HAL
open science

Constrained manifold learning for the characterization of pathological deviations from normality

Nicolas Duchateau, Mathieu de Craene, Gemma Piella, Alejandro Frangi

To cite this version:

Nicolas Duchateau, Mathieu de Craene, Gemma Piella, Alejandro Frangi. Constrained manifold learning for the characterization of pathological deviations from normality. Medical Image Analysis, 2012, 16 (8), pp.1532-1549. 10.1016/j.media.2012.07.003 . hal-02282424

HAL Id: hal-02282424

https://hal.science/hal-02282424v1

Submitted on 11 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

Constrained Manifold Learning for the Characterization of Pathological Deviations from Normality

Nicolas Duchateau*, Mathieu De Craene, Gemma Piella, Alejandro F. Frangi

Center for Computational Imaging & Simulation Technologies in Biomedicine
Information & Communication Technologies Department, Universitat Pompeu Fabra
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina
Barcelona, Spain

Abstract

This paper describes a technique to (1) learn the representation of a pathological motion pattern from a given population, and (2) compare individuals to this population. Our hypothesis is that this pattern can be modeled as a deviation from normal motion by means of non-linear embedding techniques. Each subject is represented by a 2D map of local motion abnormalities, obtained from a statistical atlas of myocardial motion built from a healthy population. The algorithm estimates a manifold from a set of patients with varying degrees of the same disease, and compares individuals to the training population using a mapping to the manifold and a distance to normality along the manifold. The approach extends recent manifold learning techniques by constraining the manifold to pass by a physiologically meaningful origin representing a normal motion pattern. Interpolation techniques using locally adjustable kernel improve the accuracy of the method. The technique is applied in the context of cardiac resynchronization therapy (CRT), focusing on a specific motion pattern of intra-ventricular dyssynchrony called septal flash (SF). We estimate the manifold from 50 CRT candidates with SF and test it on 37 CRT candidates and 21 healthy volunteers. Experiments highlight the relevance of nonlinear techniques to model a pathological pattern from the training set and compare new individuals to this pattern.

Keywords: Manifold learning, pattern analysis, cardiac atlas, myocardial motion, CRT

1. Introduction

1.1. Patient selection for CRT

In this paper, we address the issue of learning the representation of a pathological motion pattern from a given population, and its use for the comparison of individuals to this population. Our strategy is based on the fact that different grades of the same disease, for a patient or within a population, can be seen as progressive impairments of the normal condition of an organism. This notion is particularly of interest for the understanding of this disease, from the early detection of its onset to the quantification of its progression and its monitoring post-treatment. In particular, we would like to apply these considerations to cardiac resynchronization therapy (CRT) studies, supported by the following paradigm: CRT targets the correction of the dyssynchrony in the motion of the cardiac chambers, leading to improvements in the cardiac function, the patient condition, and ventricular size (St John Sutton et al., 2003); however, CRT patient selection still discards mechanical dyssynchrony as a selection criteria (Dickstein et al., 2010). There are several reasons for this: there is currently no consensus regarding the accurate characterization of mechanical dyssynchrony, its link with CRT outcome, and the way to include it within the patient selection process (Fornwalt et al., 2009; Fornwalt, 2011; Delgado and Bax, 2011; Sung and Foster, 2011).

*corresponding author.
Address: c/Tànger 122-140, E08018 Barcelona, Spain.
Tel: +34 93-542-1348, Fax: +34 93-542-2083
Email address: n.duchateaucistib@gmail.com
(Nicolas Duchateau)

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

Recently, Parsai et al. (2009) discussed the advantages of considering specific groups of mechanical dyssynchrony in the CRT selection process. Each of these groups was associated to one specific pathological pattern of myocardial motion and deformation, with different grades of abnormality with respect to a healthy cardiac function. The relevance of similar patterns of mechanical dyssynchrony was also described in De Boeck et al. (2009); Szulik et al. (2010), using more quantitative measurements. Based on these findings, a straightforward improvement of CRT patient selection would be achieved through the recognition of such patterns in new CRT candidates, as the response rate of each of these patterns is roughly known. The same concept could also be applied to quantitatively grading the severity of the disease or the response to the therapy. Nonetheless, all the above-cited methods still lack reproducible tools to perform inter-subject comparison, as discussed in Fornwalt (2011), which limits their applicability.

1.2. Robust comparison through statistical atlases

Statistical atlases were initially designed for representing instances of a given population, by modelling the statistical distribution of anatomical and functional features within this population (Toga et al., 2006; Young and Frangi, 2009). In these frameworks, the data of each subject is normalized to a common anatomical reference, which allows reproducible intra- and inter- population comparison. Atlases of motion and deformation (Rougon et al., 2004; Rao et al., 2004; Chandrashekara et al., 2005) fit for the study of cardiac dyssynchrony, namely comparing the myocardial velocities and strain of individuals to a reference population (either healthy or with the same disease) used to build the atlas. In particular, Duchateau et al. (2011b) proposed a pipeline for the characterization of abnormal patterns of ventricular dyssynchrony, in comparison with a healthy population, using an atlas-based quantification of local myocardial motion abnormalities. However, the tools used for statistically modelling the variability of healthy subjects may not be adequate if the strategy is extended to build an atlas for a specific pattern of dyssynchrony. Variations of the pattern localization within the cardiac cycle and along the myocardial wall may bias the estimation of the local variability for this population, despite its synchronization to a common system of spatiotemporal coordinates, therefore requiring the inclusion of more complex statistics or pattern analysis techniques in the atlas construction.

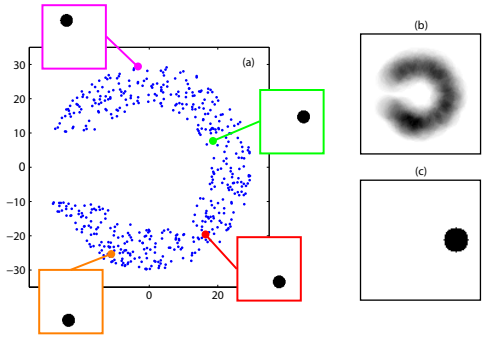


Figure 1: (a) Set of synthetic images representing a black disk over a white background. The maximum dimensionality of the space of images corresponds to the number of pixels of each image, but each synthetic image lies in a 2D subspace, the position of each disk being associated to a coordinate in a 2D space (blue dots). (b) Average image, without considering that data is arranged according to a 2D structure. (c) Fréchet mean image on the data-driven manifold, corresponding to coordinates $(25, 0)$.

1.3. Comparison to a given population

The comparison of individuals to a specific pattern of dyssynchrony can be seen as a two-steps approach: (1) finding a relevant space to model this pattern, and (2) defining a distance to this population through the mapping of any individual to this space.

The definition of an optimal space for representing and comparing populations has been extensively commented in the literature (Ashburner and Klöppel, 2011). A k-nearest neighbours comparison in the original space, or linear dimensionality reduction techniques may bias the analysis in case the modeled population lies on a non-linear space (Kim and Choi (2007); Tenenbaum et al. (2000) and Fig. 1).

In contrast, a manifold representation is particularly of interest for our application as it assumes that the global structure of the modeled population may be non-linear, and its parametrization allows inter-subject comparison along this structure.

Principal geodesic analysis (Fletcher et al., 2004) is a generalization of PCA for data lying on a manifold, but is described in cases where the manifold structure is already known and independent of the processed data.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

Manifold learning algorithms were proposed to estimate the low-dimensional structure of the studied dataset by taking into account the local properties of the input data in the dimensionality reduction process (Isomap (Tenenbaum et al., 2000), Local Linear Embedding (Roweis and Saul, 2000), Laplacian eigenmaps (Belkin and Niyogi, 2003), kernel-PCA (Schölkopf et al., 1998)).

The comparison of individuals to the modeled population requires mapping this new data to the manifold. Methods solving the "pre-image problem" (Mika et al., 1999; Kwok and Tsang, 2004; Zheng et al., 2010) estimate the image generated from the manifold that shares the same coordinates as the input image, but do not explicitly provide a formulation for the mappings between the ambient space and the manifold system of coordinates.

Several works provided an explicit formulation of the above-mentioned mappings, extending the concept of principal curves (Hastie and Stuetzle, 1989), a non-linear generalization of PCA, using kernel regression on scattered data. Gerber et al. (2009) demonstrated the analogy between these two concepts when the number of points tends to infinity. Bengio et al. (2004) proposed a framework to map new data to the manifold system of coordinates. The computation of the inverse mapping (estimating the image from the initial ambient space corresponding to a given manifold coordinate) was targeted in (Meinicke et al., 2005; Gerber et al., 2009, 2010; Etyngier et al., 2007; Davis et al., 2010).

1.4. Proposed approach

In this paper, we propose a framework for comparing the myocardial motion of individuals to a population with a specific pattern of abnormal motion, using manifold learning techniques to represent this population as a progressive deviation from normality. We extend manifold learning techniques to embed the definition of a relevant origin within the manifold, and use kernels with locally adjustable bandwidth to improve the accuracy of the mappings between the space of input images and the space of coordinates parameterizing the manifold.

The originality of our method resides in the use of 2D maps of local myocardial motion abnormalities as input, as introduced in Duchateau et al. (2011b). This highlights the presence of specific abnormal motion patterns, which can be represented by a manifold structure specific to each pattern, and allows the definition of a physiologically meaningful origin within the manifold, representing a normal motion pattern. Each pathological pattern is therefore considered a deviation from normality along a manifold structure. The proposed technique represents a step forward for patient comparison in clinical applications, as it facilitates the identification of the closest class a sample falls in (distance to the manifold) and the localization of this sample within the identified class (distance to normality along the manifold).

The method is applied in the context of CRT, comparing both healthy subjects and patients to a population with a specific pattern of intraventricular dyssynchrony called septal flash (SF) (Parsai et al., 2009). This pattern consists of a fast inward / outward radial motion of the septum during the isovolumic contraction period, contrasting with healthy hearts, which contract along the longitudinal direction (Fig. 2, 3 and 4). Gjesdal et al. (2011) interpreted SF as the result of active septal contraction during left bundle-branch block, which was shown to actively condition CRT outcome (Parsai et al., 2009; Duckett et al., 2012).

A preliminary version of this work was presented in Duchateau et al. (2011a), in which we illustrated the feasibility of such an approach. The current paper improves the whole methodology, from the atlas construction steps (use of temporal diffeomorphic free-form deformation, De Craene et al. (2012b)) to the patient comparison to the manifold (use of locally adjustable kernels) and in-depth tuning of the whole set of parameters, using both synthetic and real data.

2. Methods

The computation of a distance between individuals and a given population considered as a pathological deviation from normality consists of three steps: (1) the quantification of motion abnormalities for all the subjects in the dataset, (2) the estimation of a relevant manifold for the training population, constrained to pass by an origin representing normal motion, and (3) the mapping of any subject to the manifold.

2.1. Atlas-based computation of myocardial motion abnormalities

The input images for our method consist of 2D spatiotemporal maps of myocardial motion abnormalities, obtained from a statistical atlas of motion

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

built from healthy volunteers. The implementation used in the current paper for computing these maps improves the pipeline proposed in Duchateau et al. (2011b), as described in the following paragraphs.

2.1.1. Motion extraction

Myocardial velocities are extracted from the image sequence of each volunteer, at each point in time and space, using image-based registration. The diffeomorphic free-form deformation (FFD) implementation used in Duchateau et al. (2011b) provides piece-wise stationary velocities, which may be a limitation for the temporal synchronization step of the atlas building (Fig. 2). In our current implementation, we preferred the temporal diffeomorphic free-form deformation (TDFFD) method proposed in De Craene et al. (2012b), which enforces temporal consistency and provides differentiable velocities.

We used a multi-resolution implementation of the TDFFD, the initial grid size being of 5 x 3 control points in the spatial direction and 1 control point per frame in the temporal direction. Mean square error was used as similarity metric, combining a comparison to the first frame of the sequence and the comparison of consecutive frames (De Craene et al., 2011), with equal weights. The L-BFGS-B algorithm (Byrd et al., 1995) was chosen as optimizer for the whole registration procedure.

2.1.2. Atlas-based comparison to normality

The velocities extracted from each sequence are spatiotemporally synchronized to a common reference anatomy, using ECG matching (time) and local reorientation derived from image registration (multiscale diffeomorphic FFD, Rueckert et al. (2006)) between the first frames of each sequence (space). At each instant of the cardiac cycle and each point of the septum, the velocity vector of each individual is compared to the distribution of velocity vectors for the atlas population at this location, using the Mahalanobis distance. Average and covariance of myocardial velocities over the set of healthy volunteers encode a representation of normal motion. Assuming that the local distribution of velocities for the atlas population is Gaussian, the cumulative function associated to this distribution returns a two-tailed p-value, low p-value indicating high degree of abnormality. In contrast, a p-value close to 1 indicates normal motion.

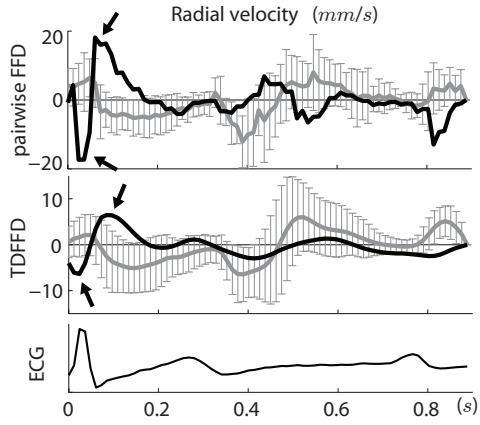


Figure 2: Radial velocities at one point of the septum (mid-inferoseptal level) during one cardiac cycle, for one CRT candidate with SF (black curve, arrows indicate the inward and outward events of SF), and the atlas of healthy volunteers (gray, average velocity ± standard deviation). Comparison between diffeomorphic FFD between each pair of consecutive frames, providing piece-wise stationary velocities (Duchateau et al., 2011b), and temporal diffeomorphic free-form deformation (TDFFD, De Craene et al. (2012b)), which enforces temporal consistency and provides differentiable velocities, all pairs of frames being considered simultaneously during the computation.

