Manifold-driven Grouping of Skeletal Muscle Fibers

Radhouene Neji, Jean-Francois Deux, Ahmed Besbes, Nikos Komodakis, Georg Langs, Mezri Maatouk, Alain Rahmouni, Guillaume Bassez, Gilles Fleury, Nikolaos Paragios

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Manifold-driven Grouping of Skeletal Muscle Fibers

Radhouène Neji — Jean-François Deux — Ahmed Besbes — Nikos Komodakis — Georg Langs — Mezri Maatouk — Alain Rahmouni — Guillaume Bassez — Gilles Fleury — Nikos Paragios

N° 6825

February 2009

Thème BIO

apport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6825--FR+ENG

Manifold-driven Grouping of Skeletal Muscle Fibers

Radhouène Neji* † , Jean-François Deux‡ , Ahmed Besbes* , Nikos Komodakis§ , Georg Langs¶ , Mezri Maatouk ‡ , Alain Rahmouni‡ , Guillaume Bassez‡ , Gilles Fleury† , Nikos Paragios*

Thème BIO — Systèmes biologiques
Projet Galen

Rapport de recherche n° 6825 — February 2009 — 19 pages

Abstract: In this report, we present a manifold clustering method for the classification of fibers obtained from diffusion tensor images (DTI) of the human skeletal muscle. To this end, we propose the use of angular Hilbertian metrics between multivariate normal distributions to define a family of distances between tensors that we generalize to fibers. The obtained metrics between fiber tracts encompasses both diffusion and localization information. As far as clustering is concerned, we use two methods. The first approach is based on diffusion maps and k-means clustering in the spectral embedding space. The second approach uses a linear programming formulation of prototype-based clustering. This formulation allows for classification over manifolds without the necessity to embed the data in low dimensional spaces and determines automatically the number of clusters. The experimental validation of the proposed framework is done using a manually annotated significant dataset of DTI of the calf muscle for healthy and diseased subjects.

Key-words: DTI, Diffusion tensor, Fiber, Hilbertian Metrics, Linear Programming, Diffusion Maps, Clustering, Human skeletal muscle

* Radhouène Neji, Ahmed Besbes and Nikos Paragios are affiliated to Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France and to Equipe GALEN, INRIA Saclay - Île-de-France, Orsay, France

† Radhouène Neji and Gilles Fleury are affiliated to Département Signaux et Systèmes Électroniques, Ecole Supérieure d'Electricité, Gif-sur-Yvette, France

‡ Jean-François Deux, Mezri Maatouk, Alain Rahmouni and Guillaume Bassez are affiliated to Centre Hospitalier Universitaire Henri Mondor, Créteil, France

§ Nikos Komodakis is affiliated to Computer Science Department, University of Crete, Crete, Greece

¶ Georg Langs is affiliated to CIR lab, Department of Radiology, Medical University of Vienna, Vienna, Austria

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 — Télécopie : +33 1 60 19 66 08

Groupement sur variétés des fibres du muscle squelettique

Résumé : Dans ce rapport, nous présentons une méthode de groupement sur variétés afin de classifier les fibres obtenues à partir d'images d'IRM de diffusion du muscle squelettique. La définition de métriques angulaires hilbertiennes entre des distributions normales multivariées permet d'obtenir une famille de distances entre tenseurs de diffusion généralisable aisément aux fibres. Les métriques obtenues tiennent en compte à la fois de l'information spatiale et de l'information de diffusion. Afin de grouper les fibres, nous utilisons deux approches. La première consiste à utiliser les plongements par diffusion et l'algorithme des k-moyennes dans l'espace du plongement. La deuxième se base sur la programmation linéaire et permet d'effectuer une classification sur variétés sans recourir à des plongements dans des espaces de basse dimension et sans préciser au préalable le nombre de groupes. La validation expérimentale est effectuée sur un ensemble significatif d'images de diffusion du mollet de sujets sains et malades qui ont été segmentées manuellement par un expert.

Mots-clés : IRM de diffusion, Tenseur de diffusion, Fibre, Métriques Hilbertiennes, Groupement, Programmation Linéaire, Plongements par Diffusion, Muscle squelettique

Contents

1 Introduction

Diffusion Tensor Imaging (DTI) has started to become more ubiquitous in other fields than brain white matter study [1]. Indeed, this modality has been used for other anatomical regions such as the tongue [2] and the human skeletal muscles [3]. The latter are of particular interest because they present an architecture of elongated myofibers with well known anatomy. Furthermore, the study of the effects of myopathies (neuromuscular diseases) on water diffusion in muscle tissues is essential to assess the possibility of the use of DTI in a diagnosis procedure and early detection of diseases. Since myopathies result in an atrophy and weakness of the muscle, we expect an alteration of the diffusion properties among diseased subjects. It is therefore important to cluster fiber tracts for local statistical analysis of diffusion information.

DTI previous studies of the human skeletal muscle [4, 5] provided a comparative study between subjects and different muscle regions of scalar values derived from tensors like trace, fractional anisotropy, etc. They also evaluated experimentally the physiological cross-sectional area (PCSA), which is an important measure of muscle architecture since it is related to the maximum muscle force. However little emphasis was put on muscle segmentation in comparison with brain white matter, where several approaches were proposed. The use of graph theory and manifold learning has been extensively explored in the previous literature. For instance, in [6] the distribution of points along each fiber tract is considered to be Gaussian, which allows to derive a Euclidean distance between each pair of fibers. Fiber bundling is done using a normalized cut. In [7], the affinity between fibers is based on the symmetrized Hausdorff distance and spectral clustering is achieved using an eigen-analysis of the affinity matrix and k-means in the embedding space. The method presented in [8] relies on Laplacian Eigenmaps and similarity between fibers is determined using their end points. In [9], the authors construct a graph-based distance between fiber tracts where both local and global dissimilarities are taken into account. The considered distance is then incorporated in a Locally Linear Embedding framework and clustering is done using k-means. Curve modeling has attracted attention and was handled in [10] by defining a spatial similarity measure between curves and using the Expectation-Maximization algorithm for clustering. The method proposed in [11] considers the simultaneous use of medoid-shift clustering and isomap-like manifold learning and proposed to include prior knowledge in the segmentation process using a white matter fiber atlas. Mean-shift was also used in [12] where each fiber is first embedded in a high dimensional space using its sequence of points, and kernels with variable bandwidths are considered in the mean-shift algorithm. More recently, fibers were represented in [13] using their differential geometry and frame transportation and a consistency measure was used for clustering. Another class of methods suggested to circumvent the limitation of unsupervised clustering where the obtained segmentation may not correspond to anatomical knowledge. They opt for supervised algorithms that try to achieve a clustering consistent with a predefined atlas. Expert manual labeling of the fibers for one subject provides an atlas in [14]. This is followed by the registration of B0 images and a hierarchical classification of fibers where the B-spline coefficients of the curves are considered to measure curve similarity. The method proposed in [7] is further extended in

[15] by means of a Nystrom approximation of the out-of-sample extension of the spectral embedding to build an atlas of fibers.

