.e. fibre tracking. The diffusion tensor (DT) represents the apparent diffusion coefficient (ADC) and can be explained as the averaging of all water spins in a voxel when applying DE gradients in several spatial directions. From the DT, anisotropy measures, such as the fractional anisotropy (FA), can be derived. The main drawback of DTI is that it can only reveal a single fibre orientation in each voxel and fails in voxels containing complex tissue architecture with more than one significant fibre orientation. One segmentation procedure based on the DT model applies a supervised clustering procedure with a collection of DTI metrics in regions of interests for the segmentation of GM, WM and CSF (Hasan and Narayana, 2006). In this method, the contrast of FA maps between CSF, WM and GM was used, based on the “principal diffusivity indices”. The CSF was segmented using its high diffusivity and low anisotropy properties. However, since this method is based on DTI, no further classification of the WM subclasses PF and CF was possible.

An approach that combines model-dependent and model-free methods for the differentiation of parallel and crossing fibre bundles based on HARDI and DTI was described by Kreher et al., (2005). In this approach a multi-diffusion tensor model was introduced, which contains one anisotropic and one isotropic diffusion tensor in order to model the tissue structures. In each voxel it is decided separately which of the two models is more appropriate for describing the

* Corresponding author. Fax: +49 761 270 3832.
E-mail address: susanne.schnell@uniklinik-freiburg.de (S. Schnell).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2009.03.003

underlying diffusion and therefore more suitable for the detection of crossing fibre bundles.

The first model-free method which used spherical harmonics for the description of the diffusion profile acquired with HARDI data was reported by Frank (2002). Spherical harmonics are functions similar to Fourier expansions, but described in spherical polar coordinates (polar angle θ and the azimuth angle φ). Every function that takes as its arguments the directions θ and φ can be expanded into spherical harmonics. A function of the signal S can be described with spherical harmonics as follows (Webster and Szego, 1930; Arfken and Weber, 1985):

$$S(\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} a_n^m Y_n^m(\theta,\varphi) \quad (1)$$

where Y is the spherical harmonic of order n (all integer $n \geq 0$), and m the azimuthal separation constant or degree (all integer m, $|m| \leq n$). The coefficients a_n^m are expressed as:

$$a_n^m = \int_{j=0}^{2\pi} \int_{\theta=0}^{\pi} Y_n^{m*}(\theta,\phi) S(\theta,\phi) \sin\theta d\theta d\varphi \quad (2)$$

with Y^* being the complex conjugate of Y. The expansion of Eq. (1) can be terminated at some n. The higher the order n the more complex the deviation from the spherical shape ($n=0$), which can be described. Then, however, a necessary condition of the sampling theorem requires that more directions be measured (see section Theory and methods and Yeo, 2005).

In Frank's (2002) approach isotropic diffusion occurring in water or CSF is described by zero order spherical harmonics, diffusion along parallel fibres by second order spherical harmonics, and diffusion in the multiple fibre case is approximated by the fourth order. The odd orders describe asymmetric components and therefore represent imaging artefacts and noise. By using a high order versus low order ratio of the spherical harmonic coefficients, Frank presented a method for differentiation between PF and CF, which is however subject to limitations. The results could include possible misclassification, especially for WM regions containing multiple crossings. These regions appeared like isotropic voxels similar to GM voxels. Differentiation between GM, CSF and background noise was thus not feasible with exclusive use of the spherical harmonic description. Descoteaux et al. (2006) extended the model in order to distinguish between isotropic, one-fibre and multi-fibre diffusion. This procedure is very promising, but automatic full image segmentation was not possible, since CF was still often misclassified as GM or noise. Alexander et al. (2002) described a method for the modelling and detection of non-Gaussian diffusion profiles also using spherical harmonics, but up to an order of eight, providing a sequence of models of increasing complexity. A statistical test was performed in order to find the simplest of the models which adequately described the data. This method was applied in a human experiment and seemed to classify isotropic (GM) and anisotropic Gaussian (WM) regions correctly as order zero and order two, respectively. It was found that on average five percent of profiles in voxels within the brain were classified as order four or above (anisotropic non-Gaussian), which, from our understanding of anatomy, would be too low a percentage. The method was validated by characterising its performance using synthetic data. It was not described how accurately GM was differentiated from CF.

Behrens et al. (2007) reviewed several recent model-free techniques. In the data shown (HARDI data in 60 DE directions) a third fibre bundle orientation could not be detected. The authors supposed that a detection of more than two orientations would be possible if more diffusion directions were to be acquired at a higher b-value. Simulations, which were performed in (Behrens et al. 2007) suggest that in order to resolve a three fibre bundle orthogonal system robustly, data with b-values above 4000 s/mm^2 has to be acquired. As will be shown below in the Theory and methods section, it is necessary to acquire more than 60 DE directions in order to fulfill the sampling theorem for spherical harmonics of order four and above (Yeo, 2005). In addition, the review by Alexander (2005) showed with noisy data synthesised from isotropic test functions that most methods generate spurious angular structure. This may explain why strong angular structures are incorrectly detected even in many GM and CSF voxels.

However, for a realistic tissue description, the existing models are rather complex and often include ill-defined parameters not adequately supported by the measurement data. All presented methods, including the model-free methods, showed that full image segmentation of microstructures and of image background was not possible. A differentiation between voxels containing PF or CF, or the differentiation between GM, CF and background noise is difficult, since this information is usually derived from some measure of diffusion anisotropy. Many publications outline methods which show potential for performing this differentiation, but so far an evaluation of their methods for fibre crossings has not been reported.

Based on the state-of-the-art described above we suggest a new data-driven analysis of multi-directional diffusion weighted MRI data, which may provide unique fingerprints for different types of tissue and image components. In addition to using a model-free approach, we employ methods developed in the field of pattern recognition. In the present case we attempt classification of six different classes: grey matter (GM), the two white matter (WM) subclasses: CF and PF, partial volume (a mixture between GM and WM), as well as cerebrospinal fluid (CSF), background noise and image artefacts (hereafter referred to as noise). First, the underlying diffusion profile per voxel of the HARDI data is described using the rotational invariants of the spherical harmonic decomposition. Then a Support Vector Machine (SVM), a computer algorithm for statistical learning which has already demonstrated robust performance in other applications (Nattkemper 2004; Quddus et al., 2005) is used for classification (Cristianini and John, 2000). The SVM is trained with the labelled image features in order to find the function for separating the classes. Afterwards the SVM is systematically tested with simulated data and then applied to six *in vivo* data sets.

Theory and methods

The support vector machine as a classifier

The field of pattern recognition is a sub-topic of machine learning. It is either based on *a priori* knowledge or on statistical information extracted from the patterns, meaning that some pattern in raw data is classified by performing some action based on some property or feature of the data. Therefore, pattern recognition can be considered as a form of data classification.

A complete pattern recognition system consists of a sensor that gathers the observations to be classified, a feature extraction mechanism, which computes numeric information from the observations, and a classifier, which does the actual job of classifying observations, relying on the extracted features. One classification method is supervised learning, which is a machine learning technique for creating a function from training data. The training data consists of pairs of input data (the feature vectors) and desired outputs (the labels). The output of the function can predict a class label of the input data. The task of the supervised learner is to predict the value of the function for input data after having seen a number of training examples. This means that for correct separation between the classes the learner has to be able to generalise from the presented data to unseen situations or data (Burges, 1998).

An example of a classifier is the support vector machine (SVM), which was introduced by Cortes and Vapnik (1995) and was found to yield good results for the problem presented in this paper. The SVM