2.1.3. Representation of p-value maps of abnormalities

We chose to represent this spatiotemporal information by means of color-coded maps, in which the horizontal axis is time (systole) and the vertical one is the position along the septum (Fig. 3). Similar displays, inspired from anatomical M-mode echocardiographic images, were previously used in strain rate imaging (Heimdal et al., 1998) or speckle tracking applications (Marwick et al., 2007). In the current paper, each pixel value corresponds to the p-value index used to locally encode abnormality, in a logarithmic scale, multiplied by the sign of the radial velocity. The color-code associates blue and red color to highly abnormal inward and outward motion of the septum, respectively. This choice was made to highlight the inward and outward events of SF, in comparison with other patterns of left ventricular dyssynchrony (Fig. 4). According to these conventions, the origin used to constrain the manifold (Sec. 2.4.3) is defined as an image having 0 value at every pixel, representing a normal motion pattern.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

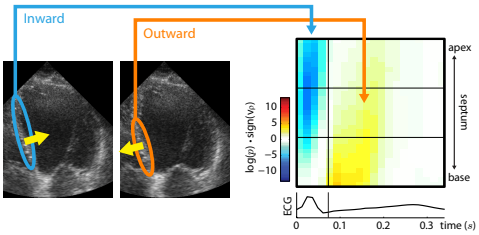


Figure 3: Map of septal motion abnormalities during systole, for one CRT candidate with SF. The color-scale encodes abnormality (p-value) in a logarithmic scale, multiplied by the sign of the radial velocity $\mathbf{v}_\rho$ to highlight the inward and outward events of SF.

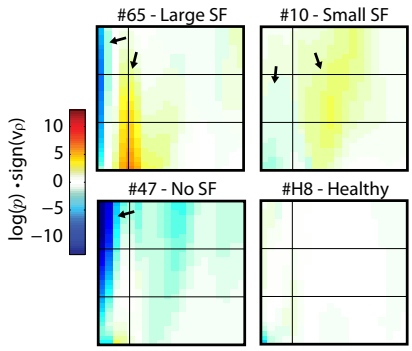


Figure 4: Abnormality maps for four different subjects: two with SF (top row, arrows indicate the inward and outward events of SF), one with left-right interaction (bottom left), a type of inter-ventricular dyssynchrony (Parsai et al., 2009) where the septum moves passively inward, as indicated by the black arrow, and one healthy volunteer (bottom right).

2.2. Definition of variables

Each of the abnormality maps described in Sec. 2.1 corresponds to one subject in the dataset, and is used as a 2D input image for the manifold learning process. The rationale of manifold learning on maps of statistical significance is discussed in Sec. 4.

All the images considered in this paper belong to an ambient space $\mathcal{A} \subset \mathbb{R}^P$, P being the number of pixels of each image. We denote $\mathcal{I} = \{\mathbf{I}_0, ..., \mathbf{I}_N\} \subset \mathcal{A}$ the dataset of $N+1$ images used for the manifold estimation. The image $\mathbf{I}_0$ corresponds to the image origin for normality. This is a synthetic image with 0 value at each pixel, as explained in Sec. 2.1.3. This image is added to the original dataset $\{\mathbf{I}_1, ..., \mathbf{I}_N\}$ before any computation. Every image $\mathbf{I}_i, \quad i > 0$ is therefore connected to $\mathbf{I}_0$ through the isomap graph resulting from the computations described below (Sec. 2.3). This amounts to considering every element of $\mathcal{I}$ as a deviation from the origin along a specific path on the manifold structure.

The coordinate space of the surface estimating the manifold is denoted $\mathcal{C} \subset \mathbb{R}^M$, M being the estimated dimensionality of the manifold. Note that $M < N+1$ due to its estimation from a training set of $N+1$ images, this size being lower than P, the number of pixels of each image. We denote $f : \mathcal{A} \to \mathcal{C}$ and $g : \mathcal{C} \to \mathcal{A}$ the correspondence functions between $\mathcal{A}$ and $\mathcal{C}$. An approximation of these functions can be obtained using interpolation techniques (Sec. 2.4), inspired from the concept of principal curves (Hastie and Stuetzle, 1989).

We preferred to keep notations general for the formulation of our problem, and we denote $S_\mathcal{A} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$ and $S_\mathcal{C} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}^+$ the metrics used to compare elements of $\mathcal{A}$ and $\mathcal{C}$, respectively. In our implementation, we used the Euclidean distance for both metrics $S_\mathcal{A}$ and $S_\mathcal{C}$, for both synthetic and CRT datasets, this choice being discussed in Sec. 4.

2.3. Manifold learning through Isomap

The isomap algorithm (Tenenbaum et al., 2000) is used to estimate the manifold (Fig. 5a). First, a graph is built for the dataset $\mathcal{I}$, based on the k-NN algorithm, connecting all the images among themselves according to the metric $S_\mathcal{A}$. The geodesic distance between two points, defined as the shortest path connecting these two points, along the graph, is denoted $d_{k\text{NN}}$. Then, Euclidean embedding of this k-NN connected graph provides a set of coordinates $\mathcal{X} = \{\mathbf{x}_0, ..., \mathbf{x}_N\} \subset \mathcal{C}$.

2.4. Mapping new patients: from $\mathcal{A}$ to $\mathcal{C}$

The estimation of $f : \mathcal{A} \to \mathcal{C}$ can be formulated as an interpolation problem, which can be re-defined on a reproducible kernel Hilbert space (RKHS) (Saitoh, 1988; Aronszajn, 1950; Glaunes, 2005) $\mathcal{F}$ of functions $\mathcal{A} \to \mathcal{C}$. The solution of this problem belongs to a subspace $\mathcal{F}_\mathcal{I} \subset \mathcal{F}$, where the pairs $(\mathbf{I}, \mathbf{c}) \in \mathcal{A} \times \mathcal{C}$ are constrained to be located in a specific set $\mathcal{I} \times \mathcal{X}$.

In our case, $\mathcal{I} = \{\mathbf{I}_i\}_{i \in [0,N]}$ is finite-dimensional. As commented in Durrleman (2010), $\mathcal{F}_\mathcal{I}$ is therefore finite-dimensional, and f can be searched in the closure of the set spanned by the vectors:

$$\forall \mathbf{J} \in \mathcal{A}, \; K_\mathcal{F}(\mathbf{J}, \mathbf{I}_i) \cdot \mathbf{c}_i, \quad (\mathbf{I}_i, \mathbf{c}_i) \in \mathcal{I} \times \mathcal{C}, \qquad (1)$$

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

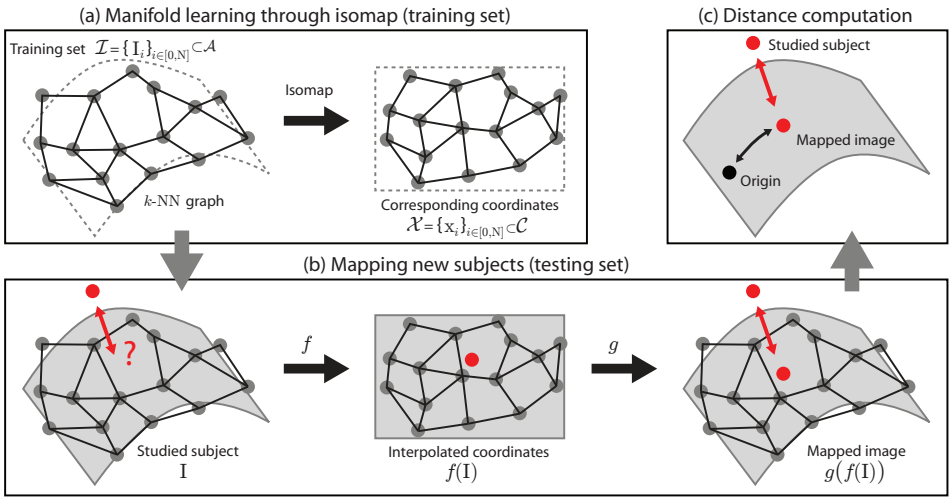


Figure 5: Pipeline for the method presented in this paper.

where $K_{\mathcal{F}} : \mathcal{A} \times \mathcal{A} \to \mathbb{M}_{M,M}$ is the reproducible kernel of $\mathcal{F}$, and $\mathbb{M}_{M,M}$ is the set of $M \times M$-dimensional real-valued matrices.

Thus, the norm in the RKHS (finite-dimensional case) can be defined as:

$$\|f\|_{\mathcal{F}} = \sum_{i=0}^{N}\sum_{j=0}^{N} \mathbf{c}_i^t \cdot K_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j) \cdot \mathbf{c}_j, \qquad (2)$$

where $.^t$ is the transposition operator.

The interpolation problem can be formulated as an exact or inexact matching problem, depending if we assume that the manifold should pass exactly by the training set, or if we tolerate some dispersion of the data around the manifold. The formulations of this problem, detailed in the following sections, have analytical solutions which can be written as expansions in terms of the training samples (nonparametric representer theorem, Schölkopf et al. (2001)). We also refer the reader to Kimeldorf and Wahba (1970) for further details on the formulation of the solution.

The choice of the kernel determines the RKHS in which the problem is solved. This point is discussed in App. A.

In the following subsections, we describe the basic formulation of this problem and the improvements made to adapt it to our framework.

2.4.1. Exact matching

In case the interpolation is formulated as an exact matching problem, we look for the optimal $f : \mathcal{A} \to \mathcal{C}$ passing by all the coordinates $\mathbf{x}_i \in \mathcal{X}$ at the data points $\mathbf{I}_i \in \mathcal{I}$:

$$\begin{cases} \underset{f\in\mathcal{F}}{\operatorname{argmin}}\left(\frac{1}{2}\|f\|_{\mathcal{F}}^2\right), \\ \text{under the constraint } \; f(\mathbf{I}_i) = \mathbf{x}_i, \; \forall i \in [0, N]. \end{cases} \qquad (3)$$

As commented in Sec. 2.4 and App. B, the solution for Eq. 3 can be written as:

$$\begin{cases} f(\mathbf{I}) = \sum_{i=0}^{N} k_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i, \\ \text{with } \mathbf{C} = \mathbf{K}_{\mathcal{I}}^{-1} \cdot \mathbf{X}, \end{cases} \qquad (4)$$

where:
- $\mathbf{C} = (\mathbf{c}_0, \ldots, \mathbf{c}_N)^t \in \mathbb{M}_{N+1,M}$, with $\mathbf{c}_i \in \mathcal{C}$,
- $\mathbf{K}_{\mathcal{I}} = \left(k_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j)\right)_{(i,j)\in[0,N]^2} \in \mathbb{M}_{N+1,N+1}$,
- $\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_N)^t \in \mathbb{M}_{N+1,M}$.

In this paper, the scalar function $k_{\mathcal{F}}$ defining the kernel $K_{\mathcal{F}}$ is chosen of the exponential form:

$$k_{\mathcal{F}}(\mathbf{I}, \mathbf{J}) = \exp\left(-S_{\mathcal{A}}(\mathbf{I}, \mathbf{J})^2/\sigma_{\mathcal{F}}^2\right), \qquad (5)$$

$\sigma_{\mathcal{F}}$ being its bandwidth and $(\mathbf{I}, \mathbf{J}) \in \mathcal{A}^2$. The choice of such a function is commented in App. A, while the choice of a relevant bandwidth is discussed in Sec. 2.4.4.

2.4.2. Inexact matching

In case we tolerate some dispersion of the data around the manifold, the previous interpolation

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

needs to be re-written as an inexact matching problem, now looking for the optimal $f : \mathcal{A} \to \mathcal{C}$ best approximating the coordinates $\mathbf{x}_i \in \mathcal{X}$ at the data points $\mathbf{I}_i \in \mathcal{I}$:

$$\underset{f \in \mathcal{F}}{\operatorname{argmin}} \Big(\frac{1}{2} \|f\|^2_{\mathcal{F}} + \frac{\gamma_f}{2} \sum_{i=0}^{N} S_{\mathcal{C}}\big(f(\mathbf{I}_i), \mathbf{x}_i\big)^2 \Big), \tag{6}$$

where γ_f is a weighting coefficient balancing the smoothness of the interpolation and the adherence to the data.

Eq. 6 has the following analytical solution:

$$\begin{cases} f(\mathbf{I}) = \sum_{i=0}^{N} k_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i, \\ \text{with } \mathbf{C} = \big(\mathbf{K}_{\mathcal{I}} + \frac{1}{\gamma_f} \mathbf{Id}_{N+1,N+1}\big)^{-1} \cdot \mathbf{X}, \end{cases} \tag{7}$$

$\mathbf{Id}_{N+1,N+1}$ being the identity matrix in $\mathbb{M}_{N+1,N+1}$.