In [16], we proposed a kernel between tensors primarily, generalized it to fiber tracts and used k-means clustering after kernel PCA and Isomap embedding. We develop the viewpoint that was proposed in [16] and build Hilbertian angular metrics between fibers. These are derived from their counterparts between tensors, providing a more general and much simpler formulation of the approach in [16]. Two approaches are studied for clustering purposes. The first approach resorts to k-means in the diffusion maps embedding space. Note that diffusion maps were used for Orientation Distribution Function (ODF) segmentation in Q-ball images in [17], where spatial coherence was imposed using the Markovian relaxation of the affinity matrix. However the fiber domain provides no straightforward spatial neighborhood relationships like those given by the nearest neighbors in the 3D image grid. We show that the proposed metrics impose spatial coherence in the fiber domain while taking into account the information provided by the tensor field.

The second clustering approach is motivated by the limits of manifold embedding methods. Indeed, the use of embeddings and common clustering techniques like k-means requires to choose the dimension of the embedding and the number of clusters. It would be preferable to obtain the number of clusters as a result of the clustering algorithm, especially when the inter-patient variability (which is rather important for skeletal muscles) may require the use of different number of clusters across patients. Moreover, selecting the embedding dimension is an issue since a too low dimension will result in information loss and a too high dimension will include an important dispersion in the data. Furthermore, clustering on the manifold directly is a tricky issue since one has to compute intrinsic means on submanifolds where an explicit expression of geodesic distances is not necessarily available. Another issue is the sensitivity of methods like k-means to initialization. Therefore, we propose a method that performs manifold clustering of fibers without resorting to manifold embeddings or computations of intrinsic means. It is based on linear programming (LP) and uses the geodesic distances in a way similar to [11] from the fibers to a reduced set of landmark fibers to perform the clustering. Unlike k-means, the algorithm provides automatically the number of clusters, is not sensitive to initialization and the class centers are chosen as examplars from the dataset.

The remainder of the report is organized as follows: in section 2, we discuss and derive the family of Hilbertian angular metrics between tensors and propose their extension to fiber tracts. In section 3, we review the diffusion map principles. In section 4, we present the LP-based clustering method and develop the geodesic clustering costs. Section 5 is dedicated to the experimental results and we discuss the perspectives of this work in section 6.

2 From Metrics on Tensors to Metrics on Fibers

In this section, we define a family of Hilbertian metrics over the space of fibers. The starting point is to consider angular distances between tensors based on Gaussian probability

densities and generalize these distances to the fiber domain. We build upon the work in [16], providing a more general theoretical approach to the problem and a simpler derivation.

2.1 Multivariate Normals: a Subset of the Exponential Distributions Family

The structure of the set of multivariate normal distributions $\mathcal{M}$ as a statistical manifold endowed with the Fisher information geometry was discussed in [18], where a closed-form solution of the geodesic distance [19] over this manifold is available for the particular case of Gaussian distributions with common mean. Here we view the multivariate normal distributions as a subset of the exponential distributions family. Let us consider a normal probability density p. In this context, given the exponential decay of the distribution, it is interesting to notice that not only p is an element of the Hilbert space L^2 of square integrable functions but any power p^α, with α a strictly positive real number is also square integrable. This motivates the use of normalized probability product kernels [20] to define a family of angular similarities between multivariate normal distributions. Indeed, considering two elements p_1 and p_2 of $\mathcal{M}$ and $\alpha \in \mathbb{R}_+^*$, we can define the following similarity $C_\alpha(p_1, p_2)$ between p_1 and p_2 as follows:

$$C_\alpha(p_1, p_2) = \frac{\int p_1(\mathbf{x})^\alpha p_2(\mathbf{x})^\alpha d\mathbf{x}}{\sqrt{\int p_1(\mathbf{x})^{2\alpha} d\mathbf{x}} \sqrt{\int p_2(\mathbf{x})^{2\alpha} d\mathbf{x}}} \tag{1}$$

C_α is simply the normalized L^2 inner product between p_1^α and p_2^α. It is therefore the cosine of the angle between p_1^α and p_2^α. It defines a Mercer kernel over the space of multivariate normal distributions, i.e. for any subset $p_{1\ldots N}$ of $\mathcal{M}$, the Gram matrix G of C_α with entries $G_{ij} = C_\alpha(p_i, p_j)$ is semi-definite positive. The Mercer property allows the construction of a mapping ϕ_α associated with the kernel C_α that provides an embedding of $\mathcal{M}$ in the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_\alpha$ such that $C_\alpha(p_1, p_2) =< \phi_\alpha(p_1), \phi_\alpha(p_2) >_{\mathcal{H}_\alpha}$, where $< ., . >_{\mathcal{H}_\alpha}$ is the inner product of $\mathcal{H}_\alpha$. This allows to have the following Hilbertian metric $d_{\alpha|\mathcal{H}_\alpha}$:

$$d_{\alpha|\mathcal{H}_\alpha}(p_1, p_2) = \sqrt{C_\alpha(p_1, p_1) - 2C_\alpha(p_1, p_2) + C_\alpha(p_2, p_2)} \tag{2}$$

Given that C_α is a normalized scalar product, i.e. $C_\alpha(p, p) = 1$, we obtain the following expression:

$$d_{\alpha|\mathcal{H}_\alpha}(p_1, p_2) = \sqrt{2 - 2C_\alpha(p_1, p_2)} \tag{3}$$

Note that C_α is generalization of the normalized Expected Likelihood kernel (for $\alpha = 1$) and of the Bhattacharya kernel (for $\alpha = \frac{1}{2}$) [20]. In the latter case, the distance $d_{\frac{1}{2}|\mathcal{H}_\alpha}$ is the Hellinger distance between probability distributions which was used in [21] to measure the similarity between 4th-order tensors for registration of HARDI data without the incorporation of spatial information. In the following subsection, we derive the closed-form expression of C_α and $d_{\alpha|\mathcal{H}_\alpha}$ for normal distributions that model a local diffusion process.

