Manifold Learning in Medical Imaging

Samuel Kadoury

Additional information is available at the end of the chapter

Abstract

Manifold learning theory has seen a surge of interest in the modeling of large and extensive datasets in medical imaging since they capture the essence of data in a way that fundamentally outperforms linear methodologies, the purpose of which is to essentially describe things that are flat. This problematic is particularly relevant with medical imaging data, where linear techniques are frequently unsuitable for capturing variations in anatomical structures. In many cases, there is enough structure in the data (CT, MRI, ultrasound) so a lower dimensional object can describe the degrees of freedom, such as in a manifold structure. Still, complex, multivariate distributions tend to demonstrate highly variable structural topologies that are impossible to capture with a single manifold learning algorithm. This chapter will present recent techniques developed in manifold theory for medical imaging analysis, to allow for statistical organ shape modeling, image segmentation and registration from the concept of navigation of manifolds, classification, as well as disease prediction models based on discriminant manifolds. We will present the theoretical basis of these works, with illustrative results on their applications from various organs and pathologies, including neurodegenerative diseases and spinal deformities.

Keywords: manifold learning, medical imaging, discriminant manifolds, piecewise geodesic regression, spine deformities, neurodegenerative diseases, shape modeling

1. Introduction

Learning on large medical imaging datasets is an emerging discipline driven from the availability of vast amounts of raw data in many of today's biomedical studies. However, challenges such as unbalanced data distributions, complex multivariate data and highly variable structural topologies demonstrated by real-world samples makes it much more difficult to efficiently learn the associated representation. An important goal of scientific data analysis in medicine, particularly in neurosciences or oncology, is to understand the behavior of biological

process or physiological/morphological alterations. This introduces the need to synthesize large amounts of multivariate data in a robust manner and raises the fundamental question of data reduction: how to discover meaningful representations from unstructured high-dimensional medical images.

Several approaches have attempted to understand how dimension reduction and regression establishes the relationship in subspaces and finally determine statistics on manifolds that optimally describe the relationships between the samples [1]. However, certain assumptions based on the representation of shapes and images using smooth manifolds are made in most cases, which frequently will not be adequate in the presence of medical imaging data and often perturbed by nuisance articulations, clutter or varying contrast.

High-dimensional classification methods have shown promise to measure subtle and spatially complex imaging patterns that have diagnostic value [2, 3]. Defining statistics on a manifold is not a straightforward process when simple statistics cannot be directly applied to general manifolds [4]. But while Euclidean estimators have been used for vector spaces, none have been adapted for multimodal data lying in different spaces. Still, there has been interest in the characterization of data in a Riemann space [5, 6]. Unfortunately, manifold-valued metrics based on the centrality theory or the geometric median [7] often lacks robustness to outliers.

A related topic lies in dimensionally reduced growth trajectories of various anatomical sites which have been investigated in neurodevelopment of newborns for example, based on geodesic shape regression to compute the diffeomorphisms with image time series of a population [8]. These regression models were also used to estimate spatiotemporal evolution of the cerebral cortex [9]. The concept of parallel transport curves in the tangent space from low-dimensional manifolds proposed by Schiratti et al. [10] was used to analyze shape morphology [11] and adapted for radiotherapy response [12]. Regression models were proposed for both cortical and subcortical structures, with 4D varifold-based learning framework with local topography shape morphing being proposed by Rekik et al. [13].

This chapter presents several manifold learning methodologies designed to address challenges encountered in medical imaging. In Section 2, we present an articulated shape inference model from nonlinear embeddings, expressing the global and local shape variations of the spine and vertebrae composing it, introduced in [14]. We then present in Section 3 a probabilistic model from discriminant manifolds to classify the neurodegenerative stage of Alzheimer's disease. Finally, a piecewise-geodesic transport curve in the tangent space from low-dimensional manifolds designed for the prediction of correction in spinal surgeries is shown in Section 4, introducing a time-warping function controlling the rate of shape evolution. We conclude this article in Section 5.

2. Shape inference through navigating manifolds

Statistical models of shape variability have been successful in addressing fundamental vision tasks such as segmentation and registration in medical imaging. However, the high dimensionality and complex nonlinear underlying structure unfortunately makes the commonly

used linear statistics inapplicable for anatomical structures. Manifold learning approaches map high-dimensional observation data that are presumed to lie on a nonlinear manifold, onto a single global coordinate system of lower dimensionality.

Inferring a model from the underlying manifold is not a novel concept but far from being trivial. In this section, we model both global statistics of the articulated model and local shape variations of vertebrae based on local measures in manifold space. We describe a spine inference/segmentation method from CT and MR images, where the model representation is optimized through a Markov Random Field (MRF) graph, balancing prior distribution with image data.

2.1. Data representation

Our spine model $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$ consists of an interconnection of L vertebrae. For each vertebra $\mathbf{s}_i$, we recover a triangular mesh with vertices $\left\{\mathbf{v}_j^i | j = 1, \ldots, V\right\}$, where the jth vertex corresponds to approximately the same location from one shape to another and V the number of vertices. Additionally, every $\mathbf{s}_i$ is annotated with landmarks on each model to rigidly register each object to its upper neighbor. Hence, an articulated deformable model (ADM) is represented by a vector of local intervertebral rigid transformations $A = [T_1, T_2, \ldots, T_L]$. To perform global shape modeling of $\mathbf{S}$, we convert A to an absolute representation $\mathbf{A}_{abs} = [T_1, T_1 \circ T_2, \ldots, T_1 \circ T_2 \circ \ldots \circ T_L]$ using recursive compositions. The transformations are expressed in the local coordinate system (LCS) of the lower vertebra. Center of transformation is the intersection of all three vertebral axes, following anteroposterior, cranial-caudal and left-right directions. Rigid transformations described here are the combination of a rotation matrix R, a translation t and scaling s. We formulate the rigid transformation $T = \{s, R, t\}$ of a triangular mesh model as $y = sRx + t$ where $x, y, t \in \mathfrak{R}^3$.

2.2. Manifold embedding

For nonlinear embeddings, we rely on the absolute vector representation $\mathbf{A}_{abs}$ as given previously. Let us now consider N articulated shape models expressed by the feature vectors $\mathbf{A}_{abs}^i$, of dimensionality D. The aim is to create a low-dimensional manifold consisting of N points $\mathbf{Y}_i$, $\mathbf{Y}_i \in \mathfrak{R}^d$, $i \in [1, N]$ where $d \ll D$ based on [15]. In such a framework, if an adequate number of data points is available, then the underlying manifold $\mathcal{M}$ is considered to be "well-sampled." Therefore, it can represent the underlying population structure. In the sub-cluster corresponding to a pathological population, each point of the training set and its neighbors would lie within a locally linear patch as illustrated in **Figure 1**.