2.4.3. Constrained problem

In this paper, the formulation of Eq. 6 is adapted to force the interpolation function to pass by the coordinates origin $\mathbf{x}_0$:

$$\begin{cases} \underset{f \in \mathcal{F}}{\operatorname{argmin}} \Big(\frac{1}{2} \|f\|^2_{\mathcal{F}} + \frac{\gamma_f}{2} \sum_{i=1}^{N} S_{\mathcal{C}}\big(f(\mathbf{I}_i), \mathbf{x}_i\big)^2 \Big), \\ \text{under the constraint } \; f(\mathbf{I}_0) = \mathbf{x}_0. \end{cases} \tag{8}$$

The analytical solution for this problem is written as:

$$\begin{cases} f(\mathbf{I}) = \sum_{i=0}^{N} k_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i, \\ \text{with } \mathbf{C} = \big(\mathbf{K}_{\mathcal{I}} + \frac{1}{\gamma_f} \mathbf{M}\big)^{-1} \cdot \mathbf{X}, \end{cases} \tag{9}$$

where $\mathbf{M} = \big(M_{i,j}\big)_{(i,j) \in [0,N]^2} \in \mathbb{M}_{N+1,N+1}$, with $M_{i,i} = 1 \; \forall i \neq 0$ and 0 otherwise.

With this formulation,

$$\mathbf{x}_i = \sum_{j=0}^{N} \big(k_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j) + \frac{1}{\gamma_f} M_{i,j} \big) \cdot \mathbf{c}_j. \tag{10}$$

Thus, $\mathbf{x}_0 = \sum_{j=0}^{N} k_{\mathcal{F}}(\mathbf{I}_0, \mathbf{I}_j) \cdot \mathbf{c}_j$, which corresponds to the exact matching formulation of Eq. 4, meaning that the constraint $f(\mathbf{I}_0) = \mathbf{x}_0$ is satisfied.

The addition of such a constraint is illustrated in Fig. 6, which displays the curve interpolated from a 1D synthetic dataset, using inexact matching, before and after forcing the curve to pass by a given point, as described in Eq. 8.

2.4.4. Use of locally adjustable kernel

In the previous formulations, no constraint is made on the kernel bandwidth, which fully conditions the accuracy of the interpolation. Previous

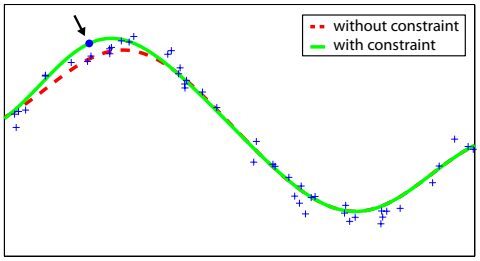


Figure 6: Interpolation of a 1D synthetic dataset using inexact matching, before and after the addition of a constraint forcing the curve to pass by the point indicated by the black arrow.

works used a fixed bandwidth, defined as average k-NN distance, namely:

$$\sigma_{\mathcal{F}} = \frac{1}{N+1} \sum_{i=0}^{N} S_{\mathcal{A}}(\mathbf{I}_i, \mathrm{nn}_k(\mathbf{I}_i)), \tag{11}$$

where $\mathrm{nn}_k(\mathbf{I}_i)$ is the kth neighbour of $\mathbf{I}_i$ (Gerber et al., 2010).

In fact, the use of a fixed bandwidth has some limitations in case the points distribution is not uniform, which is our case (Sec. 3.1.2 and 3.2.3). Indeed, a small kernel in a sparse region of the dataset would result in mapping points in this region to zero, or closer to zero than they should. On the contrary, a kernel with a too large bandwidth could result in a too planar interpolation in comparison with the manifold curvature, therefore mapping points far from the manifold structure. Similar concerns were raised in the literature about probabilistic density estimation (Scott, 1992; Silverman, 1986). The manifold learning algorithm described in Zhang et al. (2012) used a linear interpolation scheme with equal weights on each neigbourhood, proposing to locally adapt the number of k-NN to adress this issue. In the sequel, we propose to use a varying bandwidth for the interpolation kernel, which is locally adapted depending on the neighbourhood size:

$$\sigma_{\mathcal{F}}(\mathbf{I}) = \frac{1}{K^2} \sum_{k=1}^{K} \sum_{\substack{l=1 \\ l \neq k}}^{K} S_{\mathcal{A}}(\mathrm{nn}_k(\mathbf{I}), \mathrm{nn}_l(\mathbf{I})), \tag{12}$$

where the right term in the equation is the average distance between the K nearest neighbours of $\mathbf{I}$.

The advantage of using such a varying bandwidth over a formulation with a fixed bandwidth is demonstrated in Sec. 3.1.2. Note however that

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

this formulation has a limitation with respect to the interpolation problem on a RKHS, as the chosen kernel (and its bandwidth) determines the RKHS where the problem is solved (Sec. 2.4). The proposed formulation with a varying bandwidth can be seen as a deviation from the original problem formulated on a RKHS, and would require further investigation to conciliate this technical solution with the original formulation.

2.5. Mapping new patients: from $\mathcal{C}$ to $\mathcal{A}$

The formulation for the estimation of $g : \mathcal{C} \to \mathcal{A}$ is similar to the one of $f : \mathcal{A} \to \mathcal{C}$, namely a matching problem on a RKHS $\mathcal{G}$ of functions $\mathcal{C} \to \mathcal{A}$, equipped with a norm $\|.\|_{\mathcal{G}}$, which can be defined as:

$$\|g\|_{\mathcal{G}} = \sum_{i=0}^{N} \sum_{j=0}^{N} \mathbf{a}_i^t \cdot K_{\mathcal{G}}(\mathbf{x}_i, \mathbf{x}_j) \cdot \mathbf{a}_j, \tag{13}$$

where $K_{\mathcal{G}} : \mathcal{C} \times \mathcal{C} \to \mathbb{M}_{P,P}$ is the reproducible kernel of $\mathcal{G}$, and $\mathbf{a}_i \in \mathcal{A}$. The problem can be formulated in a similar manner to Sec. 2.4.3, interchanging the roles of the images and the coordinates:

$$\begin{cases} \underset{g \in \mathcal{G}}{\operatorname{argmin}} \Big(\frac{1}{2}\|g\|_{\mathcal{G}}^2 + \frac{\gamma_g}{2} \sum_{i=1}^{N} S_{\mathcal{A}}\big(g(\mathbf{x}_i), \mathbf{I}_i\big)^2 \Big), \\ \text{under the constraint } \; g(\mathbf{x}_0) = \mathbf{I}_0, \end{cases} \tag{14}$$

where γ_g is a weighting coefficient balancing the smoothness of the interpolation and the adherence to the data.

Eq. 14 has the following analytical solution:

$$\begin{cases} g(\mathbf{x}) = \sum_{i=0}^{N} k_{\mathcal{G}}(\mathbf{x}, \mathbf{x}_i) \cdot \mathbf{a}_i, \\ \text{with } \mathbf{A} = \big(\mathbf{K}_{\mathcal{X}} + \frac{1}{\gamma_g}\mathbf{M}\big)^{-1} \cdot \mathbf{Y}, \end{cases} \tag{15}$$

where:
- $\mathbf{A} = (\mathbf{a}_0, \ldots, \mathbf{a}_N)^t \in \mathbb{M}_{N+1,P}$, with $\mathbf{a}_i \in \mathcal{A}$,
- $\mathbf{K}_{\mathcal{X}} = \big(k_{\mathcal{G}}(\mathbf{x}_i, \mathbf{x}_j)\big)_{(i,j)\in[0,N]^2} \in \mathbb{M}_{N+1,N+1}$,
- and $\mathbf{Y} = (\mathbf{I}_0, \ldots, \mathbf{I}_N)^t \in \mathbb{M}_{N+1,P}$.

The scalar function $k_{\mathcal{G}}$ defining the kernel $K_{\mathcal{G}}$ is also chosen of the exponential form:

$$k_{\mathcal{G}}(\mathbf{x}, \mathbf{y}) = \exp\big(- S_{\mathcal{C}}(\mathbf{x}, \mathbf{y})^2 / \sigma_{\mathcal{G}}^2 \big), \tag{16}$$

with $(\mathbf{x}, \mathbf{y}) \in \mathcal{C}^2$, its bandwidth being defined as:

$$\sigma_{\mathcal{G}}(\mathbf{x}) = \frac{1}{K^2} \sum_{k=1}^{K} \sum_{\substack{l=1 \\ l \neq k}}^{K} S_{\mathcal{C}}(\mathrm{nn}_k(\mathbf{x}), \mathrm{nn}_l(\mathbf{x})). \tag{17}$$

2.6. Distance computation

With the previous formulations of the mappings f and g, any image $\mathbf{I} \in \mathcal{A}$ is associated to another image $\hat{\mathbf{I}}$ belonging to the manifold, by means of the composition of these mappings, using:

$$\hat{\mathbf{I}} = g\big(f(\mathbf{I})\big). \tag{18}$$

This composition allows defining a distance between any image $\mathbf{I} \in \mathcal{A}$ and the manifold (Gerber et al., 2010), namely:

$$d_P(\mathbf{I}) = S_{\mathcal{A}}(\hat{\mathbf{I}}, \mathbf{I}). \tag{19}$$

This distance is complemented by a second one, which compares individuals to normality along the manifold structure:

$$d_M(\mathbf{I}) = S_{\mathcal{C}}\big(f(\mathbf{I}), \mathbf{x}_0\big). \tag{20}$$

The relation between these two distances and the total abnormality contained in each map $\mathbf{I}$, defined as $S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0)$, is discussed in App. C.

2.7. Additional metrics

We used the objective metrics described in Davies et al. (2010) to evaluate the quality of the dimensionality reduction, in comparison with linear dimensionality reduction techniques.

2.7.1. Compactness

Compactness estimates the convergence speed of the dimensionality reduction, namely its ability to represent the studied objects by the very first dimensions. We measure compactness as:

$$C(M) = \frac{1}{\Gamma} \sum_{m=1}^{M} \lambda_m, \tag{21}$$

where M is the number of retained dimensions for the coordinate space $\mathcal{C}$, λ_m is the eigenvalue corresponding to the mth dimension, and $\Gamma = \sum_{m=1}^{N+1} \lambda_m$. Note that the formulation of Eq. 21 does not include any absolute value or norm, meaning that a monotonically increasing compactness curve reflects the absence of negative eigenvalues, as commented in Sec. 3.2.2.

The eigenvalues for the non-linear dimensionality reduction technique (ML) are obtained from the diagonalization of the k-NN distance matrix $\tau(\mathbf{D}) \in \mathbb{M}_{N+1,N+1}$ involved in the isomap process (Tenenbaum et al., 2000), where $\tau(\mathbf{D}) = -\mathbf{HDH}/2$,

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

with $\mathbf{D} = \left(d_{k\mathrm{NN}}(\mathbf{x}_i, \mathbf{x}_j)^2\right)_{(i,j)\in[0,N]^2}$, and $\mathbf{H} = \left(\delta_{ij} - 1/(N+1)\right)_{(i,j)\in[0,N]^2}$ is a centering operator. Each λ_m is associated to the mth principal direction along the manifold structure.

The eigenvalues for the linear dimensionality reduction technique (PCA) come from the diagonalization of the covariance matrix computed for the images of the training set $\mathcal{I}$. Each λ_m is associated to the mth principal direction along which the data variance is maximal.

The standard error of $C(M)$ is defined as:

$$\sigma_C(M) = \frac{1}{\Gamma}\sum_{m=1}^{M}\sqrt{\frac{2}{N+1}}\lambda_m. \tag{22}$$

2.7.2. Generalization ability

The generalization ability estimates the reconstruction error for points included within the range of noise of the training set. It represents the ability of the method to describe instances outside the training set. Each point in the training set $\mathbf{I}_i \in \mathcal{I}$ is reconstructed using leave-one-out, namely, estimating the space of reduced dimensionality from the other points in the dataset. This leads to the following reconstruction error:

$$G(\mathbf{I}_i, M) = S_{\mathcal{A}}(\hat{\mathbf{I}}_i^{LVO}, \mathbf{I}_i), \tag{23}$$

where M is the number of retained dimensions, and $\hat{\mathbf{I}}_i^{LVO}$ is the reconstruction of $\mathbf{I}_i$ using the reduced set $\mathcal{I}\backslash\mathbf{I}_i = \{\mathcal{I}_j\}_{j\neq i}$.

We computed the median, the 1st and 3rd quartiles of $G(\mathbf{I}_i, M)$ for all $\mathbf{I}_i \in \mathcal{I}$ to fully characterize the generalization ability, the normality of its distribution not being guaranteed.

Compactness and generalization ability both reflect the amount of variance explained by the model, but these metrics are not redundant. In the PCA case, the leave-one-out process apart, the reconstruction error (generalization ability) reflects the amount of variance explained by the set of of retained eigenvalues (compactness). However, as illustrated in Fig. 20, the generalization ability highly depends on the way the training set is distributed (linear or non-linear), and the validity of the above assertion (the "generalization ability reflects the amount of variance explained") is therefore guaranteed only in the linear case.

2.7.3. Specificity

Specificity characterizes the relevance of objects generated from the low dimensional coordinates,

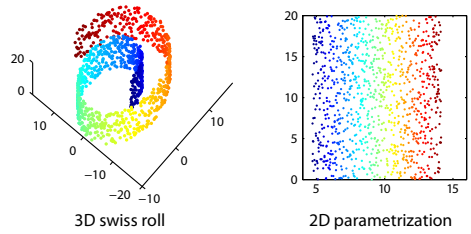


Figure 7: 3D swiss roll and 2D random distribution of points used as 2D parametrization for its construction.

with respect to the training set:

$$S(\mathbf{x}, M) = S_{\mathcal{A}}\big(g(\mathbf{x}), \mathrm{nn}_1(g(\mathbf{x}))\big), \tag{24}$$

where $g(\mathbf{x})$ is the image generated from the M-dimensional coordinate $\mathbf{x}$, and $\mathrm{nn}_1(g(\mathbf{x}))$ is its first nearest neighbour.