2.2 Explicit Derivation of the Angular Distances

Let us consider the Gaussian distribution p that models the motion distribution of water protons at a location $\mathbf{x}$ with a tensor $\mathbf{D}$. Given a diffusion time t, the probability of displacement from the position $\mathbf{x}$ to the position $\mathbf{y}$ is provided by the following equation:

$$p(\mathbf{y}|\mathbf{x}, t, \mathbf{D}) = \frac{1}{\sqrt{\det(\mathbf{D})(4\pi t)^3}} \exp(-\frac{(\mathbf{y}-\mathbf{x})^t \mathbf{D}^{-1}(\mathbf{y}-\mathbf{x})}{4t}) \tag{4}$$

We now consider two normal distributions p_1 and p_2 with parameters $(x_1, \mathbf{D}_1)$ and $(x_2, \mathbf{D}_2)$ respectively. Based on [20] and equation 4, we can see that C_α is the product of two terms:

$$C_\alpha(p_1, p_2) = C_\alpha^{tensor}(\mathbf{D}_1, \mathbf{D}_2) C_\alpha^{spatial}(p_1, p_2) \tag{5}$$

where

$$\begin{aligned} C_\alpha^{tensor}(\mathbf{D}_1, \mathbf{D}_2) &= 2\sqrt{2}\, \frac{\det(\mathbf{D}_1)^{\frac{1}{4}} \det(\mathbf{D}_2)^{\frac{1}{4}}}{\sqrt{\det(\mathbf{D}_1 + \mathbf{D}_2)}} \\ C_\alpha^{spatial}(p_1, p_2) &= \exp\left(-\frac{\alpha}{4t}(\mathbf{x}_1^t \mathbf{D}_1^{-1} \mathbf{x}_1 + \mathbf{x}_2^t \mathbf{D}_2^{-1} \mathbf{x}_2)\right) \times \\ &\quad \exp\left(\frac{\alpha}{4t}(\mathbf{D}_1^{-1}\mathbf{x}_1 + \mathbf{D}_2^{-1}\mathbf{x}_2)^t (\mathbf{D}_1^{-1} + \mathbf{D}_2^{-1})^{-1} (\mathbf{D}_1^{-1}\mathbf{x}_1 + \mathbf{D}_2^{-1}\mathbf{x}_2)\right) \end{aligned} \tag{6}$$

We notice that $C_\alpha^{spatial}$ has a much simpler expression. Indeed, using the following inversion properties

$$(\mathbf{D}_1 + \mathbf{D}_2)^{-1} = \mathbf{D}_1^{-1} - \mathbf{D}_1^{-1}(\mathbf{D}_1^{-1} + \mathbf{D}_2^{-1})^{-1}\mathbf{D}_1^{-1} \tag{7}$$

$$(\mathbf{D}_1 + \mathbf{D}_2)^{-1} = \mathbf{D}_2^{-1} - \mathbf{D}_2^{-1}(\mathbf{D}_1^{-1} + \mathbf{D}_2^{-1})^{-1}\mathbf{D}_2^{-1} \tag{8}$$

we obtain the following compact expression for $C_\alpha^{spatial}$:

$$C_\alpha^{spatial}(p_1, p_2) = \exp\left(-\frac{\alpha}{4t}(\mathbf{x}_1 - \mathbf{x}_2)^t (\mathbf{D}_1 + \mathbf{D}_2)^{-1} (\mathbf{x}_1 - \mathbf{x}_2)\right) \tag{9}$$

We can see that C^{tensor} is a tensor similarity term and is independent of the parameter α while $C_\alpha^{spatial}$ is a spatial connectivity term where appears the Mahalanobis distance between the locations $\mathbf{x}_1$ and $\mathbf{x}_2$ with respect to the sum of tensors $(\mathbf{D}_1 + \mathbf{D}_2)$. Therefore C_α takes into account the tensor affinity as well the spatial position. This is crucial since combination of spatial and diffusion information allows for a better modeling of the interactions between tensors and favors a generalization to the fiber domain, as will be discussed in the next subsection. The diffusion time t is important to weight the contribution of each term and $t \to \infty$ corresponds to the case where the spatial interaction is not taken into account. Furthermore, there is a striking similarity between the proposed family of measures since α appears as a scale parameter in the exponential function. Given the present formulation, we can conclude that changing the parameter α amounts to a rescaling of the diffusion time t. The derivation of the metrics $d_{\alpha|\mathcal{H}_\alpha}$ is handily done using equation 3.

In the next subsection, we show how the Mercer property of C_α allows the definition of angular similarities between fiber tracts.

2.3 Angular Similarities between Fibers

A fiber tract is obtained by following the principal directions of diffusion of the tensor field starting from an initial location. It is therefore natural to represent a fiber $\mathbf{F}$ as a sequence of Gaussian probability measures $(p_i)_{i=1\ldots N}$ where N is the number of points of the fiber. Every probability measure (p_i) has a pair of parameters $(\mathbf{x}_i, \mathbf{D}_i)$ where $\mathbf{x}_i$ is the spatial location and $\mathbf{D}_i$ is the tensor at $\mathbf{x}_i$ when the tensor field is supposed to be continuous. When considering the mapping ϕ_α of these measures in the RKHS $\mathcal{H}_\alpha$, we can represent $\mathbf{F}$ as a weighted average of $\phi_\alpha(p_i)_{i=1\ldots N}$, i.e. $\mathbf{F} = \sum_{i=1}^{N} w_i \phi_\alpha(p_i)$. A straightforward choice of weights is $\forall i, w_i = \frac{1}{N}$.

Let us consider a fibers $\mathbf{F}_1$ (resp. $\mathbf{F}_2$) represented using a set of probabilities $(p_i)^{(1)}_{i=1\ldots N_1}$ (resp. $(p_i)^{(2)}_{i=1\ldots N_2}$) and weights $w_i^{(1)}$ (resp. $w_i^{(2)}$). The angular similarity $\widehat{C}_\alpha$ between $\mathbf{F}_1$ and $\mathbf{F}_2$ points is defined as follows:

$$\widehat{C}_\alpha(\mathbf{F}_1, \mathbf{F}_2) = \frac{< \sum_{i=1}^{N_1} w_i^{(1)} \phi_\alpha(p_i^{(1)}), \sum_{j=1}^{N_2} w_j^{(2)} \phi_\alpha(p_j^{(2)}) >_{\mathcal{H}_\alpha}}{\left\| \sum_{i=1}^{N_1} w_i^{(1)} \phi_\alpha(p_i^{(1)}) \right\|_{\mathcal{H}_\alpha} \left\| \sum_{j=1}^{N_2} w_j^{(2)} \phi_\alpha(p_j^{(2)}) \right\|_{\mathcal{H}_\alpha}} \tag{10}$$