The main limitation of embedding algorithms is the assumption of Euclidean metrics in the ambient space to evaluate similarity between sample points. Thus, a metric in the space of articulated structures is defined so that it accommodates for anatomical spine variability and adopts the intrinsic nature of the Riemannian manifold geometry allowing us to discern between articulated shape deformations in a topological invariant framework. For each point, the K closest neighbors are selected using a distortion metric which is particularly suited for geodesics. The metric $d_M\left(\mathbf{A}_{abs}^i, \mathbf{A}_{abs}^j\right)$ estimates the distance of articulated models i, j where
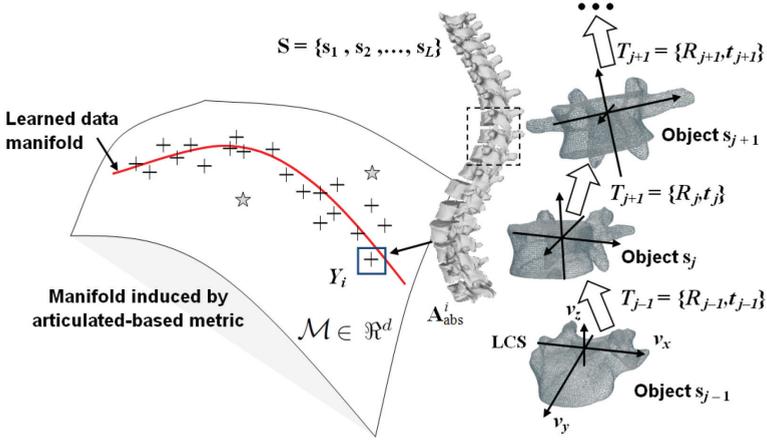
Figure 1. Representation of intervertebral transformations in manifold space.

A_{abs}^i. The distance measure for absolute representations can therefore be expressed as a sum of articulation deviations

$$d_M\left(\mathbf{A}_{abs}^i, \mathbf{A}_{abs}^j\right) = \sum_{k=1}^{L} d_M\left(T_k^i, T_k^j\right) = \sum_{k=1}^{L} \|\mathbf{t}_k^i - \mathbf{t}_k^j\| + \sum_{k=1}^{L} d_G\left(\mathbf{R}_k^i, \mathbf{R}_k^j\right). \tag{1}$$

While for the translation, the L_2 norm is chosen, geodesical distances are used between rotation neighborhoods. This is expressed as $d_G\left(\mathbf{R}_k^i, \mathbf{R}_k^j\right) = \|\log\left(\left(\mathbf{R}_k^i\right)^{-1}\mathbf{R}_k^j\right)\|_F$ where the log map is used to map a point in the manifold to the tangent plane.

Afterwards, the manifold reconstruction weights are estimated by assuming the local geometry of the patches can be described by linear coefficients that permit the reconstruction of every model point from its neighbors. In order to determine the value of the weights, the reconstruction errors are measured using the following objective function:

$$\varepsilon(\mathbf{W}) = \sum_{i=1}^{N} \left\| \mathbf{A}_{abs}^i - \sum_{j=1}^{K} W_{ij}\mathbf{A}_{abs}^j \right\|^2 \tag{2}$$

$$\text{subject to} \begin{cases} W_{ij} = 0 \text{ if } \mathbf{A}_{abs}^i \text{ not neighbor } \mathbf{A}_{abs}^j \\ \sum_j \mathbf{W}_{ij} = 1 \text{ for every } i. \end{cases} \tag{3}$$

Thus, $\varepsilon(\mathbf{W})$ sums the squared distances between all data points and their corresponding reconstructed points. The weights $\mathbf{W}_{ij}$ represent the importance of the jth data point to the reconstruction of the ith element.

The algorithm maps each high-dimensional $\mathbf{A}^i_{\text{abs}}$ to a low-dimensional $\mathbf{Y}_i$. These internal coordinates are found with a cost function minimizing the reconstruction error:

$$\Phi(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \mathbf{Y}_i - \sum_{j=1}^{K} \mathbf{W}_{ij} \mathbf{Y}_j \right\|^2$$
$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{M}_{ij} \mathbf{Y}_i^T \mathbf{Y}_j \tag{4}$$

with $\mathbf{M}$ as a sparse and symmetric $N \times N$ matrix enclosing the reconstruction weights $\mathbf{W}_{ij}$ such that $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$, and $\mathbf{Y}$ spanning the $\mathbf{Y}_i$'s. The optimal embedding, up to a global rotation, is obtained from the bottom $d + 1$ eigenvectors of $\mathbf{M}$ and helps to minimize the cost function $\Phi(\mathbf{Y})$ as a simple eigenvalue problem. The d eigenvectors form the d embedding coordinates. The coordinates $\mathbf{Y}_i$ can be translated by a constant displacement without affecting the overall cost $\Phi(\mathbf{Y})$. The eigenvector corresponding to the smallest eigenvalue corresponds to the mean value of the embedded data $\mathbf{Y}^0 = \{\mathbf{y}_1, \ldots, \mathbf{y}_d\}$, $\mathbf{y}_i = 0, \forall i$. This can be discarded with $\sum \mathbf{Y}_i = 0$ to obtain an embedding centered at the origin. Hence, a new ADM can be inferred in the embedded d-space as a low-dimensional point $\mathbf{Y}^{\text{new}}$ by finding its optimal manifold coordinates $\mathbf{y}_i$.

To obtain the articulation vector for a new embedded point in the ambient space (image domain), one has to determine the representation in high-dimensional space based on its intrinsic coordinates. We first assume an explicit mapping $f : \mathcal{M} \rightarrow \mathfrak{R}^D$ from manifold space $\mathcal{M}$ to the ambient space $\mathfrak{R}^D$. The inverse mapping of $\mathbf{Y}_i$ is then performed by estimating the relationship between $\mathfrak{R}^D$ and $\mathcal{M}$ as a joint distribution, such there exists a smooth functional which belongs to a local neighborhood. Theoretically the manifold should follow the conditional expectation:

$$f(Y_i) \equiv E\big(\mathbf{A}^i_{\text{abs}} | \mathcal{M}(A_i) = \mathbf{Y}_i\big) = \int A_i \frac{p(\mathbf{Y}_i, A_i)}{p_{\mathcal{M}(A_i)}(\mathbf{Y}_i)} dD \tag{5}$$

which captures the overall trend of the data in D-space. Here, both $p_{\mathcal{M}(A_i)}(\mathbf{Y}_i)$ (marginal density of $\mathcal{M}(A_i)$) and $p(\mathbf{Y}_i, A_i)$ (joint density) are unknown. Based on the *Nadaraya-Watson* kernel regression [16], we replace densities by kernel functions as $p_{\mathcal{M}(A_i)}(\mathbf{Y}_i) = \frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j)$ and $p(\mathbf{Y}_i, A_i) = \frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j) G_g(A_i, A_j)$ [17]. The Gaussian regression kernels G require the neighbors $\mathbf{A}^j_{\text{abs}}$ of $j\in\mathcal{N}(i)$ to determine the bandwidths h, g so it includes all K data points ($\mathcal{N}(i)$ representing the neighborhood of i). Plugging these estimates in Eq.(5), this gives:

$$f_{\text{NW}}(\mathbf{Y}_i) = \int A_i \frac{\frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j) G_g(A_i, A_j)}{\frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j)} dD. \tag{6}$$

By assuming G is symmetric about the origin, we propose to integrate in the kernel regression estimator, the manifold-based distortion metric d_M which is particularly suited for geodesic metrics and articulated diffeomorphisms. This generalizes the expectation such that the observations $\mathbf{Y}$ are defined in manifold space $\mathcal{M}$:

$$f_{\mathrm{NW}}(\mathbf{Y}_i) = \underset{\mathbf{A}_{\mathrm{abs}}^i}{\operatorname{argmin}} \frac{\sum_{j \in \mathcal{N}(i)} G(\mathbf{Y}_i, \mathbf{Y}_j) d_M\left(\mathbf{A}_{\mathrm{abs}}^i, \mathbf{A}_{\mathrm{abs}}^j\right)}{\sum_{j \in \mathcal{N}(i)} G(\mathbf{Y}_i, \mathbf{Y}_j)} \tag{7}$$

which integrates the distance metric $d_M\left(\mathbf{A}_{\mathrm{abs}}^i, \mathbf{A}_{\mathrm{abs}}^j\right)$ defined in Eq. (1) and updates $f_{\mathrm{NW}}(\mathbf{Y}_i)$ using the closest neighbors of point $\mathbf{Y}_i$ in the manifold space. This constrains the regression to be valid for similar data points in its vicinity since locality around $\mathbf{Y}_i$ preserves locality in $\mathbf{A}_{\mathrm{abs}}^i$.