We computed the median, the 1st and 3rd quartiles of $S(\mathbf{x}, M)$ over a set of randomly generated coordinates $\mathbf{x}$ to fully characterize the specificity, the normality of its distribution not being guaranteed.

3. Experiments

In the sequel, we detail the experiments designed for tuning the parameters of our method, namely: the manifold dimensionality, the number of k-NN, the bandwidth of the interpolation kernels, and the inexact matching weights γ_f and γ_g. We first use a synthetic dataset to understand the behaviour of some specific parameters. Then, these parameters are tuned for a real dataset of CRT candidates. The algorithm with optimal values is finally applied to compare a set of individuals to a specific population.

3.1. Parameter tuning - Synthetic data

We created a three dimensional dataset of 1000 points, arranged according to a 2D structure in the 3D space (swiss roll, Fig. 7), defined as:

$$\begin{cases} I_{i,1} = \cos\big(3\pi/2\cdot(1+2x_{i,1})\big), \\ I_{i,2} = \sin\big(3\pi/2\cdot(1+2x_{i,2})\big), \\ I_{i,3} = x_{i,2} \in [0, 20], \end{cases} \tag{25}$$

$(I_{i,1}, I_{i,2}, I_{i,3}) = \mathbf{I}_i$ referring to a point in the 3D space, obtained from the coordinate $(x_{i,1}^{GT}, x_{i,2}^{GT}) = \mathbf{x}_i^{GT}$, randomly generated from a uniform distribution. In the following, we denote $g^{GT} : \mathcal{C} \to \mathcal{A}$

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

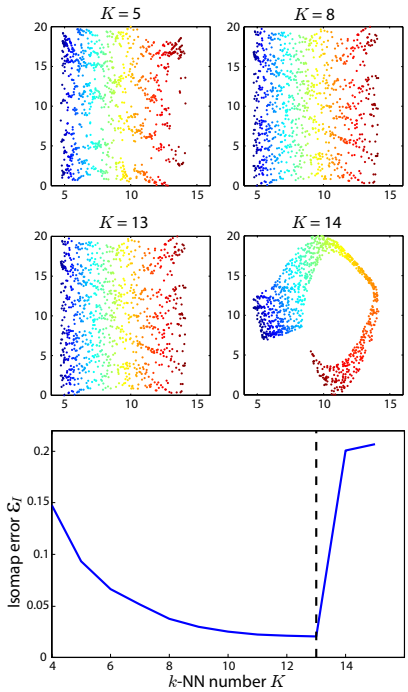


Figure 8: Influence of the number of k-NN on the isomap output, for the 3D swiss roll data. Top: isomap output for $K = \{5, 8, 13, 14\}$. Bottom: evolution of the isomap error ϵ_I (Eq. 26) with the number of k-NN. Mean error over all the pairs $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}^2$. A jump is present on this curve for $K = 14$, reflecting the apparition of a short-circuit in the k-NN graph.

the parametrization function allowing the generation of points in the 3D space from ground truth coordinates, and Eq. 25 can be rewritten as $\mathbf{I}_i = g^{GT}(\mathbf{x}_i^{GT})$.

For this synthetic dataset, the Euclidean distance $\|.\|$ was used for both metrics $S_{\mathcal{A}}$ and $S_{\mathcal{C}}$.

3.1.1. Number of k-NN

The output from the isomap algorithm is a low-dimensional approximation of a Riemannian space. We used the metric described in Tenenbaum and Langford (2002) to estimate the quality of this approximation. The error in the approximation of the path between two coordinates $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}^2$ is defined as the relative difference between the geodesic distance $d_{k\mathrm{NN}}$, and the distance defined by $S_{\mathcal{C}}$:

$$\epsilon_I(\mathbf{x}_i, \mathbf{x}_j) = 1 - \frac{S_{\mathcal{C}}(\mathbf{x}_i, \mathbf{x}_j)}{d_{k\mathrm{NN}}(\mathbf{x}_i, \mathbf{x}_j)}. \qquad (26)$$

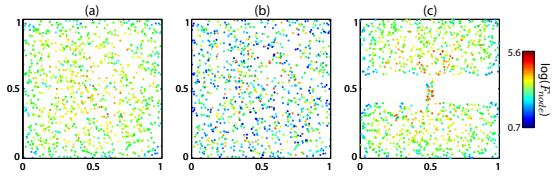


Figure 9: Influence of the graph number of k-NN and the spatial arrangement of the graph nodes on the the node flow distribution. Synthetic dataset of 1000 points within a given spatial domain. (a) Initial spatial domain with $K = 20$. (b) Same spatial domain with $K = 5$. Large local variations of the node flow are locally observed, reflecting a too low connectivity within the graph. (c) Modified spatial domain with $K = 20$. Node flow is higher where the spatial domain is narrow, reflecting that a large number of shortest paths along the graph pass within this region.

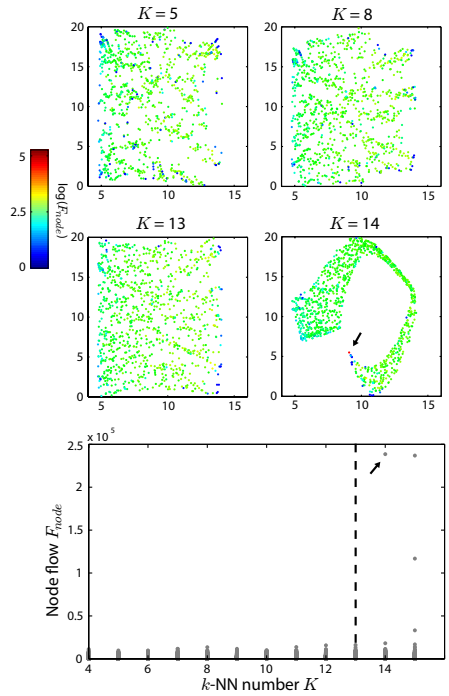


Figure 10: Influence of the number of k-NN on the node flow, for the 3D swiss roll data. Top: isomap output for $K = \{5, 8, 13, 14\}$, where each node is colored according to its flow (Eq. 27). For all points, the variability of the flow distribution is low when $K \leq 13$, while the apparition of a short-circuit for $K = 14$ makes one specific point having almost the highest flow possible (black arrow). Bottom: evolution of the node flow distribution with the number of k-NN. Black arrow indicates the value of the node flow for the point where the short-circuit appears.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

We evaluated the influence of the number of k-NN on the isomap output using this error. The experiment is illustrated in Fig. 8, and displays the mean error over all the pairs $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}^2$. When the number of k-NN increases, the 2D distribution of points estimated by the isomap algorithm gets closer to the 2D parametrization used for building the swiss roll, until a short-circuit appears for $K = 14$, visible both in the curve representing the isomap error ϵ_I and in the subfigure of the isomap output. The optimal number of k-NN for this dataset is therefore $K = 13$.

The error ϵ_I could also be used for parameter tuning in case the dimensionality M is unknown (Sec. 3.2), but this strategy may not be optimal to assess the presence or absence of a short-circuit, in particular if the number of samples in the training set is low. We adapted the measurement of node flow on a graph (Choi and Choi, 2007) to assess the apparition of a short-circuit in the k-NN graph. We first defined the flow of a given edge $F_{edge}(\text{edge}_{\mathbf{x}_i \mapsto \mathbf{x}_j})$ as the number of shortest paths $\left(\text{path}_{\mathbf{x}_r \mapsto \mathbf{x}_s}\right)_{(r,s) \in [0,N]^2}$ of the graph passing on the edge. The total flow at a node $\mathbf{x}_i \in \mathcal{X}$ is therefore defined as:

$$F_{node}(\mathbf{x}_i) = \sum_{k=1}^{K} F_{edge}(\text{edge}_{\mathbf{x}_i \mapsto \text{nn}_k(\mathbf{x}_i)}). \tag{27}$$

The node flow reflects the spatial arrangement of the nodes of the graph, conditioned by the number of k-NN used. Optimal K should minimize ϵ_I within the spatial domain occupied by the nodes of the graph. The uniformity of the spatial arrangement of the nodes (K being set to its optimal value) leads to the uniformity of the node flow distribution. In the absence of short-circuit, the local density of the graph and local variations of the spatial domain occupied by the nodes of the graph may change the node flow locally (Fig. 9). The evolution of the node flow distribution with the number of k-NN, for all the points of the k-NN graph, is shown in Fig. 10. Low variability of the node flow distribution is observed when the estimated coordinates $\mathbf{x}_i$ tend to be uniformly distributed (K getting closer to 13). Note that the points on the border of the graph have slightly lower node flow, due to a lower probability that a shortest path passes by these points. When a short-circuit appears at a specific point ($K = 14$, black arrow), a majority of the shortest paths pass by this point, which has therefore a much higher node flow.

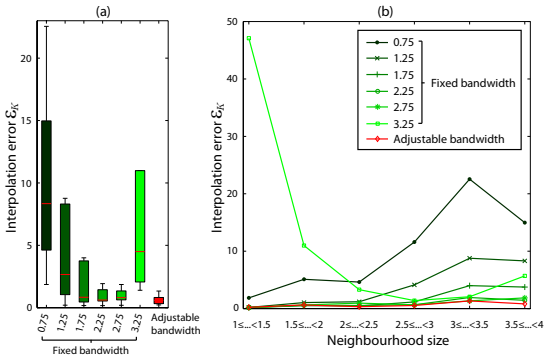


Figure 11: Comparison of locally adjustable kernel (red) and fixed bandwidth kernels (green levels) in terms of interpolation error ϵ_K (Eq. 29), for 100 testing points belonging to the swiss roll, generated using Eq. 25. (a) Median and 1st/3rd quartiles for each kernel size. (b) Median of the interpolation error ϵ_K for the points within a given interval of neighbourhood size (right term in Eq. 12), highlighting the accuracy of the locally adjustable kernel for each type of neighbourhood.

3.1.2. Varying kernel

The accuracy of an interpolation based on a locally adjustable kernel (Sec. 2.4.4) was compared to the one of kernels with fixed bandwidth. We first generated a testing set of 100 points belonging to the swiss roll (note that these points are different from the already existing set of 1000 points generated in Sec. 3.1), as follows:

1. We first generated 100 ground truth coordinates $\mathbf{x}^{GT}$, randomly obtained from a uniform distribution.
2. For each ground truth coordinate, we computed the weighted center of mass of its neighbourhood:

$$W(\mathbf{x}^{GT}, \mathbf{p}) = \sum_{k=1}^{K} p_k \cdot \text{nn}_k(\mathbf{x}^{GT}), \tag{28}$$

   where $\mathbf{p}$ is a vector of weights p_k, randomly generated for each coordinate $\mathbf{x}^{GT}$ from a uniform distribution, such that $\sum_{k=1}^{K} p_k = 1$.
3. Finally, we computed the point on the swiss roll associated to this center of mass, using the parametrization described in Sec. 3.1, namely: $\mathbf{I} = g^{GT}(W(\mathbf{x}^{GT}, \mathbf{p}))$.

The kernel interpolation is accurate if the point $\mathbf{I} = g^{GT}(W(\mathbf{x}^{GT}, \mathbf{p}))$ is mapped to the weighted center of mass of the coordinates $f(\text{nn}_k(\mathbf{I}))$, with the same weights p_k. We defined the interpolation

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

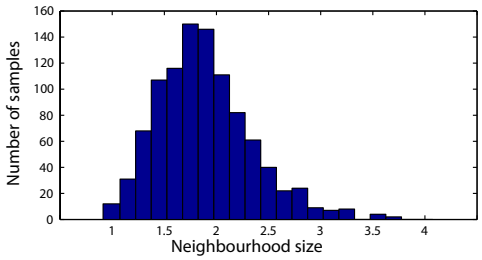


Figure 12: Histogram of the distribution of neighbourhood sizes (right term in Eq. 12) for the 3D swiss roll dataset. Non-uniformity in the data distribution supports the use of a kernel with varying bandwidth for this dataset.

error as:

$$\epsilon_K(\mathbf{I}) = S_{\mathcal{C}}\Big(f(\mathbf{I}), \sum_{k=1}^{K} p_k f(\mathrm{nn}_k(\mathbf{I}))\Big). \tag{29}$$

The median and 1st/3rd quartiles of the interpolation error ϵ_K over the generated set of 100 points, for each kernel size, is represented in Fig. 11a, and is complemented by Fig. 11b, representing the median of the interpolation error ϵ_K for the points within a given interval of neighbourhood size (right term in Eq. 12). As commented in Sec. 2.4.4, small kernels (dark green) introduce more errors for large neighbourhoods, while on the contrary, larger kernels (light green) are less accurate for small neighbourhoods. In contrast, the proposed kernel (red) with adjustable bandwidth results in an accurate interpolation for any neighbourhood size. The performance of kernels with fixed bandwidth of 2.25 and 2.75 is very similar, probably as a consequence of this specific dataset (100 randomly selected points, using randomly generated testing weights p_k).

The relevance of using a kernel with varying bandwidth depends on the extent of non-uniformity in the data distribution. This was assessed by representing the histogram of the distribution of neighbourhood sizes (right term in Eq. 12) over the studied dataset, which supports the use of such a kernel (Fig. 12).

3.2. Parameter tuning - CRT data

3.2.1. Dataset description

Using the method presented in Sec. 2.3, a manifold was estimated from a population of 50 CRT candidates with SF. This manifold is expected to represent pathological deviations from normal motion, each point of the manifold being a SF pattern. The parameter optimization described in the following sections was performed on this population. Justifications about the size of this population are given in Sec. 3.2.5. Note that this number is conditioned by the number of patients undergoing CRT in the hospital from which the data was collected (Hospital Clínic, Barcelona, Spain), and among them, by the presence of SF for these patients (generally half of the population undergoing CRT). As a comparison, the clinical study of Parsai et al. (2009), which motivated our approach, considered a population of 161 patients, 87 of which had SF.