Using the bilinearity of the inner product $< ., . >_{\mathcal{H}_\alpha}$, we can express $\widehat{C}_\alpha$ using C_α :

$$\widehat{C}_\alpha(\mathbf{F}_1, \mathbf{F}_2) = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} w_i^{(1)} w_j^{(2)} C_\alpha(p_i^{(1)}, p_j^{(2)})}{\left\| \sum_{i=1}^{N_1} w_i^{(1)} \phi_\alpha(p_i^{(1)}) \right\|_{\mathcal{H}_\alpha} \left\| \sum_{j=1}^{N_2} w_j^{(2)} \phi_\alpha(p_j^{(2)}) \right\|_{\mathcal{H}_\alpha}} \tag{11}$$

where $\left\| \sum_{i=1}^{N_k} w_i^{(k)} \phi_\alpha(p_i^{(k)}) \right\|_{\mathcal{H}_\alpha} = \sqrt{\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} w_i^{(k)} w_j^{(k)} C_\alpha(p_i^{(k)}, p_j^{(k)})}$ for $k = \{1, 2\}$. Again the corresponding Hilbertian metric between fibers is derived in a similar way to equation 3.

In the following, we review the theory of diffusion-maps based clustering.

3 Diffusion Maps

Diffusion maps [22] are a spectral embedding of a set X of n nodes, for which *local geometries* are defined by a kernel $k : X \times X \to \mathbf{R}$. The kernel k satisfies $k(x, y) \geq 0$, and $k(x, y) = k(y, x)$. This kernel can be interpreted as an *affinity* between nodes. The resulting graph (an edge between x and y carries the weights $k(x, y)$) can be transformed into a reversible Markov chain by the socalled normalized graph Laplacian construction. In [23] a related construction was used to define a geometry on a set of observations, or trajectories. We define

$$s(x) = \sum_y k(x, y) \quad \text{and} \quad p(x, y) = \frac{k(x, y)}{s(x)}. \tag{12}$$

This new kernel is no longer symmetric, but it satisfies

$$\forall x,\ \sum_y p(x,y) = 1. \tag{13}$$

Therefore it can be interpreted as the probability of the transition from node x to node y in one time step, or a *transition kernel* of a Markov chain. It gives a diffusion operator

$$Pf(x) = \sum a(x,y) f(y) d\mu(y), \tag{14}$$

P is the Markov matrix with the entries $p(x,y)$ and its powers P^τ allow to propagate information through the Markov chain in τ timesteps according to the transition kernels. According to P^τ we can define a family of *diffusion distances* parameterized by τ on the set of nodes

$$D_\tau(x,y) = \sum_{l=1,\ldots,m} \frac{(p_\tau(x,l) - p_\tau(y,l))^2}{\pi(l)} \tag{15}$$

where $\pi(y) = s(x)/\sum_j s(y)$ is the probability of the node x in the unique stationary distribution (the uniqueness is fulfilled if the graph is connected). D_τ is an L^2 distance between the posterior distributions of reaching x or y from all points l in the graph. It captures the connectivity in the Markov chain, summing over all possible paths from x to y. It is low if there is a large number of paths of length τ with high transition probabilities between the nodes x and y.

The operator P defines a geometry which can be mapped to an Euclidean geometry by an eigenvalue decomposition of P. The latter results in a sequence of eigenvalues $\lambda_1, \lambda_2 \ldots$ and corresponding eigenfunctions $\Psi_1, \Psi_2, \ldots$ that fulfill $P\Psi_i = \lambda_i \Psi_i$. The diffusion map after τ timesteps $\mathbf{\Psi}_\tau : X \to \mathbf{R}^w$ embeds each node $i = 1, \ldots, n$ in the Markov chain into a w dimensional Euclidean space where the clustering of the data points can be done using k-means

$$i \mapsto \mathbf{\Psi}_\tau(i) \triangleq \begin{pmatrix} \lambda_1^\tau \Psi_1(i) \\ \lambda_2^\tau \Psi_2(i) \\ \vdots \\ \lambda_w^\tau \Psi_w(i) \end{pmatrix} \tag{16}$$

In this space, the Euclidean distance reflects the distances (parameterized by τ) defined by the Diffusion distance D_τ.

$$\|\mathbf{\Psi}_\tau(i) - \mathbf{\Psi}_\tau(j)\| = D_\tau(i,j). \tag{17}$$

Note that a common choice for the kernel $k(.,.)$ is the Gaussian kernel, i.e. $k(x,y) = \exp\left(-\frac{d^2(x,y)}{2\sigma^2}\right)$, where d is a distance over the set X and σ a scale factor. In our case, d corresponds to the fiber metric defined in section 2.

In the next section, we introduce the LP-based manifold clustering where there is no need to embed the data in low-dimensional spaces or to precise beforehand the number of clusters.

4 Manifold Clustering via Linear Programming

Clustering refers to the process of organizing a set of objects into groups such that the members of each group are as similar to each other as possible. A common way of tackling this problem is to formulate it as the following optimization task: given a set of objects $\mathcal{V} = \{p_1, \ldots, p_n\}$, endowed with a distance function $d(\cdot, \cdot)$ that measures dissimilarity between objects, the goal of clustering is to choose K objects from $\mathcal{V}$, say, $\{q_1, \ldots, q_K\}$ (these will be referred to as cluster centers hereafter) such that the obtained sum of distances between each object and its nearest center is minimized, or:

$$\min_{q_1,\ldots,q_K \in \mathcal{V}} \sum_{p \in \mathcal{V}} \min_i d(p, q_i) \ . \tag{18}$$

An important drawback of the above formulation is that it requires the number of clusters K to be provided beforehand, which is problematic as this number is very often not known in advance. Note that a wrong value for K may have a very negative effect on the final outcome. One would thus prefer K to be automatically estimated by the algorithm as a byproduct of the optimization process. To address this issue, we will let K be a variable here, and, instead of (18), we will use the following modified objective function, which additionally assigns a penalty $g(q_i)$ to each one of the chosen cluster centers q_i:

$$\min_K \min_{q_1,\ldots,q_K \in \mathcal{V}} \left(\sum_{p \in \mathcal{V}} \min_i d(p, q_i) + \sum_{i=1}^{K} g(q_i) \right) \ . \tag{19}$$