2.3. Optimization on manifold

Once an appropriate modeling of spine shape variations is determined with a manifold, a successful inference between the image and manifold must be accomplished. We describe here how a new model is generated. We search the optimal embedded manifold point $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_d)$ of the global spine model. Such a strategy offers an ideal compromise between the prior constraints, as well as the individual shape variations described by the weight vector $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ in a localized sub-patch. The energy E of inferring the model $\mathbf{S}$ in the image $\mathcal{I}$ is a function of the set of displacement vectors Δ in the manifold space for global shape representation. This involves: (a) a data-related term expressing the image cost and (b) a global prior term measuring deformation between low-dimensional vectors with shape models. The third term represents (c) a higher-order term which is expressed by the reconstruction weights Ω for local vertebra modeling. The energy E can be expressed as the following combination of a global and local optimization:

$$E\left(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega\right) = V\left(\mathbf{Y}^0 + \Delta, \mathcal{I}\right) + \alpha\, V(\mathbf{N}, \Delta) + \beta\, V(\mathbf{H}, \Delta, \Omega). \tag{8}$$

The global alignment of the model with the target image primarily drives the deformation of the model. The purpose is to estimate the set of articulations describing the global spine model by determining its optimal representation $\mathbf{Y}^0$ in the embedded space. This is performed by obtaining the global representation using the mapping in (7) so that: $f_{\mathrm{NW}}(Y_i + \Delta) = f_{\mathrm{NW}}\left(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \boldsymbol{\delta}_d\}\right)$. This allows to optimize the model in manifold space coordinates while retrieving the articulations in $\mathcal{I}$. The global cost can be expressed as:

$$V\left(\mathbf{Y}^0 + \Delta, \mathcal{I}\right) = V\left(f_{\mathrm{NW}}\left(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \boldsymbol{\delta}_d\}\right), \mathcal{I}\right). \tag{9}$$

The inverse transform allows to obtain $\mathbf{A}_{\mathrm{abs}}^i + D$, with D as deformations in the image space. Since the transformations T_i are implicitly modeled in the absolute representation $\mathbf{A}_{\mathrm{abs}}^0$, we can formally consider the singleton image-related term as a summation of costs associated with each L vertebra of the model:

$$V\left(\mathbf{A}^0_{\text{abs}} + \boldsymbol{D}, \mathcal{I}\right) = \sum_{i=1}^{L} V_i\left(\mathbf{s}_i^*\left(T_i^0 + d_i\right), \mathcal{I}\right) \tag{10}$$

where $V_i(\mathbf{s}, \mathcal{I}) = \sum_{\mathbf{v}_i \in \mathbf{s}} \mathbf{n}_i^T(\mathbf{v}_i) \nabla \mathcal{I}(\mathbf{v}_i)$ minimizes the distance between mesh vertices of the inferred shape and gradient image $\mathcal{I}$ by a rigid transformation. Here, $\mathbf{n}_i$ is the normal pointing outwards and $\nabla \mathcal{I}(\mathbf{v}_i)$ the image gradient at $\mathbf{v}_i$.

The prior constraint for the rigid alignment are pairwise potentials between neighboring models y_i such that the difference in manifold coordinates is minimal with regards to a prior distribution of neighboring distances P:

$$\alpha V(\mathbf{N}, \Delta) = \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}\left(y_i^0 + \boldsymbol{\delta}_i, y_j^0 + \boldsymbol{\delta}_j, P\right). \tag{11}$$

This term represents the smoothness term of the global cost function to ensure that the deformation $\boldsymbol{\delta}_i$ applied to point coordinates are regular, with $V_{ij} = (0, 1)$ a distance assigning function based on the distances to P.

One can integrate the global data and prior terms along with local shape terms parameterized as the higher-order cliques, by combining (9), (11):

$$\begin{aligned}
E\left(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega\right) = {} & V\left(f_{\text{NW}}\left(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \boldsymbol{\delta_d}\}\right), \mathcal{I}\right)\right) \\
& + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}\left(y_i^0 + \boldsymbol{\delta}_i, y_j^0 + \boldsymbol{\delta}_j\right) + \beta \sum_{c \in \mathcal{C}} V_c\left(\mathbf{w}_c^0 + \omega_c\right).
\end{aligned} \tag{12}$$

The optimization strategy of the resulting MRF (12) in the continuous domain is not a straightforward problem. The convexity of the solution domain is not guaranteed, while gradient-descent optimization approaches are prone to nonlinearity and local minimums. We seek to assign the optimal labels $\mathcal{L}^\Delta = \{l_1, \ldots, l_d\}$ and $\mathcal{L}^\Omega = \{l_1, \ldots, l_n\}$ which are associated to the quantized space Δ of displacements and local weight parameters Ω respectively. We consider that displacing the coordinates of point y_i^0 by $\boldsymbol{\delta}^{l_i}$ is equivalent to assigning label l_i to y_i^0. An incremental approach is adopted where in each iteration t we look for the set of labels that improves the current solution s.t. $y_i^t = y_i^0 + \sum_t \boldsymbol{\delta}^{l_i t}$, which is a temporal minimization problem. Then (12) can be rewritten as:

$$\begin{aligned}
E^t\left(\mathcal{L}^\Delta, \mathcal{L}^\Omega\right) = {} & V\left(f_{\text{NW}}\left(\{y_1^{t-1}, l_1^\Delta, \ldots, y_d^{t-1}, l_d^\Delta\}\right), \mathcal{I}\right)\right) \\
& + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}\left(y_i^{t-1}, y_j^{t-1}, l_i^\Delta, l_j^\Delta\right) + \beta \sum_{c \in \mathcal{C}} V_c\left(\mathbf{w}_c^{t-1}, l_c^\Omega\right).
\end{aligned} \tag{13}$$

We solve the minimization of the higher-order cliques in (13) by transforming them into quadratic functions [18]. We apply the FastPD method [19] which solves the problem by formulating the duality theory in linear programming.

2.4. Results

Manifold learning. The manifold was built from a database containing 711 scoliotic spines demonstrating several types of deformities. Each spine model in the database was obtained from biplanar radiographic stereo-reconstructions. It is modeled with 12 thoracic and 5 lumbar vertebrae (17 in total), represented by 6 landmarks on each vertebra (4 pedicle extremities and 2 endplate center points) which were manually identified by an expert on the radiographic images. The resulting manifold is shown in **Figure 2**.