A second population was used for testing the distances proposed in Sec. 2.6, as described in Sec. 3.3. This population was made of 37 CRT candidates (6 having SF and 31 without SF) and 21 healthy volunteers. All patient data was acquired before the implantation of the CRT device. The presence of SF was assessed by two experienced cardiologists, from the visual inspection of echocardiographic M-mode images, as described in Parsai et al. (2009).

A 2D spatiotemporal map of myocardial motion abnormalities obtained from a statistical atlas of motion (Duchateau et al., 2011b) was associated to each subject, as explained in Sec. 2.1. The atlas was built from the set of 21 healthy volunteers. Abnormality maps for the set of volunteers used for the atlas construction were computed using leave-one-out on this population. The abnormality maps had a size of 20×31 pixels, corresponding to the sampling of the systolic period (horizontal dimension) and the septum along its medial line (vertical dimension), respectively.

For this dataset, the Euclidean distance $\|.\|$ was used for both metrics $S_{\mathcal{A}}$ and $S_{\mathcal{C}}$. The choice of this distance for $S_{\mathcal{A}}$ is discussed in Sec. 4. This distance was used for the metric $S_{\mathcal{C}} : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}^+$ due to the Euclidean embedding of the coordinate space provided by the isomap algorithm (Sec. 2.3).

3.2.2. Dimensionality reduction and k-NN

An overview of methods for estimating the intrinsic dimensionality of a dataset was given in Camastra (2003), but there is no standard manner of performing this step. For the CRT dataset, both optimal dimensionality and number of k-NN were unknown, and were determined using the same experimental design as in Sec. 3.1.1.

We first computed the evolution of the node flow distribution (Eq. 27) with the number of k-NN, as represented in Fig. 13. Note that this measurement is directly performed on the k-NN graph, before

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

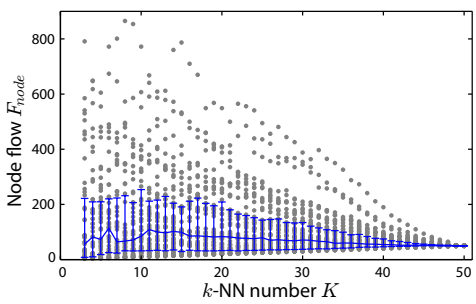


Figure 13: Evolution of the node flow distribution with the number of k-NN. Blue errorbars represent the median and 1st/3rd quartiles over the training set of CRT candidates with SF.

the dimensionality reduction step, and is therefore independent of the retained dimensionality M. No jump in the node flow distribution was observed when the number of k-NN increased, meaning that no short-circuit had been introduced. We therefore chose the dimensionality looking at the value of M minimizing the isomap error ϵ_I for each value of K (white crosses on Fig. 14a), taking its median value as final value for M. Then, we determined the value of K from the evolution of ϵ_I with the number of k-NN, when M is set to its optimal value (Fig. 14b). According to this experiment, we set $M = 4$ and $K = 5$, as hardly any influence on ϵ_I is observed for values of $K < 30$, and high values of K represent a substantial increase in terms of computational time.

A 2D embedding of the computed manifold (output of isomap) is represented in Fig. 15 for illustration purposes, showing the link in the coordinate space between each image and its nearest neighbours. We can qualitatively observe that subjects are arranged in the 2D space according to the pattern present on the map. In particular, subjects with high abnormal motion patterns are located on the border zone of the graph, while the closest subjects to normality are located at the center. However, further investigation is required to determine which characteristic of the SF abnormality pattern (e.g. magnitude, temporal location or spread, spatial location or spread, etc.) is modeled by each of the principal directions of the manifold dataset, as also observed in Fig. 22.

The presence of negative eigenvalues during the isomap process should be carefully checked, as they may not reflect the amount of explained variance (Laub and Müller, 2004). This is not the case for our dataset, as assessed by the compactness curve in Fig. 19, which is strictly increasing.

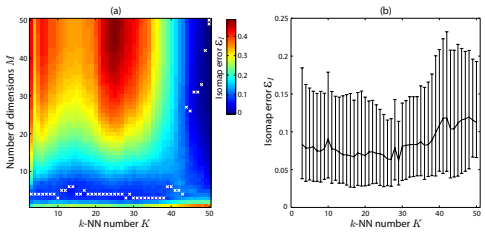


Figure 14: (a) Evolution of the isomap error ϵ_I with the dimensionality M and the number of k-NN, for the training set of CRT candidates with SF. White crosses indicate the minimum value of ϵ_I for each value of k-NN. (b) Evolution of the isomap error ϵ_I with the number of k-NN, for $M = 4$.

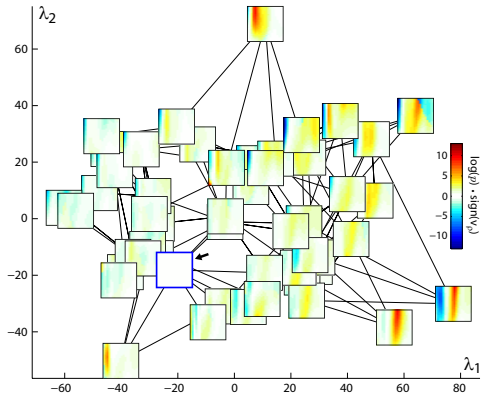


Figure 15: 2D embedding of the manifold of SF p-value maps (output of isomap) according to its two first dimensions. The black arrow indicates the origin image used to constrain the manifold, representing a normal motion pattern.

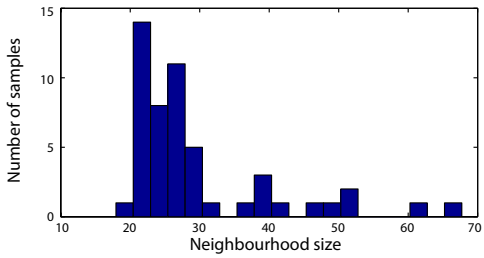


Figure 16: Histogram of the distribution of neighbourhood sizes (right term in Eq. 12) for the training set of CRT candidates with SF. Non-uniformity in the data distribution supports the use of a kernel with varying bandwidth for this dataset.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

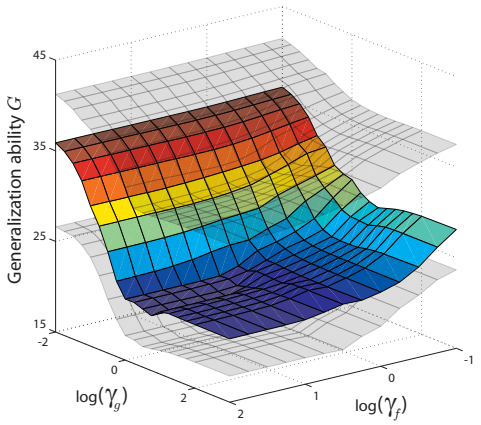


Figure 17: Generalization ability in function of the weighting terms γ_f and γ_g. Median value (color) and 1st/3rd quartiles (gray) over the training set of CRT candidates with SF.

3.2.3. Kernel bandwidth

We used a locally adjustable kernel for the interpolation performed by $f : \mathcal{A} \to \mathcal{C}$ and $g : \mathcal{C} \to \mathcal{A}$, as described in Sec. 2.4.4. The relevance of using such a kernel for the CRT dataset is illustrated in Fig. 16, through the histogram of the distribution of neighbourhood sizes, similarly to the experiment performed in Fig. 12 for the synthetic dataset. The accuracy of such a kernel, in comparison with kernels of fixed bandwidth, was already demonstrated on synthetic data (Sec. 3.1.2). This experiment cannot be repeated for the CRT dataset due to the lack of ground truth parametrization of the manifold.

3.2.4. Weighting the closeness to the data

The generalization ability metric (Sec. 2.7.2) was used to determine the optimal values of the weighting terms γ_f and γ_g in Eq. 8 and 14, respectively. Both weights were determined jointly, as illustrated in Fig. 17. Optimal values were those that minimized the median generalization ability. These values were found to be $\log(\gamma_f) = 1$ and $\log(\gamma_g) = 0.5$ for our dataset.

Fig. 18 represents the reconstruction of five patients from the training set (using leave-one-out), for the optimal values of γ_f and γ_g.

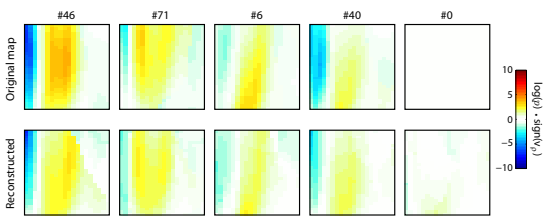


Figure 18: Reconstruction of five patients from the training set of CRT candidates with SF (using leave-one-out), for the optimal values of γ_f and γ_g.

3.2.5. Performance of the dimensionality reduction

We used the objective metrics described in Sec. 2.7 to evaluate the performance of the non-linear dimensionality reduction technique (manifold learning, ML) we used, in comparison with a linear technique (PCA). This comparison is presented in Fig. 19.

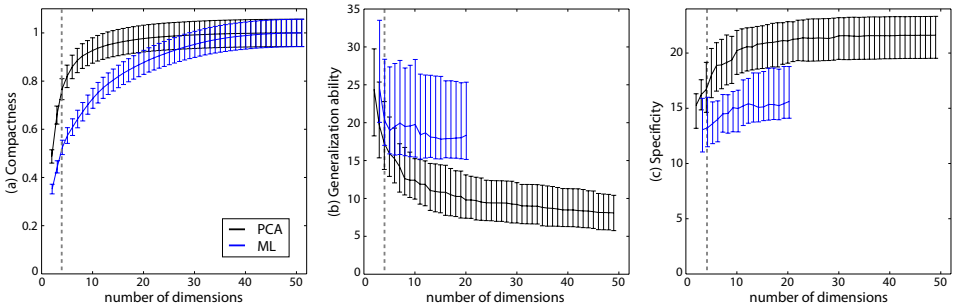
Note that in PCA and ML methods, the eigenvalues used for the estimation of compactness correspond to different objects (comparison of images using the Euclidean distance [linear case] or the geodesic distances [Isomap]). This limits the value of comparing the PCA and ML methods in terms of compactness, despite the fact that compactness is a normalized measure (Eq. 21).

The results in Fig. 19 indicate that the PCA approach is more compact than the ML one, and has a lower generalization ability and higher specificity. We provide in the following some elements of interpretation of these results. Main concerns are (1) the relevance of using non-linear techniques to estimate the manifold structure, and (2) preventing from over-fitting.

Relevance of using a non-linear model. First, consider the dataset represented in Fig. 20. When computing the generalization ability, one sample is removed (black dot), and the ability of the model to reconstruct it measured. In the ML case (left), this reconstruction will be affected by the changes in the local structure of the estimated manifold (dashed line) induced by the leave-one-out process. In contrast, few changes are expected on the PCA coordinates if such a point is left out. The generalization ability will therefore have lower values in the PCA case.

Now, consider the dataset represented in Fig. 21. When computing the specificity, synthetic data is generated from the manifold coordinates, and their closeness to the training set is measured. Points generated from the model coordinates (black dots #1, #2 and #3) will still be close to the training set (red dots) in the non-linear case (left), while

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953



	ML	PCA	p-value
Compactness	0.52 ± 0.03	0.77 ± 0.04	$0.000^{\dagger}$
Generalization ability	$20.27(17.01 - 28.37)$	$17.26(13.83 - 22.78)$	$0.011^{\ddagger}$
Specificity	$13.23(11.52 - 16.00)$	$16.71(14.63 - 19.16)$	$0.000^{\ddagger}$

† unpaired Student's t-test; ‡ Mann-Whitney U-test.

Figure 19: Comparison between non-linear (ML) and linear (PCA) dimensionality reduction techniques using the objective metrics described in Sec. 2.7. Mean $\pm$ standard deviation (compactness) or median and 1st/3rd quartiles (generalization ability and specificity) over the training set of CRT candidates with SF. Generalization ability and specificity experiments were performed up to $M = 20$ due to the poor relevance of using ML for $M > 20$ (Fig. 14). Values for $M = 1$ were not represented. Vertical dashed line indicates the retained dimensionality $M = 4$. Top: Values for a range of dimensions. Optimal values for γ_f and γ_g were determined for each tested dimensionality using the framework of Sec. 3.2.4 (ML case). Bottom: Values for the retained dimensionality $M = 4$. Unpaired Student's t-test ($\dagger$) and Mann-Whitney U-test ($\ddagger$) were used for inter-groups comparison (last column), depending if normal distribution of the values can be assumed or not.

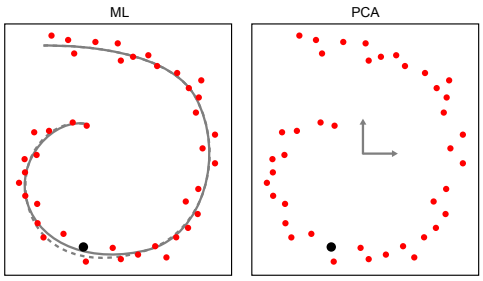


Figure 20: Influence of the removal of one sample (black dot) on the ability of the model to reconstruct the training set (red dots).

nothing prevents them to be far from it in the linear case (right). The specificity will therefore have higher values in the PCA case.

These illustrative examples may help interpreting the values obtained in Fig. 19, but nothing guarantees that the manifold estimated for the SF patients has a structure similar as the one represented in Fig. 20 and 21.