But, even if K is known, another serious drawback of many of the existing optimization-based techniques for clustering is that they are particularly sensitive to initialization and thus may get easily trapped in bad local minima. For instance, K-means (one of the most commonly used clustering methods) is doomed to fail if its initial cluster centers happen not to be near the actual cluster centers. To deal with that, here we will rely on a recently proposed clustering algorithm [24], which has been shown to yield approximately optimal solutions to the NP-hard problem (19). This algorithm relies on reformulating (19) as an equivalent integer program, whose LP-relaxation (denoted as PRIMAL hereafter) has the following form:

$$\text{PRIMAL} \equiv \min_{\mathbf{x}} \sum_{p,q \in \mathcal{V}, p \neq q} d(p, q) x_{pq} + \sum_{q \in \mathcal{V}} g(q) x_{qq} \tag{20}$$

$$\text{s.t.} \sum\nolimits_{q \in \mathcal{V}} x_{pq} = 1, \ x_{pq} \leq x_{qq}, \ x_{pq} \geq 0 \tag{21}$$

If constraints $x_{pq} \geq 0$ are replaced with $x_{pq} \in \{0, 1\}$, then the resulting integer program is equivalent to clustering problem (19). In this case, each binary variable x_{pq} with $p \neq q$ indicates whether object p has been assigned to cluster center q or not, while binary variable x_{qq} indicates whether object q has been chosen as a cluster center or not. Constraints

$\sum_{q\in\mathcal{V}} x_{pq} = 1$ simply express the fact that each object must be assigned to exactly one center, while constraints $x_{pq} \leq x_{qq}$ require that if p has been assigned to q then object q must obviously be chosen as a center. The most crucial issue for tackling this integer LP is setting the variables x_{qq} correctly, *i.e*, deciding which objects will be chosen as centers. To this end, the so-called *stability* of an object has been introduced in [24]. This is a measure which, intuitively, tries to quantitatively answer the following question: how much does one need to further penalize an object to ensure that it will never be selected as an optimal cluster center? For having a practical algorithm, an efficient way of estimating object stabilities is required. It turns out that this can indeed be done very fast by moving to the dual domain and appropriately updating a solution of a dual relaxation to PRIMAL. Since each dual cost provides a lower bound to the cost of the optimal clustering, an additional advantage of working in the dual domain is the ability to avoid bad local minima. We refer the reader to [24] for more details.

We now discuss the case where the objects lie on a manifold. This implies the use of the geodesic distance as a similarity measure. Ideally this distance should correspond to the pairwise cost $d(p, q)$ for $p \neq q$ in the linear programming formulation proposed in equation 20. As proposed in [11], a first possible choice is to compute the geodesic distances between all the pairs of points using the Dijkstra algorithm on a k-NN graph, as is done in the Isomap algorithm. The shortest path is found using a local approximation of the geodesic distance, for example the angular Hilbertian metric between fibers derived in section 2. The pairwise cost $d(p, q)$ is set to $d(p, q) = d_g(p, q)$ where d_g is the corresponding geodesic distance. Instead, inspired by the landmark Isomap algorithm [25], we can compute the geodesic distances from all the data points to a reduced set of randomly selected landmarks. This will reduce the computational load that a full computation of the geodesic distances between every pair of data points would entail. Let $(l_m)_{m=1\ldots n_l}$ be a set of such chosen n_l landmarks. We would like to replace $d_g(p, q)$ by a reasonable approximation. Given that the geodesic distance between two points is the length of the shortest path linking these points, we note the following $\forall m \in [1 \ldots n_l]$, $|d_g(p, l_m) - d_g(q, l_m)| \leq d_g(p, q) \leq d_g(p, l_m) + d_g(q, l_m)$, which implies

$$\sup_m |d_g(p, l_m) - d_g(q, l_m)| \leq d_g(p, q) \leq \inf_m (d_g(p, l_m) + d_g(q, l_m)) \tag{22}$$

This provides a lower bound and an upper bound to the cost $d_g(p, q)$ in the case where only the geodesic distances to some landmarks are computed. Note that in the particular case where p and q are landmarks $d_g(p, q) = \sup_m |d_g(p, l_m) - d_g(q, l_m)| = \inf_m (d_g(p, l_m) + d_g(q, l_m))$. On the other hand we can also note that

$$\inf_m (d_g(p, l_m) + d_g(q, l_m)) - 2\eta \leq \quad d_g(p, q) \tag{23}$$

$$d_g(p, q) \leq \quad \sup_m |d_g(p, l_m) - d_g(q, l_m)| + 2\eta \tag{24}$$

where $\eta = \inf_m \min(d_g(p, l_m), d_g(q, l_m))$. Therefore it makes sense to replace the cost $d_g(p, q)$ whether by its upper bound or its lower bound, since both approximate the cost up to 2η.

It is interesting to note in this setting that the lower bound is the L^∞ norm between the distance-to-landmarks representation of p and q. Indeed, let $\mathbf{u_p}$ (resp. $\mathbf{u}_q$) be the n_l-dimensional vector of geodesic distances of p (resp. q) to the landmarks

$$\mathbf{u}_p = [d_g(p, l_1), \ldots, d_g(p, l_{n_l})]^t \quad , \quad \mathbf{u}_q = [d_g(q, l_1), \ldots, d_g(q, l_{n_l})]^t \tag{25}$$

By definition, $\sup_m |d_g(p, l_m) - d_g(q, l_m)| = ||\mathbf{u}_p - \mathbf{u}_q||_\infty$. Thus the lower bound approximation has the advantage of defining a metric cost. Intuitively, for a number of landmarks sufficiently larger than the intrinsic dimension of the manifold, the distance vector representation will provide a good characterization of the points on the manifold.

5 Experimental Results

Thirty subjects (twenty healthy subjects and ten patients affected by myopathies) underwent a diffusion tensor imaging of the calf muscle using a 1.5 T MRI scanner. The following acquisition parameters were used : repetition time (TR)= 3600 ms, echo time (TE) = 70 ms, slice thickness = 7 mm and a b value of 700 $s.mm^{-2}$ with 12 gradient directions and 13 repetitions. The size of the obtained volumes is $64 \times 64 \times 20$ voxels with a voxel resolution of $3.125\, mm \times 3.125\, mm \times 7\, mm$. We acquired simultaneously high-resolution T1-weighted images that were segmented manually by an expert into seven muscle groups to provide the ground truth. To give an idea about the muscle architecture in the calf, we present in [Fig.2 (a)] a manual segmentation overlaid on an axial slice of a high-resolution T1-weighted image. The following muscle groups are considered: the soleus (SOL), lateral gastrocnemius (LG), medial gastrocnemius (MG), posterior tibialis (PT), anterior tibialis (AT), extensor digitorum longus (EDL), and the peroneus longus (PL). We manually delineated a region of interest (ROI) for fiber tracking [26] and the fibers with a majority of points lying outside of the ROI were discarded. The obtained manual segmentations of the T1-weighted images were downsampled to the resolution of the diffusion images and used to provide a ground-truth segmentation of the fiber tracts as follows: for each fiber, the number of voxels crossed by the fiber and belonging to each muscle group were counted. Then the fiber was assigned to the class with the majority vote. In our experiments we set the diffusion time to $t = 2\ 10^4$ and the parameter α in the fiber metric to $\alpha = 1$, both for diffusion maps and LP-based clustering. The weights w_i of each fiber $\mathbf{F}$ in (11) were chosen as the inverse of the number of points in $\mathbf{F}$.