Adaptation of the articulated model was done on two different data sets. The first consisted of volumetric CT scans ($512 \times 512 \times 251$, resolution: 0.8×0.8 mm, thickness: $1 - 2$ mm) of the lumbar and main thoracic regions obtained from 21 different patients acquired for operative planning purposes. The MR dataset comprised multi-parametric volumetric data ($256 \times 256 \times 160$, resolution: 1.3×0.9 mm, thickness: 1 mm) of 8 patients acquired for diagnostic purposes. For this study, only the T1 sequence was selected for the experiments. All patients on both datasets (29 in total) had 12 thoracic and 5 lumbar vertebrae. Both CT and MR data were manually annotated with 3D landmarks by an expert in radiology, corresponding to left and right pedicle tips as well as midpoints of the vertebral body. Segmentation of the vertebrae from the CT and MR slices were also made by the same operator.

CT imaging experiments. We first evaluated the model accuracy in CT images by computing the correspondence of the inferred vertebral mesh models to the segmented target structures. As a preprocessing step, a rough thresholding was performed on the whole volume to filter out noise artifacts. The overall surface-to-surface comparison results between the inferred 3D vertebral models issued from the articulated model and from known segmentations were first
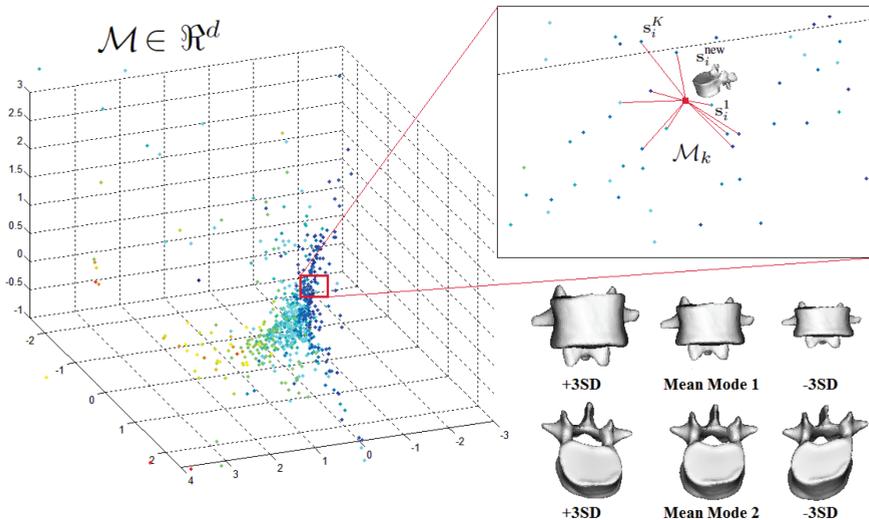
Figure 2. Low-dimensional manifold embedding of the spine dataset comprising 711 models exhibiting various types of deformities. The sub-domain was used to estimate both the global shape pose costs and individual shape instances based on local neighborhoods.

calculated. The mean errors are 2.2 ± 1.5 mm (range: $0.6 - 5.4$ mm) for thoracic vertebra and 2.8 ± 1.9 mm (range: $0.7 - 8.1$ mm) for lumbar vertebra.

MR imaging experiments. For the experiments involving the segmentation of 3D spine models from MR images, the surface-to-surface comparison showed encouraging results (thoracic: 2.9 ± 1.8 mm, lumbar: 3.0 ± 1.9 mm) based on differences to ground-truth. As in the previous experiments with CT imaging, ground-truth data was generated by manually segmenting the structures models which were validated by an expert in radiology. As difficult as the CT inference is, the MR problem represent an even greater challenge as the image resolution is more limited and interslice spacing is increased compared to CT. Modeling of the statistical properties of the shape variations and global pose becomes even more important in this case, as it relies heavily in the nonlinear distribution of the patient morphology.

3. Probabilistic modeling of discriminant nonlinear manifolds in the identification of Alzheimer's

Neurodegenerative pathologies, such as Alzheimer's disease (AD), are linked with morphological and metabolic alterations which can be assessed from medical imaging and biological data. Recent advances in machine learning have helped to improve classification and prognosis rates, but lack a probabilistic framework to measure uncertainty in the data. In this section, we present a method to identify progressive mild cognitive impairment (MCI) and predict their conversion to AD from MRI and positron emitting tomography (PET) images. We show a discriminative probabilistic manifold embedding where locally linear mappings transform data points in low-dimensional space to corresponding points in high-dimensional space. A discriminant adjacency matrix is constructed to maximize the separation between different clinical groups, including MCI converters and nonconverters, while minimizing the distance in latent variables belonging to the same class.

3.1. Probabilistic model for discriminant manifolds

Manifold learning algorithms are based on the premise that data are often of artificially high dimension and can be embedded in a lower dimensional space. However the presence of outliers and multiclass information can on the other hand affect the discrimination and/or generalization ability of the manifold. We propose to learn the optimal separation between four classes (1) normal controls, (2) nonconverter MCI patients, (3) converter MCI patients and (4) AD patients, by using a discriminant graph-embedding. Here, n labeled points $\mathbb{Y} = \left\{ (y_i, l_i) \right\}_{i=1}^{n}$ defined in $\mathbb{R}^D$ are generated from the underlying manifold $\mathcal{M}$, where l_i denotes the label (NC, cMCI, nMCI or AD). For the labeled data, there exists a low-dimensional (latent) representation of the high-dimensional samples such that $\mathbb{X} = \{ (\mathbf{x}_i, l_i) \}_{i=1}^{n}$ defined in $\mathbb{R}^d$. We assume here that the mapping $\mathbf{M}_i \in \mathbb{R}^{D \times d}$ between high and low-dimensional spaces is locally linear, such that tangent spaces in local neighborhoods can be estimated with $\mathbf{y}_j - \mathbf{y}_i$ and

$\mathbf{x}_j - \mathbf{x}_i$, representing the pairwise differences between connected neighbors i, j. Therefore the relationship can be established as $\mathbf{y}_j - \mathbf{y}_i \approx \mathbf{M}_i(\mathbf{x}_j - \mathbf{x}_i)$.

In order to effectively discover the low-dimensional embedding, it is necessary to maintain the local structure of the data in the new embedding. The graph $G = (V, W)$ is an undirected similarity graph, with a collection of nodes V connected by edges, and the symmetric matrix W with elements describing the relationships between the nodes. The diagonal matrix D and the Laplacian matrix L are defined as $L = D - W$, with $D(i, i) = \sum_{j \neq i} W_{ij} \forall i$.

Using the theoretical framework from [20], we can determine a distribution of linear maps associated with the low-dimensional representation to describe the data likelihood for a specific model:

$$\log p(\mathbb{Y}|G) = \log \iint p(\mathbb{Y}, \mathbf{M}, \mathbb{X}|G) d\mathbf{x} d\mathbf{M} \tag{14}$$

This joint distribution can be separated into three prior terms: the linear maps, latent variables and the likelihood of the high dimensional points $\mathbb{Y}$:

$$p(\mathbb{Y}, \mathbf{M}, \mathbb{X}|G) = p(\mathbb{Y}|\mathbf{M}, \mathbb{X}, G)p(\mathbf{M}|G)p(\mathbb{X}|G) \tag{15}$$

We now define the discriminant similarity graphs establishing neighborhood relationships, as well define each of the three prior terms included in the joint distribution.