The experiment illustrated in Fig. 22 goes in the direction of the above interpretation. It represents synthetic images generated from the model coordinates (progressive deviations from the map used as origin for normality $\mathbf{I}_0$, along the two first principal directions of the manifold dataset $\mathcal{I}$), obtained using either PCA or ML. As indicated by the black arrows, PCA does not guarantee that the computed maps still contain the characteristic inward and outward events of SF, while this pattern is preserved by the use of ML. This supports the relevance of using non-linear techniques to estimate the manifold structure.

Prevention of over-fitting. An over-fitting model has bad generalization ability (high values for $G(\mathbf{I}_i, M)$), and good specificity (low values of $S(\mathbf{x}, M)$). The results of Fig. 19 may raise the pos-

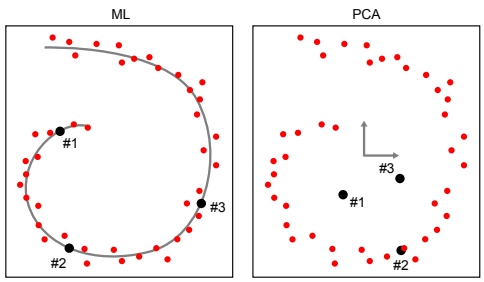


Figure 21: Comparison of three points generated from the model coordinates (black dots #1, #2 and #3) to the training set (red dots).

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

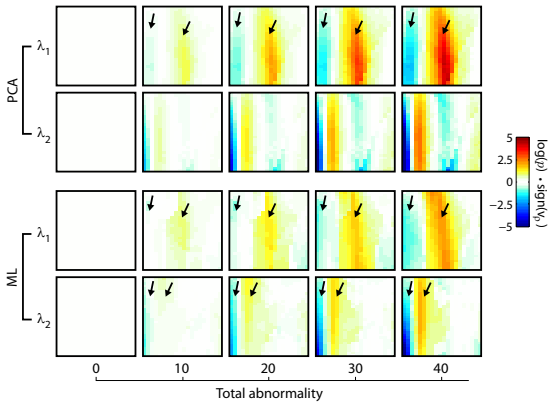
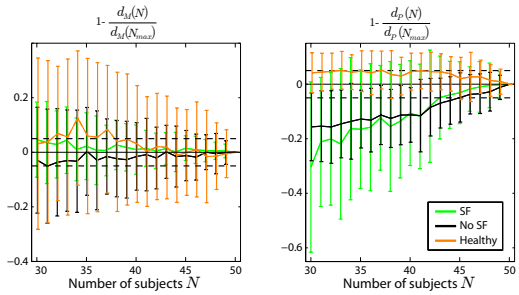


Figure 22: Progressive deviations from the map used as origin for normality $\mathbf{I}_0$, along the two first principal directions of the manifold dataset $\mathcal{I}$, obtained using either PCA or manifold learning (ML). Arrows indicate the characteristic inward and outward motion events of SF. Total abnormality is defined as $S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0)$ (Sec. 2.6).



	SF	No SF	Healthy
d_M	39.2 ± 7.5	41.7 ± 5.8	44.4 ± 5.1
d_P	43.7 ± 7.0	39.0 ± 6.6	40.7 ± 8.4

Figure 23: Top: Normalized evolution of the distances d_M and d_P for subject of each tested population, versus the size of the training population (number of SF patients N). Error bars represent the standard deviation over 100 random combinations of $N_s < N$ subjects. Bottom: values above which this evolution stabilizes to its final value $\pm 5\%$ (dashed line). Average $\pm$ standard deviation values over each tested population.

sibility that our model is over-fitting (higher generalization and lower specificity for the ML method, compared to PCA). However, the experiment illustrated in Fig. 17 aimed at determining the optimal weights γ_f and γ_g to fit to the data, optimality being associated to the lowest generalization ability values. This limits over-fitting, which would correspond to high values of this metric.

Size of the training population. To justify that the subject comparison is not biased due to the size of the training population ($N = 50 + 1$ synthetic image), we computed the evolution of the distances d_M and d_P for a training population made of $N_s < N$ patients. This experiment is summarized in Fig. 23. Little variation was observed for the learning parameters (M, K, γ_f and γ_g) with training population sizes $30 \geq N_s \geq 50$, and we kept for this experiment the parameters estimated for $N_s = N$. Curves were normalized against the value obtained for the largest training population, so that the evolution is represented in the same magnitude scale (%). The plot on the top represents this evolution for one subject of each tested population. For each value of $N_s < N$, the experiment was repeated for 100 random combinations of N_s subjects (vertical error bars). For each tested population and each distance (d_M and d_P), the number of subjects above which this evolution stabilizes to its final value $\pm 5\%$ is summarized in the table of Fig. 23 (average $\pm$ standard deviation values over each tested population). Based on this convergence, we can reasonably trust a manifold estimation with the 50 patients of the training population.

3.3. Patient analysis - CRT data

Fig. 24a represents the distance between all the subjects involved in this study and the manifold. We separated the analysis between d_M and d_P for interpretation purposes. The patients from the training set have low d_P (distance to the manifold), which corresponds to the reconstruction error inherent to the estimation of f and g using an inexact matching formulation (Eq. 8 and 14), and largely span the space associated to d_M (distance to normality along the manifold). As the training population size is finite, the density in the space of coordinates around patients with the most abnormal patterns is lower, and these patients have higher reconstruction error, namely higher d_P. Few patients from the training set are close to the origin according to d_M, in comparison with the healthy volunteers. This may come from the accuracy of the patient selection process using M-mode images (Parsai et al., 2009), small SF being harder to identify, and from the accuracy of the abnormality maps to detect low abnormalities (Duchateau et al., 2011b). Among the testing subjects, patients having SF are closer to the manifold than patients without SF, according to d_P, and almost within the range of the reconstruction error for the training set. Higher val-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

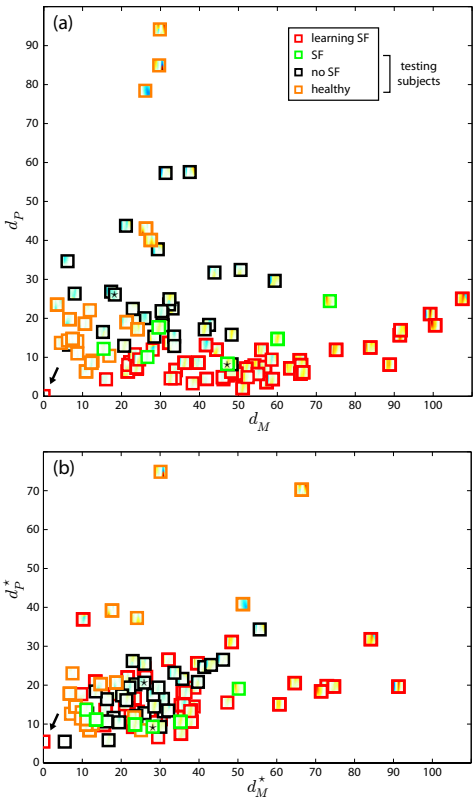


Figure 24: (a) Subject ordering according to d_M and d_P (ML method, Sec. 2.6), used as horizontal and vertical axis, respectively. (b) Subject ordering according to $d_M^\star$ and $d_P^\star$ (PCA, Eq. 30), used as horizontal and vertical axis, respectively. The black arrow indicates the image used to constrain the manifold, representing a normal motion pattern, and used as origin. Asterisk-marked miniatures refer to two subjects with the same amount of total abnormality, but belonging to different populations (cf. App. C).

ues of d_M are observed in the subjects having higher SF abnormalities on the maps. Large values of d_P can be observed for some volunteers. These subjects actually have high velocities during the cardiac cycle, reflected by high values of abnormality on their associated 2D maps. Their pattern of motion abnormality cannot be accurately reconstructed from the training population, which does not include any healthy subject. Thus, their corresponding image on the manifold $\hat{\mathbf{I}}$ tends to be close to $\mathbf{I}_0$, resulting in a low d_M and a high d_P.

Fig. 24b uses a similar display to represent the ordering of subjects obtained using PCA, separating the analysis between $d_M^\star$ and $d_P^\star$, defined as:

$$\begin{cases} d_P^\star(\mathbf{I}) = S_{\mathcal{A}}(\hat{\mathbf{I}}^\star, \mathbf{I}), \\ d_M^\star(\mathbf{I}) = S_{\mathcal{C}}(\mathbf{x}^\star, \mathbf{x}_0^\star), \end{cases} \tag{30}$$

where $\hat{\mathbf{I}}^\star$ is the reconstruction of $\mathbf{I}$ using the first M principal directions obtained from PCA, and $\mathbf{x}^\star$ are the M-dimensional coordinates of $\mathbf{I}$ in the PCA space.

This figure highlights the limitations of PCA to perform patient comparison to a population with a specific abnormal pattern. Little discrimination is observed between the healthy volunteers, the testing patients having SF, and patients without SF. Both patients with and without SF appear equally distant from the SF training set according to $d_P^\star$, and volunteers do not have necessarily low values of $d_M^\star$. The PCA-based approach therefore does not guarantee that our main objectives for CRT studies (characterizing patients according to their distance to patterns of dyssynchrony for which the response rate is roughly known, and grading of the disease severity) are fulfilled, in comparison with the manifold learning approach.

3.4. Total distance to normality

Each 2D map processed in this study locally contains a measure of abnormality, and the total abnormality contained in each map is therefore defined as $S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0)$ (Sec. 2.6).

It is therefore interesting to look for the link between $S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0)$ and the distances d_M and d_P defined in Sec. 2.6. We can show that the dimensionality reduction inherent to the manifold estimation preserves the concept of abnormality embedded in the input maps, but also takes into account the geometry of the training set. This interpretation and the comparison between the ML and PCA cases are detailed in App. C.

4. Discussion

We have described a complete pipeline to compare individuals and a population with a specific pattern of abnormal motion in terms of myocardial motion. The extension of manifold learning techniques to embed the definition of a physiologically meaningful origin allowed representing the learnt population as progressive deviations from normality along a manifold structure. The originality of our work resides in using 2D maps of motion abnormalities as input, obtained from a statistical atlas of

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

myocardial motion built from a set of healthy volunteers, which facilitates the definition of an origin for normality. Our experiments demonstrated the relevance of manifold learning techniques to learn a population with a specific pathological pattern, and to compare individuals to this pattern. We first selected the optimal values for the parameters involved in our method, using a synthetic dataset and the training set of CRT candidates for which the manifold is learnt. Then, we demonstrated the performance of our method to characterize both training and testing datasets.

As described in Sec. 3.2, we used the Euclidean distance for the metric $S_{\mathcal{A}}$. This metric evaluates pixel-wise the difference between two maps of abnormality, being considered as 2D images. The relationship between this metric and the statistical information locally contained in each map will be addressed in further work. Choosing the Euclidean distance may introduce some bias towards global and local shifts of the observed patterns. However, the use of TDFFD (De Craene et al., 2012b) guarantees the velocities computed for each subject to be differentiable, and therefore introduces smoothness on the abnormality maps, which limits the above-mentioned bias. We preferred to choose a simple metric for the sake of computational speed. Metrics based on image registration (as used in other manifold learning applications working directly on real images (Gerber et al., 2010)) or image correlation-based metrics cannot be applied in our case as they require the processed images to share the same topology. Indeed, a large variety of patterns are observed on the maps of abnormality we used either for the manifold estimation or for the comparison to the manifold, as visible in Fig. 4. In particular, these alternative metrics would certainly fail in case of the lack of abnormal pattern, typically for the maps of healthy volunteers.

We preferred a strategy in which the manifold is estimated for one specific population only (SF patients). This approach differs from classic methods for population comparison (Ashburner and Klöppel, 2011), which generally consider all subjects as part of a single dataset, and look for the space of reduced dimensionality that allows the best discrimination between different pre-identified groups. Nonetheless, the flexibility of these techniques is limited when a new subject or a new population is added to the existing dataset, as dimensionality reduction is applied to the whole set of studied subjects. In addition, the dimensionality reduction could be biased towards certain populations if they show higher variability. An alternative for moving beyond these limitations consists in separating the analysis for each coherent group of subjects, which is the strategy we have opted for.

In this work, we assumed that the training set contained members close to the origin, namely close to normality. This allows modeling the pathological pattern of SF as a progressive degeneration from normality. In case this methodology is applied to other populations where this assumption does not hold, the ML process can be made simpler (without the additional constraint of exact matching for one point), as formulated in Sec. 2.4.2.

The approach we proposed represents a potential step forward to improve patient selection in the context of CRT. As highlighted in Fornwalt (2011), current approaches lack of reproducible tools to perform patient comparison pre- and post- therapy. Our method is ready to be used in both cases: (1) using baseline data, it would allow grading a pattern (distance to normality) and estimating the ability of a patient to respond (distance to patterns for which the response rate is known (Parsai et al., 2009)); (2) using baseline and follow-up data, it would improve the understanding of the link between the evolution of abnormal patterns and CRT outcome. Both aspects will be studied in further work as part of a thorough clinical study.

Limitations. The pipeline presented in this study was applied to the characterization of individuals against patients with SF, as this pattern is clearly defined on the maps we used as input, which were validated in Duchateau et al. (2011b). Besides, this pattern has been shown to highly condition CRT outcome (Parsai et al., 2009; Duckett et al., 2012).

The maps were computed from 2D echocardiographic sequences, as it is currently the most clinically widespread modality with sufficient temporal resolution to accurately quantify SF.

The study focused on motion abnormalities only. Its extension to strain abnormalities may refine the analysis of the patterns of cardiac dyssynchrony by allowing the characterization of locally infarcted segments, which may affect CRT outcome (Khan et al., 2010).