5.1 Diffusion Maps Clustering

In order to quantitatively evaluate the diffusion maps clustering, we measure the dice overlap coefficient of the obtained segmentation with the ground-truth segmentation provided by the clinician. We tested the clustering method at two levels: for 7 and 10 clusters. The dimension of the diffusion maps embedding w was set to the number of clusters. The number of timesteps τ in (16) was set to $\tau = 1$. We tested two values for the scale parameter of the Gaussian kernel: $\sigma = \{0.5, 1\}$. The clustering in the embedding space is done using k-means

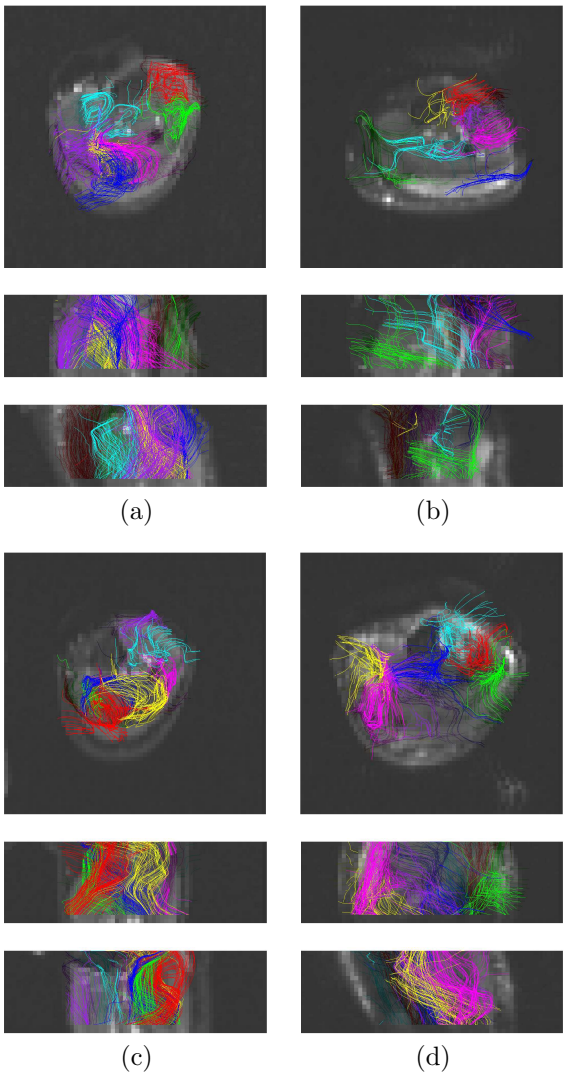
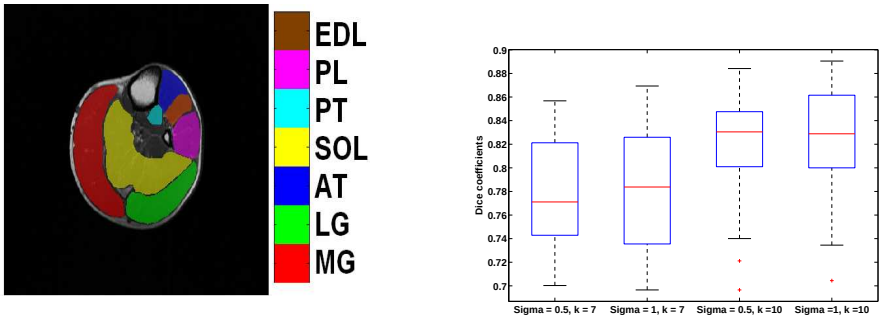


Figure 1: Axial, coronal and saggital views of diffusion maps fiber segmentation in 7 classes overlaid on diffusion-free (B0) images for (a), (c) two healthy subjects (b), (d) two diseased subjects

with 50 restarts and taking the clustering result with the least distortion. Distortion is computed as the ratio of intra-class and inter-class variances.



(a) (b)

Figure 2: (a) An axial slice of a high-resolution T1-weighted image showing a manual segmentation of the calf muscle in seven groups. (b) Boxplots of dice overlap coefficients for the thirty patients with different values of σ and k (the number of clusters). The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

In [Fig.2 (b)], we present the boxplots of the dice overlap coefficients for the thirty subjects, using the different values of σ and for 7 and 10 clusters. We can note that the quantitative results are rather satisfactory. For example, with the parameter σ set to $\sigma = 1$, we obtain a mean dice overlap coefficient of 0.78 (resp. 0.82) and a standard deviation of 0.05 (resp. 0.045) for 7 (resp. 10) classes . For a qualitative assessment, we show in [Fig.1] the obtained fiber classification in 7 clusters for two healthy and two diseased subjects. It is important to note in this setting that the diffusion images of the skeletal muscle are intrinsically noisier than brain images due to the short spin-spin (T2) relaxation time of the muscle tissue. Therefore the obtained fiber tracts are rather noisy, especially for diseased patients where the fat artifact is stronger. Despite the low quality of the tractography, our algorithm could still segment it in coherent fiber bundles.