Within and between similarity graphs: In our work, the geometrical structure of $\mathcal{M}$ can be modeled by building a within-class similarity graph W_w for feature vectors of same group and a between-class similarity graph W_b, to separate features from all four classes. When constructing the discriminant locally linear latent variable embedding, elements are partitioned into W_w and W_b classes. The intrinsic graph G is first created by assigning edges only to samples of the same class (ex: nMCI). Each sample is therefore reconstructed only from feature vectors of the same clinical group. Local reconstruction coefficients are incorporated in the within-class similarity graph, such that W_w is defined as:

$$W_{w_{i,j}} = \begin{cases} 1 & \text{if } \mathbf{y}_i \in \mathcal{N}_w(\mathbf{y}_j) \text{ or } \mathbf{y}_j \in \mathcal{N}_w(\mathbf{y}_i) \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

with $\mathcal{N}_w$ containing neighbors of the same class. Conversely, W_b depicts the statistical properties to be avoided in the inference process. Distances between samples from different clinical groups are computed as:

$$W_{b_{i,j}} = \begin{cases} 1 & \text{if } \mathbf{y}_i \in \mathcal{N}_b(\mathbf{y}_j) \text{ or } \mathbf{y}_j \in \mathcal{N}_b(\mathbf{y}_i) \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

with $\mathcal{N}_b$ containing neighbors having different class labels from the ith sample. The objective is to transform points to a new manifold $\mathcal{M}$ of dimensionality d, i.e., $\mathbf{y}_i \rightarrow \mathbf{x}_i$, by mapping connected samples from the same group in W_w as close as possible to the class cluster, while moving NC, nMCI, cMCI and AD samples of W_b as far away from one another.

Model components: The prior added on the latent variables $\mathbb{X}$ are located at the origin of the low-dimensional domain, while minimizing the Euclidean distance of neighboring points that are associated with the neighborhood of high-dimensional points and maximizing the distance between coordinates of different classes. In order to set the variables with an expected scale α and H representing the probability density function, the following log prior is defined:

$$\log p(\mathbb{X}|\boldsymbol{W},\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\left(\alpha\|\mathbf{x}_i\| + \sum_{j=1}^{n}W_{w_{i,j}}\|\mathbf{y}_i - \mathbf{y}_j\|^2 - \sum_{j=1}^{n}W_{b_{i,j}}\|\mathbf{y}_i - \mathbf{y}_j\|^2\right) - \log H_{\mathbb{X}} \quad (18)$$

The prior added to the linear maps defines how the tangent planes described in low and high dimensional spaces are similar based on the Frobenius norm. This prior ensures smooth manifolds:

$$\log p(\mathbf{M}|\boldsymbol{W}) = -\frac{1}{2}\left(\left\|\sum_{i=1}^{n}\mathbf{x}_i\right\|_F^2 - \sum_{i=1}^{n}\sum_{j=1}^{n}\left(W_{w_{i,j}} - W_{b_{i,j}}\right)\|\mathbf{M}_i - \mathbf{M}_j\|_F^2\right) - \log H_{\mathbf{M}} \quad (19)$$

Finally, approximation errors from the linear mapping $\mathbf{M}_i$ between low and high-dimensional domains are penalized by including the following log likelihood:

$$\begin{aligned}\log p(\mathbb{Y}|\mathbb{X},\mathbf{W},\gamma) = {}& \|\sum_{i=1}^{n}\mathbf{y}_i\|^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{w_{i,j}}\Delta(i,j)^{\mathrm{T}}\gamma\mathbf{I}\Delta(i,j)\\ & + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{b_{i,j}}\Delta(i,j)^{\mathrm{T}}\gamma\mathbf{I}\Delta(i,j) - \log H_{\mathbf{y}}\end{aligned} \quad (20)$$

with $\Delta(i,j)$ the difference in Euclidean distance between pairs of neighbors in high and low-dimensional space and γ the update parameter for the EM inference. Samples of $\mathbf{y}$ are drawn from a multivariate normal distribution.

3.2. Variational inference

The objective is to infer the low-dimensional coordinates and linear mapping function for the described model, as well as the intrinsic parameters of the model $\Phi = (\alpha,\gamma)$. This is achieved by maximizing the marginal likelihood of:

$$\log p(\mathbb{Y}|\boldsymbol{W},\Phi) = \iint \rho(\mathbf{M},\mathbb{X})\log\frac{p(\mathbb{Y},\mathbf{M},\mathbb{X}|\boldsymbol{W},\Phi)}{\rho(\mathbf{M},\mathbb{X})}\,d\mathbf{x}\,d\mathbf{M}. \quad (21)$$

By assuming the posterior $\rho(\mathbf{M},\mathbb{X})$ can be factored in separate terms $\rho(\mathbf{M})$ and $\rho(\mathbb{X})$, a variational expectation maximization algorithm can be used to determine the model's parameters, which are initialized with Φ. The E-step updates the independent posteriors $\rho(\mathbb{X})$ and $\rho(\mathbf{M})$, while the parameters of Φ are updated in the M-step by maximizing Eq. (21).

The discriminant latent variable model can then be used to perform the mapping of new image feature vectors to the manifold. The variational EM algorithm described in the previous section

can be used to transform a set of new input points $\mathbf{y}_q$ without changing the overall neighborhood graph structure, by finding the distribution of the local linear map $\mathbf{y}_q$ and it is low-dimensional coordinate using the E-step explained above. Once the manifold representation $\mathbf{x}_q$ is obtained, a cluster analysis finds the corresponding class in the manifold, yielding a prediction of the input feature vector $\mathbf{y}_q$.

3.3. Experiments

We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) database with 1.5 or 3.0 T structural MR images (adni.loni.usc.edu) and FDG-PET images. For this study, 187 subjects with both MRI and PET images during a 24 month period were used to train the probabilistic manifold model, including 46 AD patients, 94 MCI patients, and 47 normal controls. During the follow-up period, 43 MCI subjects converted to AD and 56 remained stable. All groups are matched approximately by age (mean of 76.7 ± 5.4) and gender. Images were non-rigidly registered to a standard template, which was then segmented using FSL-FIRST automatic segmentation [21].