Nonetheless, the methodology described in the present work is generic, and therefore not specific to SF, or abnormality maps obtained from 2D echocardiography. Other potential strategies for building the maps of abnormality could include other imag-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

ing modalities (3D echocardiography, magnetic resonance and tagged magnetic resonance as recently proposed (De Craene et al., 2012a)) and other pathological cardiac mechanisms such as the classes of mechanical dyssynchrony identified in Parsai et al. (2009).

The quality of the maps of abnormality is primordial for the accuracy of the proposed method. Both the echocardiographic acquisition and the atlas construction steps (velocity extraction and spatiotemporal synchronization) may influence the patterns observed on the maps and the subject comparison, within the manifold learning process (construction of the k-NN graph and intra-manifold distance) and when mapping patients to the manifold.

The current strategy, namely first compute maps of abnormality and then perform ML, may not be optimal as both steps were considered independent. However, using the abnormality maps as input has the advantage of providing a representation of what would be a perfectly "normal" motion pattern, namely the synthetic image with 0 value at each pixel that we added to our training set ("normality" being conditioned by the relevance of the healthy population used to compute the maps, see Duchateau et al. (2011b) for further details on this point). This facilitates the procedure of modeling a pathology as a deviation from normality, reducing a problem involving two populations (healthy and the studied pattern) to a problem with the studied pattern and a relevant origin.

An improvement of our method could combine both steps, learning directly from 2D maps of velocities. This would require to carefully choose a new similarity metric between the images, with consequences on both the representation of the training populations (the population with the pattern to learn during the isomap procedure, and the population of healthy volunteers that define normal motion) and the comparison to these populations. The optimality of the method should however be carefully defined, depending on the objective (population modeling or classification, for instance).

5. Conclusion

We have proposed a method for representing a specific pathological motion pattern as a deviation from normality along a manifold structure, normality being by construction the manifold origin. The method was used to characterize individuals according to their distance to normality, and to the pathological pattern used to estimate the manifold. We first evaluated the optimal set of parameters involved in our pipeline. Then, we illustrated the performance of such an approach in the context of CRT, learning the manifold for a set of patients with SF, a specific pattern of intra-ventricular dyssynchrony, and comparing both healthy volunteers and CRT candidates to this population. Experiments demonstrated the advantage of non-linear embedding of the training set, and the relevance of the proposed method for grading different stages of motion abnormalities and comparing subjects to a specific pathological pattern.

Acknowledgments

This work was supported by the Spanish Industrial and Technological Development Center (cvREMOD CEN-20091044), the Spanish Ministry of Science and Innovation, Plan E and ERDF (STIMATH TIN2009-14536-C02-01), and the European Commission's Seventh Framework Program (euHeart FP7-ICT-224495). The authors gratefully acknowledge the support and advice for the atlas construction and the data collection of Dr. E. Silva, Dr. A. Doltra and Dr. M. Sitges from the Thorax Institute, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Universitat de Barcelona, Spain; and Prof. B.H. Bijnens from the Universitat Pompeu Fabra, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain.

Conflict of interest

None.

Appendix A. Choice of the kernel.

The theory of RKHS provides specific results about the link between the RKHS and the function used to define the kernel. Glaunes (2005) extensively discusses this, focusing on characteristics of the kernel such as regularity and invariance by rigid transforms, and the definition of an observation scale.

Taking as example the matching problem of Sec. 2.4.1, simple parametric kernels are of the form $K_{\mathcal{F}}(\mathbf{I},\mathbf{J}) = h_{\mathcal{F}}(S_{\mathcal{A}}(\mathbf{I},\mathbf{J})) \cdot \mathbf{Id}_{M,M}$, $\forall \mathbf{I},\mathbf{J} \in \mathcal{A}$, namely isotropic and invariant by rigid transforms, where $h_{\mathcal{F}} : \mathbb{R}^+ \to \mathbb{R}$ is a given scalar function, $S_{\mathcal{A}} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$ is the metric used to compare

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

elements of $\mathcal{A}$, and $\mathbf{Id}_{M,M}$ the identity matrix in $\mathbb{M}_{M,M}$.

Examples of such kernels that are semi-definite positive, given $x \in \mathbb{R}^+$, are defined by the following functions $h_{\mathcal{F}}$:

- Gaussian kernels: $h_{\mathcal{F}}(x) = \exp\big(-x^2/\sigma_{\mathcal{F}}^2\big)$,
- Cauchy kernels: $h_{\mathcal{F}}(x) = \big(1 + x^2/\sigma_{\mathcal{F}}^2\big)^{-1}$,
- Sobolev kernels, for which $h_{\mathcal{F}}$ is the inverse Fourier transform of $(1+x^2)^{-s}$, $s > P + 1/2$,
- Bessel kernels (Coifman and Lafon, 2006; Glaunes, 2005).

Gaussian and Cauchy kernels are particularly of interest as they introduce an observation scale $\sigma_{\mathcal{F}}$. However, choosing the optimal function $h_{\mathcal{F}}$ defining the kernel is still an open issue. In our application, we opted for Gaussian kernels due to their wide use in the machine learning community.

Appendix B. Formulation of the exact matching problem.

According to the formulation of Eq. 3 (Sec. 2.4.1), the exact matching problem has the following analytical solution:

$$f(\mathbf{I}) = \sum_{i=0}^{N} K_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i, \tag{31}$$

where the vectors $\mathbf{c}_i \in \mathcal{C}$ are solutions of the following linear system:

$$\sum_{j=0}^{N} K_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j) \cdot \mathbf{c}_j = \mathbf{x}_i, \ \forall i \in [0, N]. \tag{32}$$

Note that $K_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j) \in \mathbb{M}_{M,M}$ with this formulation.

Eq. 32 can be condensed using the product of block matrices:

$$\mathbf{K}_{\mathcal{I}} \cdot \mathbf{C} = \mathbf{X}, \tag{33}$$

where:

- $\mathbf{C}^t = (\mathbf{c}_0^t, \ldots, \mathbf{c}_N^t) \in \mathbb{R}^{(N+1)M}$,
- $\mathbf{K}_{\mathcal{I}} \in \mathbb{M}_{(N+1)M,(N+1)M}$ is the matrix made of the $M \times M$-dimensional blocks $K_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j)$, $(i,j) \in [0,N]^2$,
- $\mathbf{X}^t = (\mathbf{x}_0^t, \ldots, \mathbf{x}_N^t) \in \mathbb{R}^{(N+1)M}$.

If we choose a reproducing kernel of the form (App. A):

$$K_{\mathcal{F}}(\mathbf{I}, \mathbf{J}) = k_{\mathcal{F}}(\mathbf{I}, \mathbf{J}) \cdot \mathbf{Id}_{M,M}, \tag{34}$$

where $\mathbf{Id}_{M,M}$ is the identity matrix in $\mathbb{M}_{M,M}$, and $k_{\mathcal{F}} = h_{\mathcal{F}} \circ S_{\mathcal{A}}$, the solution for Eq. 3 can be rewritten in a simpler form, namely:

$$\begin{cases} f(\mathbf{I}) = \sum_{i=0}^{N} k_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i, \\ \text{with } \mathbf{C} = \mathbf{K}_{\mathcal{I}}^{-1} \cdot \mathbf{X}, \end{cases} \tag{35}$$

where:

- $\mathbf{C} = (\mathbf{c}_0, \ldots, \mathbf{c}_N)^t \in \mathbb{M}_{N+1,M}$,
- $\mathbf{K}_{\mathcal{I}} = \big(k_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j)\big)_{(i,j)\in[0,N]^2} \in \mathbb{M}_{N+1,N+1}$,
- $\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_N)^t \in \mathbb{M}_{N+1,M}$.

Appendix C. Link between the total abnormality in each map and proposed distances.

In this paper, we used the Euclidean distance for both metrics $S_{\mathcal{A}}$ and $S_{\mathcal{C}}$. Thus, the total abnormality contained in each map is:

$$S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0) = \|\mathbf{I} - \mathbf{I}_0\| = \|\mathbf{I}\|. \tag{36}$$

In the PCA case, we can demonstrate that:

$$\begin{aligned} \|\mathbf{I}\|^2 &= \|\mathbf{I} - \hat{\mathbf{I}}^\star\|^2 + \|\hat{\mathbf{I}}^\star - \hat{\mathbf{I}}_0^\star\|^2 \\ &\quad + \|\mathbf{I}_0 - \hat{\mathbf{I}}_0^\star\|^2 - 2 \sum_{m=M+1}^{N+1} x_m^\star \cdot x_{0,m}^\star, \end{aligned} \tag{37}$$

where $\hat{\mathbf{I}}^\star$ and $\hat{\mathbf{I}}_0^\star$ are the PCA-based reconstruction of $\mathbf{I}$ and $\mathbf{I}_0$ using the first M principal directions, and $x_m^\star$ and $x_{0,m}^\star$ correspond to the mth component of $\mathbf{x}^\star$ and $\mathbf{x}_0^\star$, respectively.

Proof: We denote $\mathbf{e}_m^\star$ the eigenvector of the PCA basis corresponding to the mth component. We can write from the PCA definition:

$$\mathbf{I} = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{I}_i + \sum_{m=1}^{N+1} x_m^\star \mathbf{e}_m^\star. \tag{38}$$

From the dimensionality reduction process, we also have that:

$$\hat{\mathbf{I}}^\star = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{I}_i + \sum_{m=1}^{M} x_m^\star \mathbf{e}_m^\star. \tag{39}$$

Thus, for the left side of Eq. 37:

$$\begin{aligned} \|\mathbf{I}\|^2 &= \|\mathbf{I} - \mathbf{I}_0\|^2, \\ &= \sum_{m=1}^{N+1} (x_m^\star - x_{0,m}^\star)^2, \\ &= \sum_{m=1}^{N+1} \big((x_m^\star)^2 + (x_{0,m}^\star)^2 - 2 \cdot x_m^\star \cdot x_{0,m}^\star\big). \end{aligned} \tag{40}$$

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

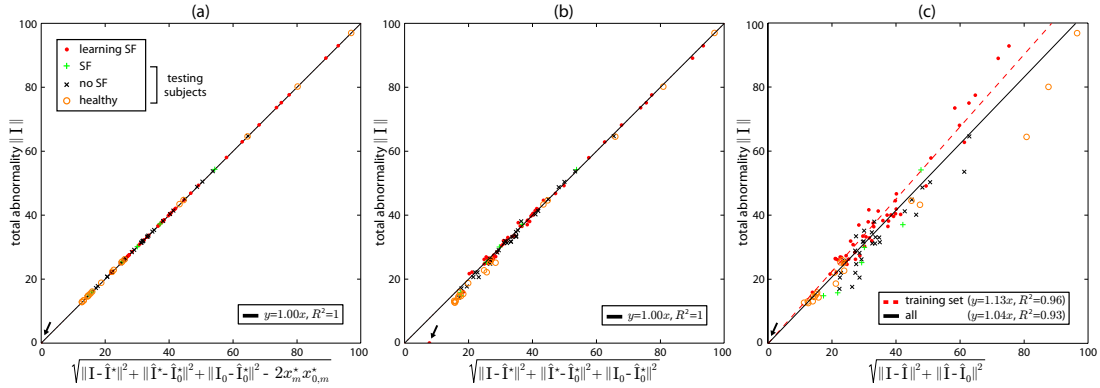


Figure 25: Subject ordering according to the amount of total abnormality contained in each map and (a) the PCA-based distance to normality with residual, or (b) without residual, or (c) the ML-based total distance to normality. The black arrow indicates the image used to constrain the manifold, representing a normal motion pattern, and used as origin.

And for each term of the right side of Eq. 37:

$$\begin{aligned}\|\mathbf{I}-\hat{\mathbf{I}}^\star\|^2 &= \|\sum_{m=1}^{N+1} x_m^\star \mathbf{e}_m^\star - \sum_{m=1}^{M} x_m^\star \mathbf{e}_m^\star\|^2, \\ &= \sum_{m=M+1}^{N+1} (x_m^\star)^2.\end{aligned} \tag{41}$$

Similarly, we have:

$$\|\mathbf{I}_0-\hat{\mathbf{I}}_0^\star\|^2 = \sum_{m=M+1}^{N+1} (x_{0,m}^\star)^2. \tag{42}$$

And finally:

$$\begin{aligned}\|\hat{\mathbf{I}}^\star-\hat{\mathbf{I}}_0^\star\|^2 &= \|\sum_{m=1}^{M} x_m^\star \mathbf{e}_m^\star - \sum_{m=1}^{M} x_{0,m}^\star \mathbf{e}_m^\star\|^2, \\ &= \sum_{m=1}^{M} \left((x_m^\star)^2 + (x_{0,m}^\star)^2 - 2\cdot x_m^\star \cdot x_{0,m}^\star\right).\end{aligned} \tag{43}$$

The result of Eq. 37 is therefore straightforward.