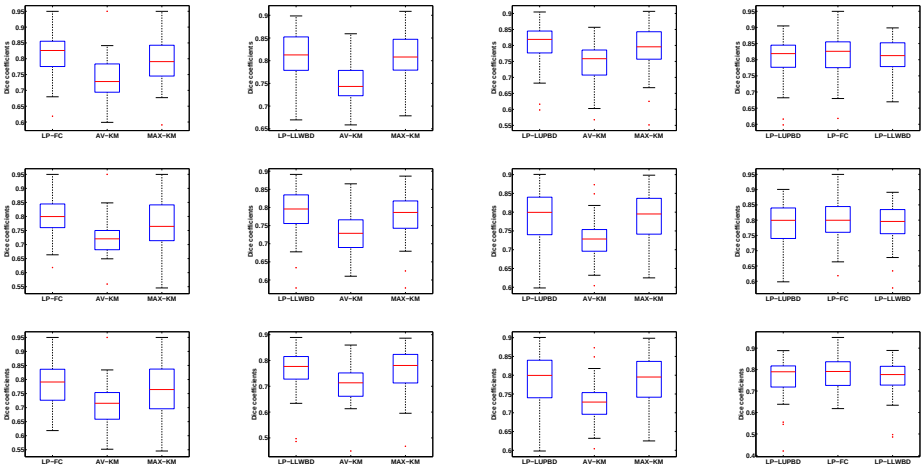
5.2 LP-based Clustering

We selected 30% of the fibers as landmarks and for the computation of the geodesic distances using the Dijkstra algorithm, we considered a k-NN graph where k set to $k = 12$. The cost $g(\mathbf{F})$ of choosing a fiber $\mathbf{F}$ as a class center in (20) was set to a constant $g = \beta\, \mu_{\frac{1}{2}}\left(d_g(\mathbf{F}_i, \mathbf{F}_j)_{i \neq j}\right)$ where $\mu_{\frac{1}{2}}$ is the statistical median. We tested the following values of β: $\{7, 10, 13\}$. For the sake of comparison, we evaluate also the performance of k-means clustering using the same metric and a manifold embedding. The dimensionality of the embedding is chosen to be the number of clusters obtained by our method, which is a common choice

in embedding-based approaches. The k-means algorithm is run 50 times and each time we compute the dice overlap of the clustering result with the ground-truth segmentation. We consider both the average dice coefficients over the restarts of the k-means algorithm and the dice coefficient of the clustering with the least distortion. We run the following experiments:

1. We compute all the geodesic distances between every pair of points and use them for linear programming clustering. We compare the obtained result with an Isomap embedding followed by k-means.

2. We compute the geodesic distances to a set of landmarks and use the lower (resp. upper) bound approximation for linear programming clustering. We compare the obtained result with a landmark-Isomap embedding followed by k-means.



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

Figure 3: Boxplots of dice overlap coefficients for the thirty patients. Each row corresponds to a value of β, from top to bottom β takes the following values 7 , 10 and 13. (a) LP clustering using full computation of distances (LP-FC), comparison is done with respect to the average score of k-means (AV-KM) and the score of the k-means clustering with least distortion (MAX-KM) after manifold embedding. (b) LP clustering using lower bound approximation (LP-LLWBD). (c) LP clustering using upper bound approximation (LP-LUPBD). (d) Comparison between LP-FC, LP-LLWBD and LP-LUPBD.

We provide in [Fig.3 (a), (b), (c)] the boxplots showing the distributions of the dice coefficients for the thirty patients using different values of β for our algorithm, compared with k-means after manifold embedding. We can note that linear programming clustering performs significantly better than the average score achieved by k-means both for a full and

landmark-based computation of the geodesic distances. Furthermore, it achieves results equivalent to the best k-means with an average dice coefficient of approximately 0.8 and in some cases it improves marginally the dice overlap. The advantage is that our result is reproducible, i.e. unlike k-means it is not sensitive to initialization. When comparing the three versions of linear programming clustering, we can see in [Fig.3 (d)] that the lower bound and upper bound approximations perform similarly apart from the case $\beta = 10$ where the lower bound approximation performed better, which may be explained by the metricity of the corresponding cost. The full computation yields slightly better results than the approximations. This corroborates the analysis provided is section 4. For qualitative evaluation, we show in [Fig.4(a)] (resp. [Fig.4(b)]) a clustering result obtained for a healthy (resp. diseased) subject for $\beta = 10$. Ground truth segmentation for the healthy patient is provided in [Fig.4(c)]. There are too few fibers in [Fig.4(b)] because the tractography fails to recover fibers through the manual region of interest. This is due to the presence of tensors with very low determinant (low diffusion). It is interesting to note that with the same parameter $\beta = 10$, the algorithm found ten clusters for the healthy subject while it found only three for the diseased patient, which seems to reflect the advantage of letting the number of clusters a variable of the optimization problem. Note also how the soleus (in cyan in [Fig.4(c)]) is subdivided in an anterior and a posterior part in [Fig.4(a)], which is consistent with its anatomy of oblique fibers converging towards a central aponeurosis.

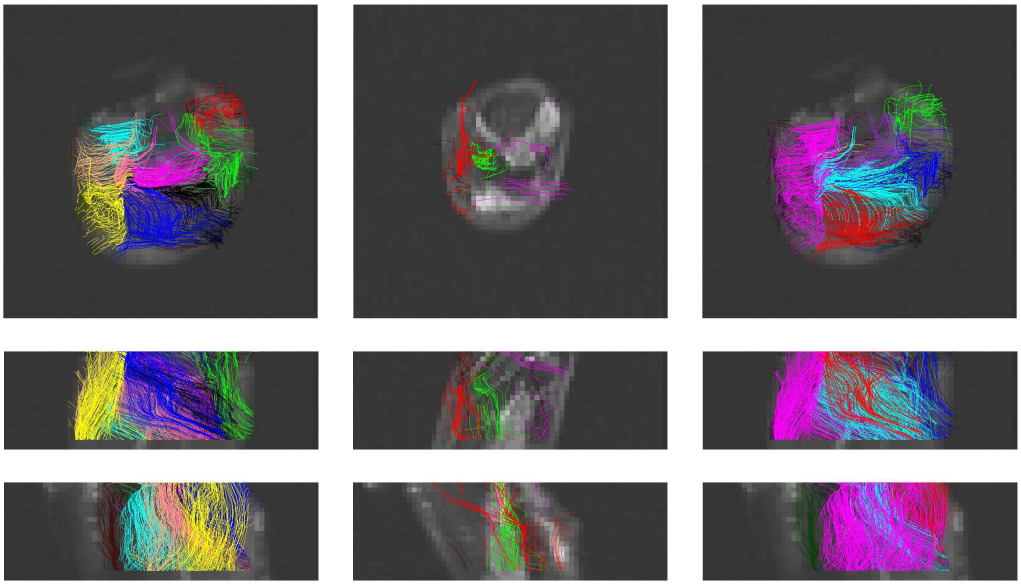
6 Conclusion

In this report, we proposed a skeletal fiber clustering framework. The main ingredient is the definition of a family of metrics between fibers that encompasses spatial and diffusion information. For clustering purposes, we used diffusion maps and also a novel manifold clustering method where there is no need to perform an embedding in a low dimensional space or to select the number of clusters and applied the method to the bundling of the fibers of the human skeletal muscle. While the manifold assumption seems to hold experimentally for the muscle, the case of multiple disconnected manifolds [27] should be explored. A procedure of landmark selection should also be investigated for manifold LP-based clustering, as well as other metrics between fibers. Based for example on the metric in (3), the clustering can also be done at the tensor level.