A 9-fold cross-validation was performed to assess the performance of the method. The optimal manifold dimensionality was set at $d = 8$, when the trend of the nonlinear residual reconstruction error stabilized for the entire training set. We evaluated the classification performance of the proposed method for discriminating between cMCI and nMCI patients, by training the model with MRI, PET and with MRI + PET biomarkers from the ROIs illustrated in **Figure 3**. **Figure 4** presents ROC curves obtained by the proposed and comparative methods such as SVM (nonlinear RBF kernel), LLE and LL-LVM [20]. The discriminative nature of the proposed framework clearly shows an improvement to standard learning approaches models which were trained using MRI only, PET only and combined multimodal features. It illustrates that increased accuracy (77.4%) can be achieved by combining MRI and PET features, showing the benefit of extracting complementary features from the dataset for prediction purposes. When comparing the performance of the proposed method to the other learning methods (SVM, LLE,
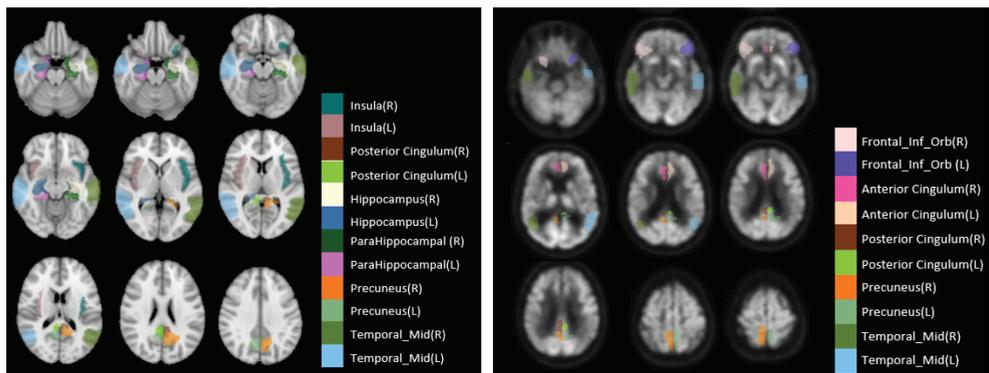
Figure 3. Selected FSL segmented brain regions for feature selection on (left) MRI and (right) PET images.
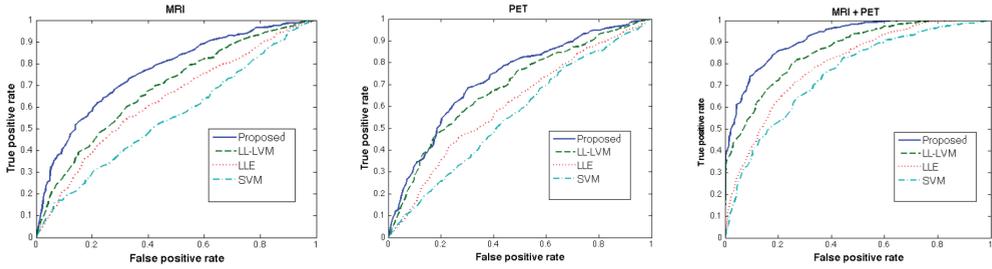
Figure 4. ROC curves comparing the SVM, LLE and LL-LVM with the proposed method for cMCI/nMCI prediction using MRI, PET and multimodality data.

LL-LVM), the probabilistic model integrating similarity graphs shows a statistically significant improvement ($p < 0.01$) to all three approaches based on paired t-test.

4. Spatiotemporal manifold prediction model for surgery prediction

In this final section, we present a statistical framework for predicting the surgical outcomes following spine surgery of adolescents with idiopathic scoliosis. A discriminant manifold is first constructed to maximize the separation between responsive and nonresponsive groups of patients. The model then uses subject-specific correction trajectories based on articulated transformations in order to map spine correction profiles to a group-average piecewise-geodesic path. Spine correction trajectories are described in a piecewise-geodesic fashion to account for varying times at follow-up exams, regressing the curve via a quadratic optimization process. To predict the evolution of correction, a baseline reconstruction is projected onto the manifold, from which a spatiotemporal regression model is built from parallel transport curves inferred from neighboring exemplars (**Figure 5**).
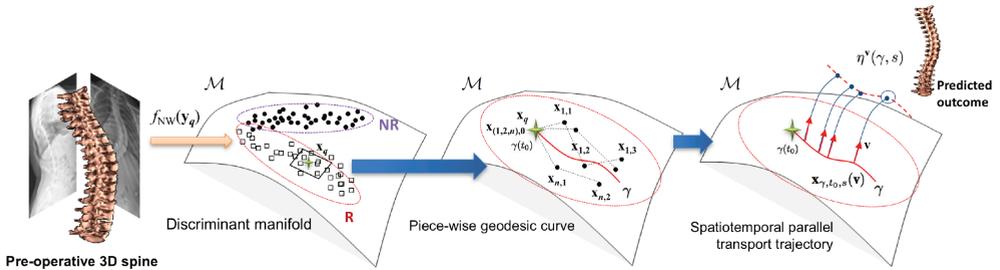
Figure 5. Proposed prediction framework for spine surgery outcomes. In the training phase, a dataset of spine models are embedded in a spatiotemporal manifold $\mathcal{M}$, into responsive (R) or nonresponsive (NR) groups. During testing, an unseen baseline 3D spine reconstruction $\mathbf{y}_q$ is projected on $\mathcal{M}$ using f_{NW} based on Nadaraya-Watson kernels. The closest samples to the projected point $\mathbf{x}$ are selected to regress the spatiotemporal curve γ used for predicting the correction due with surgery.

4.1. Discriminant embedding of spine models

We propose to embed a collection of nonresponsive (NR) and (2) responsive (R) patients to surgery which will offer a maximal separation between the classes, by using a discriminant graph-embedding. Here, n labeled points $\mathbb{Y} = \{(\mathbf{y}_i, l_i, t_i)\}_{i=1}^n$ defined in $\mathbb{R}^D$ are embedded in the low-dimensional manifold $\mathcal{M}$, where l_i describes the label (NR or R) and t_i defines the time of follow-up. We assume that for the sampled data, an underlying manifold of the high-dimensional data exists such that $\mathbb{X} = \{(\mathbf{x}_i, l_i, t_i)\}_{i=1}^n$ defined in $\mathbb{R}^d$. We rely on the assumption that a locally linear mapping $\mathbf{M}_i \in \mathbb{R}^{D \times d}$ exists, where local neighborhoods are defined as tangent planes estimated with $\mathbf{y}_j - \mathbf{y}_i$ and $\mathbf{x}_j - \mathbf{x}_i$, describing the paired distances between linked neighbors i, j. Hence, the relationship can be established as $\mathbf{y}_j - \mathbf{y}_i \approx \mathbf{M}_i(\mathbf{x}_j - \mathbf{x}_i)$.

Because the discriminant manifold structure in $\mathbb{R}^d$ requires to maintain the local structure of the underlying data, a undirected similarity graph $\mathcal{G} = (V, W)$ is built, where each node V are connected to each other with edges that are weighted with the graph W. The overall structure of $\mathcal{M}$ is therefore defined with W_w for feature vectors belonging to the same class and W_b, which separate features from both classes. During the embedding of the discriminant locally linear latent manifold, data samples are divided between W_w and W_b.

4.2. Piecewise-geodesic spatiotemporal manifold

Once sample points $\mathbf{x}_i$ are in manifold space, the objective is to regress a regular and smooth piecewise-geodesic curve $\gamma : [t_1, t_N]$ that accurately fits the embedded data describing the spatiotemporal correction following surgery within a 2 year period. For each sample data $\mathbf{x}_i$, the K closest individuals demonstrating similar baseline features are identified from the embedded data, creating neighborhoods $\mathcal{N}(\mathbf{x}_q)$ with measurements at different time points, thus creating a low-dimensional Riemannian manifold where data points $\mathbf{x}_{i,j}$, with i denoting a particular individual, j the time-point measurement and $j = 0$ the preoperative model. By assuming the manifold domain is complete and piecewise-geodesic curves are defined for each time trajectories, time-labeled data can be regressed continuously in $\mathbb{R}^D$, thereby creating smooth curves in time intervals described by samples in $\mathbb{R}^d$.

However, due to the fact the representation of the continuous curve is a variational problem of infinite dimensional space, the implementation follows a discretization process which is derived from the procedure in [22], such that:

$$
\begin{aligned}
E(\gamma) = {} & \frac{1}{K_d} \sum_{i=1}^{K_d} \sum_{j=0}^{t_N} w_i \| \gamma(t_{i,j}) - (\mathbf{x}_{i,j} - (\mathbf{x}_{i,0} - \mathbf{x}_q)) \|^2 \\
& + \frac{\lambda}{2} \sum_{i=1}^{K_d} \alpha_i \| v_i \|^2 + \frac{\mu}{2} \sum_{i=1}^{K_d} \beta_i \| a_i \|^2 .
\end{aligned}
\tag{22}
$$

This minimization process simplifies the problem to a quadratic optimization, solved with LU decomposition. The piecewise nature is represented by the term $K_d \in \mathcal{N}(\mathbf{x}_q)$, defined as

samples along γ. The first component of Eq.(22) is a penalty term to minimize the geodesic distance between samples $\mathbf{x}_{i,j}$ and the regressed curve, where w_i are weight variables based on sample distances. This helps regress a curve that will lie close to $\mathbf{x}_{i,j}$, shifted by $\mathbf{x}_q$ in order to have the initial reconstructions co-registered. The second term represents the velocity of the curve (defined by v_i, approximating $\dot{\gamma}(t_i)$), minimizing the L_2 distance of the 1st derivative of γ. By minimizing the value of the curve's first derivatives, this prohibits any discontinuities or rapid transitions of the curve's direction, and is modulated by α_i. Finally, an acceleration penalty term (defined by a_i) focuses on the 2nd derivative of γ with respect to t_i by minimizing the L_2 norm. The acceleration is modulated by β_i. Estimates for v_i and a_i (weighted by $\{\lambda, \mu\}$, respectively), are generated using geometric finite differences. These estimates dictates the forward and backward step-size on the regressed curve, leading to directional vectors in $\mathcal{M}$ as shown in [22]. In order to minimize $E(\gamma)$, a nonlinear conjugate gradient technique defined in the low-dimensional space $\mathbb{R}^d$ is used, thus avoiding convergence and speed issues. The regressed curve γ is therefore defined for all time points, originating at t_0. The curve creates a group average of spatiotemporal transformations based on individual correction trajectories.

4.3. Prediction of spine correction

Finally, to predict the evolution of spine correction from an unseen preoperative spine model, we use the geodesic curve $\gamma : \mathbb{R}^D \to \mathcal{M}$ modeling the spatiotemporal changes of the spine, where each point $\mathbf{x} \in \mathcal{M}$ is associated to a speed vector $\mathbf{v}$ defined with a tangent plane on the manifold such that $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}$.

Based on Riemannian theory, an exponential mapping function at $\mathbf{x}$ with velocity $\mathbf{v}$ can be defined from the geodesics such that $e_{\mathbf{x}}^{\mathcal{M}}(\mathbf{v})$. Using this concept, parallel transport curves defined in $T_{\mathbf{x}}$ can help define a series of time-index vectors along γ as proposed by [10]. The collection of parallel transport curves allows to generate an average trajectory in ambient space $\mathbb{R}^D$, describing the spine changes due to the corrective forces of tethering. The general goal is to begin the process at the preoperative sample, and navigate the piecewise-geodesic curve describing correction evolution in time, where one can extract the appearance at any point (time) in $\mathbb{R}^D$ using the exponential mapping. For implementation purposes, the parallel transport curve are constrained within a smooth tubular boundary perpendicular to the curve (from an ICA) to generate the spatiotemporal evolution in the coordinate system of the preoperative model.

Hence, given the manifold at time t_0 with $\mathbf{v}$ defined in the tangent plane and the regressed piecewise-geodesic curve γ, the parallel curve is obtained as:

$$\eta^{\mathbf{v}}(\gamma, s) = e_{\gamma(s)}^{\mathcal{M}}\big(\mathbf{x}_{\gamma, t_0, s}(\mathbf{v})\big), \ \ s \in \mathbb{R}^d. \tag{23}$$

Therefore by repeating this mapping for manifold points seen as samples of individual progression trajectories along $\gamma(s)$, an evolution model can be generated. Whenever a new sample is embedded, new samples points along $\gamma(s)$, denoted as $\eta^{\mathbf{v}}(\gamma, \cdot)$ can be generated parallel to the regressed piecewise curve in $\mathcal{M}$, capturing the spatiotemporal changes in correction.

A time warp function allowing s to vary along the geodesic curve is described as $\phi_i(t) = \theta_i(t - t_0 - \tau_i) + t_0$. Here, we propose to incorporate a personalized acceleration factor based on the spine maturity and flexibility derived from the spine bending radiographs and Risser grade. A coefficient $\theta_i = C_i \times R_i$ describing the change in Cobb angle C_i between poses, and modulated by the Risser grade R_i. This coefficient regulates the rate of correction based on the K neighboring samples. Finally, to take under account the relative differences between the group-wise samples and the query model once mapped onto the regressed curve, a time-shift parameter τ_i is incorporated in the warp function.

For spine correction evolution, displacement vectors $\mathbf{v}_i$ are obtained by a PCA of the hyperplane crossing $T_{\mathbf{x}_i}\mathcal{M}$ in manifold $\mathcal{M}$ [10]. Hence, for any query sample $\mathbf{x}_q$ which represents the mapped preoperative 3D reconstruction (prior to surgery), the predicted model at time t_k can be regressed from the piecewise-geodesic curve generated from embedded samples $\mathbf{x}$ in $\mathcal{N}(\mathbf{x}_q)$ such that:

$$\mathbf{y}_{q,t_k} = \eta^{\mathbf{v}_q}\left(\gamma, \phi_i(t_k)\right) + \varepsilon_{q,t_k} \tag{24}$$

which yields a predicted postoperative model $\mathbf{y}_{q,t_k}$ in high-dimensional space $\mathbb{R}^D$, and ε_{q,t_k} a zero-mean Gaussian distribution. The generated model offers a complete constellation of interconnected vertebral models composing the spine shape $\mathbf{S}$, at first-erect (FE), 1 or 2-year visits, including landmarks on vertebral endplates and pedicle extremities, which can be used to capture the local shape morphology with the correction process.