It would be interesting as a future research direction to perform statistical analysis of diffusion properties within each fiber bundle and assess the changes induced by myopathies.

Acknowledgments

This work was partially supported by Association Française contre les Myopathies (AFM: `http://www.afm-france.org`) under the DTI-MUSCLE project.



(a) (b) (c)

Figure 4: Axial, coronal and saggital views of fiber segmentation obtained with the lower bound approximation for (a) a healthy subject in 10 classes (b) a diseased subject in 3 classes. The parameter β was set to 10 in both cases. In (c) the ground truth segmentation of (a) with the following muscles: the soleus (cyan) , lateral gastrocnemius (red), medial gastrocnemius (magenta), posterior tibialis (yellow, for this patient it is barely visible because it is too small and has too few fibers), anterior tibialis (green), extensor digitorum longus (purple), and the peroneus longus (blue).

References

[1] Denis Le Bihan, Jean-Francois Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabrait, "Diffusion tensor imaging: Concepts and applications", *Journal of Magnetic Resonance Imaging*, vol. 13, pp. 534–546, 2001.

[2] Richard J. Gilbert and Vitaly J. Napadow, "Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging", *Dysphagia*, vol. 20, pp. 1–7, 2005.

[3] Usha Sinha and Lawrence Yao, "In vivo diffusion tensor imaging of human calf muscle", *Journal of Magnetic Resonance Imaging*, vol. 15, no. 1, pp. 87–95, 2002.

[4] Craig J. Galban, Stefan Maderwald, Kai Uffmann, Armin de Greiff, and Mark E. Ladd, "Diffusive sensitivity to muscle architecture: a magnetic resonance diffusion tensor imaging study of the human calf", *European Journal of Applied Physiology*, vol. 93, no. 3, pp. 253 – 262, 2004.

[5] B.M. Damon, Z. Ding, A.W. Anderson, A.S. Freyer, and J.C. Gore, "Validation of diffusion tensor MRI-based muscle fiber tracking", *Magnetic Resonance in Medicine*, vol. 48, pp. 97–104, 2002.

[6] A. Brun, H. Knutsson, H. J. Park, M. E. Shenton, and C.-F. Westin, "Clustering fiber tracts using normalized cuts", in *MICCAI*, 2004.

[7] Lauren ODonnell and Carl-Fredrik Westin, "White matter tract clustering and correspondence in populations", in *MICCAI*, 2005.

[8] Anders Brun, Hae-Jeong Park, Hans Knutsson, and Carl-Fredrik Westin, "Coloring of DT-MRI fiber traces using Laplacian eigenmaps", in *EUROCAST*, 2003.

[9] Andy Tsai, Carl-Fredrik Westin, Alfred O. Hero, and Alan S. Willsky, "Fiber tract clustering on manifolds with dual rooted-graphs", in *CVPR*, 2007.

[10] M. Maddah, W. Grimson, S. Warfield, and W. Wells, "A unified framework for clustering and quantitative analysis of white matter fiber tracts", *Medical Image Analysis*, vol. 12, no. 2, pp. 191–202, 2008.

[11] Demian Wassermann and Rachid Deriche, "Simultaneous manifold learning and clustering: Grouping white matter fiber tracts using a volumetric white matter atlas", in *MICCAI 2008 Workshop - Manifolds in Medical Imaging: Metrics, Learning and Beyond.*, 2008.

[12] Orly Zvitia, Arnaldo Mayer, and Hayit Greenspan, "Adaptive mean-shift registration of white matter tractographies", in *ISBI*, 2008.

[13] Peter Savadjiev, Jennifer S. W. Campbell, G. Bruce Pike, and Kaleem Siddiqi, "Streamline flows for white matter fibre pathway segmentation in diffusion MRI", in *MICCAI*, 2008.

[14] Mahnaz Maddah, Andrea U. J. Mewes, Steven Haker, W. Eric L. Grimson, and Simon K. Warfield, "Automated atlas-based clustering of white matter fiber tracts from DTMRI", in *MICCAI*, 2005.

[15] Lauren ODonnell and Carl-Fredrik Westin, "Automatic tractography segmentation using a high-dimensional white matter atlas", *IEEE TMI*, vol. 26, no. 11, pp. 1562–1575, 2007.

[16] R. Neji, J.-F. Deux, G. Fleury, M. Maatouk, G. Langs, J.-Ph. Thiran, G. Bassez, A. Rahmouni, and N. Paragios, "A kernel-based approach to diffusion tensor and fiber clustering in the human skeletal muscle", INRIA research report 6686, 2008.

[17] D. Wassermann, M. Descoteaux, and R. Deriche, "Diffusion maps clustering for magnetic resonance q-ball imaging segmentation", *International Journal on Biomedical Imaging*, 2008.

[18] Rachid Deriche, David Tschumperlé, Christophe Lenglet, and Mikaël Rousson, "Variational Approaches to the Estimation, Regularization and Segmentation of Diffusion Tensor Images", in *Mathematical Models of Computer Vision: The Handbook*, Paragios, Chen, and Faugeras, Eds. Springer, 2005.

[19] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A Riemannian framework for tensor computing", *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.

[20] Tony Jebara, Risi Kondor, and Andrew Howard, "Probability product kernels", *Journal of Machine Learning Research*, vol. 5, pp. 819–844, 2004.

[21] A. Barmpoutis, B. C. Vemuri, and J. R. Forder, "Registration of high angular resolution diffusion MRI images using 4th order tensors", in *MICCAI*, 2007.

[22] Ronald R. Coifman and Stéphane Lafon, "Diffusion maps", *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 5–30, 2006.

[23] Georg Langs and Nikos Paragios, "Modeling the structure of multivariate manifolds: Shape maps", in *CVPR*, 2008.

[24] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas, "Clustering via LP-based stabilities", in *NIPS*, 2008.

[25] Vin de Silva and Joshua B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction", in *NIPS*, 2002.

[26] Pierre Fillard, Nicolas Toussaint, and Xavier Pennec, "Medinria: DT-MRI processing and visualization software", Similar Tensor Workshop, 2006.

[27] Alvina Goh and Rene Vidal, "Segmenting fiber bundles in diffusion tensor images", in *ECCV*, 2008.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249-6399