4.4. Experiments

The discriminant manifold was trained from a database of 438 3D spine reconstructions generated from biplanar images [23], originating from 131 patients demonstrating several types of deformities with immediate follow-up (FE), 1 and 2 year visits. Patients were recruited from a single center prospective study. Patients were divided in two groups, with the first group composed of 94 responsive patients showing a reduction in Cobb angle over or equal to 10° between the FE and follow-up visit. The second group was composed of 37 nonresponsive (NP) patients with a reduction of less than 10°. We evaluated the geometrical accuracy of the predictive manifold for 56 unseen surgical patients (mean age 12 ± 3, average main Cobb

	FE visit			1-year visit			2-year visit		
	3D RMS	Dice	Cobb	3D RMS	Dice	Cobb	3D RMS	Dice	Cobb
Biomec. sim	3.3 ± 1.1	85 ± 3.4	2.8 ± 0.8	3.6 ± 1.2	84 ± 3.6	3.2 ± 0.9	4.1 ± 2.3	82 ± 3.9	3.6 ± 1.0
LL-LVM [20]	3.6 ± 1.4	83 ± 4.0	3.8 ± 1.5	4.7 ± 3.3	79 ± 4.4	5.5 ± 2.6	6.6 ± 4.4	71 ± 5.9	7.0 ± 3.9
Deep AE [24]	4.1 ± 1.5	80 ± 4.4	5.1 ± 2.7	5.0 ± 1.9	77 ± 4.9	5.8 ± 3.0	6.3 ± 4.6	72 ± 5.7	6.6 ± 4.2
Proposed	2.4 ± 0.8	92 ± 2.7	1.8 ± 0.5	2.9 ± 0.9	90 ± 2.8	2.0 ± 0.7	3.2 ± 1.3	87 ± 3.1	2.1 ± 0.6

Predictions are evaluated at FE, 1 and 2-years.

Table 1. 3D RMS errors (mm), dice (%) and cobb angles (°) for the proposed method, and compared with biomechanical simulations, locally linear latent variable models (LL-LVM) and deep auto-encoders (AE).

angle on the frontal plane at the first visit was $47 \pm 10°$), with predictions at $t = 0$ (FE), $t = 12$ and $t = 24$ months. For the predicted models, we evaluated the 3D root-mean-square difference of the vertebral landmarks generated, the Dice coefficients of the vertebral shapes and in the main Cobb angle. The results are shown in **Table 1**. Results were confronted to other techniques such as biomechanical simulations performed on each subject using finite element modeling with ex-vivo parameters [25], a locally linear latent variable model [20] and a deep auto-encoder network [24]. Results from the predicted geometrical models show the regressed spatiotemporal geodesic curve yields anatomically coherent structures, with accurate local vertebral morphology.

5. Discussion

Algorithms capable of extracting clinically relevant and meaningful descriptions from medical imaging datasets have become of widespread interest to theoreticians as well as practitioners in the medical field, accelerating the pace in recent years involving varied fields such as in machine learning, geometry, statistics and genomics to propose new insights for the analysis of imaging and biologic datasets. Towards this end, manifold learning has demonstrated a tremendous potential to learn the underlying representation of high-dimensional, complex imaging datasets.

We presented frameworks describing longitudinal, multimodal image features from neuroimaging data using a Bayesian model for discriminant nonlinear manifolds to predict the conversion of progressive MCI to Alzheimer's disease. This probabilistic method introduces class-dependent latent variables which is based on the concept that local structure is transformed from manifold to the high-dimensional domain. This variational learning method can ultimately assess uncertainty within the manifold domain, which can lead to a better understanding of relationships between converters and nonconverters for patients with MCI.

Finally, a prediction method for the outcomes of spine surgery using geodesic parallel transport curves generated from probabilistic manifold models was presented. The mathematical models allow to describe patterns in a nonlinear and discriminant Riemannian framework by first distinguishing nonprogressive and progressive cases, followed by a prediction of structural evolution. The proposed model provides a way to analyze longitudinal samples from a geodesic curve in manifold space, thus simplifying the mixed effects when studying group-average trajectories.

Author details

Samuel Kadoury

Address all correspondence to: samuel.kadoury@polymtl.ca

Polytechnique Montreal, Succ. Centre-Ville, Montreal, QC, Canada

References

[1] Yang Y, Dunson DB, et al. Bayesian manifold regression. The Annals of Statistics. 2016; **44**(2):876-905

[2] Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. NeuroImage. 2008;**41**(4): 1220-1227

[3] Li S, Shi F, Pu F, Li X, Jiang T, Xie S, Wang Y. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. American Journal of Neuroradiology. 2007; **28**(7):1339-1345

[4] Beg MF, Miller MI, Trouvé A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision. 2005; **61**(2):139-157

[5] Fletcher PT, Lu C, Joshi S. Statistics of shape via principal geodesic analysis on lie groups. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 1. IEEE; 2003. pp. I-I

[6] Pennec X. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision. 2006;**25**(1):127

[7] Fletcher PT, Venkatasubramanian S, Joshi S. The geometric median on Riemannian manifolds with application to robust atlas estimation. NeuroImage. 2009;**45**(1):S143-S152

[8] Singh N, Hinkle J, Joshi S, Fletcher PT. A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: International Conference on Information Processing in Medical Imaging. Springer; 2013. pp. 560-571

[9] Fishbaugh J, Prastawa M, Gerig G, Durrleman S. Geodesic regression of image and shape data for improved modeling of 4D trajectories. In: 2014 International Symposium on Biomedical Imaging. IEEE; 2014. pp. 385-388

[10] Schiratti JB, Allassonniere S, Colliot O, Durrleman S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: Advances in Neural Information Processing Systems. 2015:2404-2412

[11] Kadoury S, Mandel W, Roy-Beaudry, Nault ML, Parent S. 3-D morphology prediction of progressive spinal deformities from probabilistic modeling of discriminant manifolds. IEEE Transactions on Medical Imaging. 2017;**36**(5):1194-1204

[12] Chevallier J, Oudard S, Allassonnière S. Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA; 2017

[13] Rekik I, Li G, Lin W, Shen D. Predicting infant cortical surface development using a 4D varifold-based learning framework and local topography-based shape morphing. Medical Image Analysis. 2016;**28**:1-12

[14] Kadoury S, Labelle H, Paragios N. Spine segmentation in medical images using manifold embeddings and higher-order MRFs. IEEE Transactions on Medical Imaging. 2013;**32**: 1227-1238

[15] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000;**290**:2323-2326

[16] Nadaraya EA. On estimating regression. Theory of Probability and its Applications. 1964; **10**:186-190

[17] Davis B, Fletcher P, Bullitt E, Joshi S. Population shape regression from random design data. In: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007. IEEE. 2007;**1**:1-8

[18] Rother C, Kohli P, Feng W, Jia J. Minimizing sparse higher order energy functions of discrete variables. In: Conference on Computer Vision and Pattern Recognition; 2009. pp. 1382-1389

[19] Komodakis N, Tziritas G, Paragios N. Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal dual strategies. Computer Vision and Image Understanding. 2008;**112**(1):14-29

[20] Park M, Jitkrittum W, Qamar A, Szabó Z, Buesing L, Sahani M. Bayesian manifold learning: The locally linear latent variable model (LL-LVM). In: Advances in Neural Information Processing Systems. 2015:154-162

[21] Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage. 2011;**56**(3):907-922

[22] Boumal N, Absil PA. A discrete regression method on manifolds and its application to data on SO (n). IFAC Proceedings Volumes. 2011;**44**(1):2284-2289

[23] Humbert L, de Guise J, Aubert B, Godbout B, Skalli W. 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences. Medical Engineering & Physics. 2009;**31**(6):681-687

[24] Thong W, Parent S, Wu J, Aubin CE, Labelle H, Kadoury S. Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models. European Spine Journal. 2016;**25**(10):3104-3113

[25] Cobetto N, Parent S, Aubin CE. 3D correction over 2 years with anterior vertebral body growth modulation: A finite element analysis of screw positioning, cable tensioning and postop functional activities. Clinical Biomechanics. 2018;**51**:26-33