In the PCA case, and according to the computations above, we also have:

$$\begin{aligned}\|\hat{\mathbf{I}}^\star-\hat{\mathbf{I}}_0^\star\|^2 &= \sum_{m=1}^{M} (x_m^\star - x_{0,m}^\star)^2, \\ &= S_{\mathcal{C}}(\mathbf{x}^\star, \mathbf{x}_0^\star)^2,\end{aligned} \tag{44}$$

and Eq. 37 can be rewritten as:

$$\begin{aligned}\|\mathbf{I}\|^2 &= d_P^\star(\mathbf{I})^2 + d_M^\star(\mathbf{I})^2 \\ &+ d_P^\star(\mathbf{I}_0)^2 - 2\sum_{m=M+1}^{N+1} x_m^\star \cdot x_{0,m}^\star.\end{aligned} \tag{45}$$

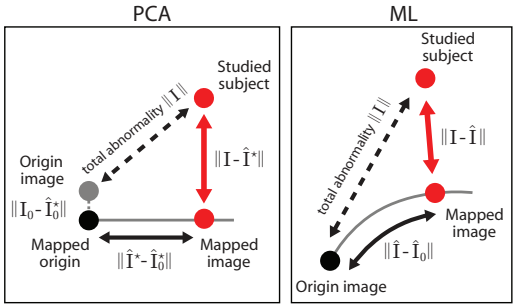
This is illustrated in Fig. 25a, confirming that both expressions are equal. Note that d_M and d_P calculate the Euclidean norm of M- and P- dimensional vectors, respectively. The validity of using such distances in the same equation is therefore ensured by the residual term $2\cdot\sum_{m=M+1}^{N+1} x_m^\star \cdot x_{0,m}^\star$.

Fig. 25b compares the amount of total abnormality to the right term of Eq. 37 without this residual term, namely to $\sqrt{\|\mathbf{I}-\hat{\mathbf{I}}^\star\|^2 + \|\hat{\mathbf{I}}^\star-\hat{\mathbf{I}}_0^\star\|^2 + \|\mathbf{I}_0-\hat{\mathbf{I}}_0^\star\|^2}$, showing this mostly affects the images close to normality.

We performed the same comparison in the ML case, for which the manifold is constrained to pass by $\mathbf{I}_0$, and therefore $\hat{\mathbf{I}}_0^\star = \mathbf{I}_0$, namely $d_P(\mathbf{I}_0) = 0$. As visible in Fig. 25c, the property discussed above is not satisfied in this case, linear regression over the plotted data leading to slope coefficients of 1.13 (manifold data only, dashed red line) and 1.04 (whole data, black line), associated to R^2 coefficients of 0.96 and 0.93, respectively. The dimensionality reduction inherent to the manifold estimation still preserves the concept of abnormality embedded in the input maps, but also takes into account the geometry of the training set, as summarized in Fig. 26.

These observations are also supported by the fact that some images may contain the same amount of total abnormality, but be clearly discriminated by the proposed distances d_M and d_P. Fig. 27 illustrates this point for two subjects with different patterns on the maps (one with SF, and the other without SF). Both maps contain the same amount of total abnormality despite the fact that they be-

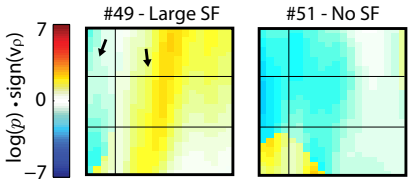
Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953



$$(\text{total abnormality})^2 \approx \|\mathbf{I}-\hat{\mathbf{I}}^\star\|^2 + \|\hat{\mathbf{I}}^\star-\hat{\mathbf{I}}_0^\star\|^2 + \|\mathbf{I}_0-\hat{\mathbf{I}}_0^\star\|^2$$

$$(\text{total abnormality})^2 \neq \|\mathbf{I}-\hat{\mathbf{I}}\|^2 + \|\hat{\mathbf{I}}-\hat{\mathbf{I}}_0\|^2$$

Figure 26: Relationship between the total abnormality contained in each map and the proposed distances for comparing subjects to the training population.



	#49	#51
Total abnormality	30.06	31.79
d_M	47.16	18.27
d_P	8.32	26.11

Figure 27: Example of two patients having the same amount of total abnormality $S_{\mathcal{A}}(\mathbf{I}, \mathbf{I}_0)$, but different patterns (presence and absence of SF, respectively). These subjects are clearly discriminated by the use of the proposed distances d_M and d_P (cf. table values and asterisks-marked miniatures on Fig. 24).

long to different groups, but are correctly discriminated by the distances we propose (Fig. 27 and asterisk-marked miniatures in Fig. 24).

References

Aronszajn, N., 1950. Theory of reproducing kernels. Transactions of the American Mathematical Society 68, 337–404.

Ashburner, J., Klöppel, S., 2011. Multivariate models of inter-subject anatomical variability. Neuroimage 56, 422–439.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15, 1373–1396.

Bengio, Y., Paiement, J.F., Vincent, P., Delalleau, O., Le Roux, N., Ouimet, M., 2004. Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering. Advances in Neural Information Processing Systems 16, 177–184.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16, 1190–1208.

Camastra, F., 2003. Data dimensionality estimation methods: A survey. Pattern Recognition 36, 2945–2954.

Chandrashekara, R., Mohiaddin, R., Rueckert, D., 2005. Comparison of cardiac motion fields from tagged and untagged MR images using nonrigid registration, in: Proc. International Conference on Functional Imaging and Modeling of the Heart (FIMH), LNCS vol. 3504, pp. 425–433.

Choi, H., Choi, S., 2007. Robust kernel isomap. Pattern Recognition 40, 853–862.

Coifman, R.R., Lafon, S., 2006. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. Applied and Computational Harmonic Analysis 21, 31–52.

Davies, R.H., Twining, C.J., Cootes, T.F., Taylor, C.J., 2010. Building 3-d statistical shape models by direct optimization. IEEE Transactions on Medical Imaging 29, 961–981.

Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2010. Population shape regression from random design data. International Journal of Computer Vision 90, 255–266.

De Boeck, B.W.L., Teske, A.J., Meine, M., Leenders, G.E., Cramer, M.J., Prinzen, F.W., Doevendans, P.A., 2009. Septal rebound stretch reflects the functional substrate to cardiac resynchronization therapy and predicts volumetric and neurohormonal response. European Journal of Heart Failure 11, 863–871.

De Craene, M., Duchateau, N., Tobon-Gomez, C., Ghafaryasl, B., Piella, G., Rhode, K.S., Frangi, A.F., 2012a. SPM to the heart: Mapping of 4D continuous velocities for motion abnormality quantification, in: Proc. IEEE International Symposium on Biomedical Imaging, pp. 454–457.

De Craene, M., Piella, G., Camara, O., Duchateau, N., Silva, E., Doltra, A., Dhooge, J., Brugada, J., Sitges, M., Frangi, A.F., 2012b. Spatiotemporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography. Medical Image Analysis 16, 427–450.

De Craene, M., Tobon-Gomez, C., Butakoff, C., Duchateau, N., Piella, G., Rhode, K.S., Frangi, A.F., 2011. Temporal diffeomorphic free form deformation (TDFFD) applied to motion and deformation quantification of tagged MRI sequences, in: Proc. Statistical Atlases and Computational Models of the Heart (STACOM), MICCAI'11 Workshop, pp. 247–256.

Delgado, V., Bax, J.J., 2011. Assessment of systolic dyssynchrony for cardiac resynchronization therapy is clinically useful. Circulation 123, 640–655.

Dickstein, K., Vardas, P.E., Auricchio, A., Daubert, J.C., Linde, C., McMurray, J., Ponikowski, P., Priori, S.G., Sutton, R., van Veldhuisen, D.J., 2010. Focused update of ESC guidelines on device therapy in heart failure: an update of the 2008 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure and the 2007 ESC guidelines for cardiac and resynchronization therapy. European Heart Journal 31, 2677–2687.

Duchateau, N., De Craene, M., Piella, G., Frangi, A.F., 2011a. Characterizing pathological deviations from normality using constrained manifold-learning, in: Proc. In-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

ternational Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS vol. 6893, pp. 256–263.

Duchateau, N., De Craene, M., Piella, G., Silva, E., Doltra, A., Sitges, M., Bijnens, B.H., Frangi, A.F., 2011b. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. Medical Image Analysis 15, 316–328.

Duckett, S.G., Camara, O., Ginks, M.R., Bostock, J., Chinchapatnam, P., Sermesant, M., Pashaei, A., Lambiase, P.D., Gill, J.S., Carr-White, G.S., Frangi, A.F., Razavi, R., Bijnens, B.H., Rinaldi, C.A., 2012. Relationship between endocardial activation sequences defined by high-density mapping to early septal contraction (septal flash) in patients with left bundle branch block undergoing cardiac resynchronization therapy. Europace 14, 99–106.

Durrleman, S., 2010. Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Phd thesis. Université de Nice-Sophia Antipolis.

Etyngier, P., Keriven, R., Segonne, F., 2007. Projection onto a shape manifold for image segmentation with prior, in: Proc. IEEE International Conference on Image Processing (ICIP), pp. IV361–364.

Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging 23, 995–1005.

Fornwalt, B.K., 2011. The dyssynchrony in predicting response to cardiac resynchronization therapy: A call for change. Journal of the American Society of Echocardiography 24, 180–184.

Fornwalt, B.K., Delfino, J.G., Sprague, W.W., Oshinski, J.N., 2009. It's time for a paradigm shift in the quantitative evaluation of left ventricular dyssynchrony. Journal of the American Society of Echocardiography 22, 672–676.

Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R., 2010. Manifold modeling for brain population analysis. Medical Image Analysis 14, 643–653.

Gerber, S., Tasdizen, T., Whitaker, R., 2009. Dimensionality reduction and principal surfaces via kernel map manifolds, in: Proc. IEEE International Conference on Computer Vision, pp. 529–536.

Gjesdal, O., Remme, E.W., Opdahl, A., Skulstad, H., Russell, K., Kongsgaard, E., Edvardsen, T., Smiseth, O.A., 2011. Mechanisms of abnormal systolic motion of the interventricular septum during left bundle-branch block. Circulation Cardiovascular Imaging 4, 264–273.

Glaunes, J.A., 2005. Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. Phd thesis. Université Paris 13.

Hastie, T., Stuetzle, W., 1989. Principal curves. Journal of the American Statistical Association 84, 502–516.

Heimdal, A., Stylen, A., Torp, H., Skjaerpe, T., 1998. Real-time strain rate imaging of the left ventricle by ultrasound. Journal of the American Society of Echocardiography 11, 1013–1019.

Khan, F.Z., Virdee, M.S., Read, P.A., Pugh, P.J., O'Halloran, D., Fahey, M., Elsik, M., Begley, D., Fynn, S.P., Dutka, D.P., 2010. Effect of low-amplitude two-dimensional radial strain at left ventricular pacing sites on response to cardiac resynchronization therapy. Journal of the American Society of Echocardiography 23, 1168–1176.

Kim, K.H., Choi, S., 2007. Neighbor search with global geometry: a minimax message passing algorithm, in: Proc. International Conference on Machine Learning (ICML), pp. 410–408.

Kimeldorf, G.S., Wahba, G., 1970. A correspondance between bayesian estimation on stochastic processes and smoothing by splines. Annals of Mathematical Statistics 41, 495–502.

Kwok, J.T.Y., Tsang, I.W.H., 2004. The pre-image problem in kernel methods. IEEE Transactions on Neural Networks 15, 1517–1525.

Laub, J., Müller, K.R., 2004. Feature discovery in non-metric pairwise data. Journal of Machine Learning Research 5, 808–818.

Marwick, T.H., Yu, C.M., Sun, J.P., 2007. Myocardial Imaging: Tissue Doppler and Speckle Tracking. Wiley-Blackwell.

Meinicke, P., Klanke, S., Memisevic, R., Ritter, H., 2005. Principal curves from unsupervised kernel regression. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1379–1391.

Mika, S., Schlkopf, B., Smola, A., Mller, K.R., Scholz, M., Rtsch, G., 1999. Kernel PCA and de-noising in feature spaces. Advances in Neural Information Processing Systems 11, 536–542.

Parsai, C., Bijnens, B.H., Sutherland, G.R., Baltabaeva, A., Claus, P., Marciniak, M., Paul, V., Scheffer, M., Donal, E., Derumeaux, G., Anderson, L., 2009. Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms. European Heart Journal 30, 940–949.

Rao, A., Chandrashekara, R., Sanchez-Ortiz, G.I., Mohiaddin, R., Aljabar, P., Hajnal, J.V., Puri, B.K., Rueckert, D., 2004. Spatial transformation of motion and deformation fields using nonrigid registration. IEEE Transactions on Medical Imaging 23, 1065–1076.

Rougon, N.F., Petitjean, C., Prêteux, F.J., 2004. Building and using a statistical 3D motion atlas for analyzing myocardial contraction in MRI, in: Proc. SPIE Conference on Medical Imaging, pp. 253–264.

Roweis, S., Saul, L., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A., 2006. Diffeomorphic registration using B-splines, in: Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS vol. 4191, pp. 702–709.

Saitoh, S., 1988. Theory of Reproducing Kernels and its Applications. Pitman Research Notes in Mathematical Series, 189. Wiley.

Schölkopf, B., Herbrich, R., Smola, A.J., 2001. A generalized representer theorem. Computational Learning Theory 2111, 416–426.

Schölkopf, B., Smola, A., Müller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299–1319.

Scott, D.W., 1992. Multivariate density estimation: theory, practice, and visualization (Wiley Series in Probability and Statistics). Wiley.

Silverman, B.W., 1986. Density estimation for statistics and data analysis (Chapman & Hall / CRC Monographs on Statistics & Applied Probability). Chapman and Hall / CRC.

St John Sutton, M.G., Plappert, T., Abraham, W.T., Smith, A.L., DeLurgio, D.B., Leon, A.R., Loh, E., Kocovic, D.Z., Fisher, W.G., Ellestad, M., Messenger, J., Kruger, K.,

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841512000953

Hilpisch, K.E., Hill, M.R., Multicenter InSync Randomized Clinical Evaluation (MIRACLE) Study Group, 2003. Effect of cardiac resynchronization therapy on left ventricular size and function in chronic heart failure. Circulation 107, 1985–1990.

Sung, R.K., Foster, E., 2011. Assessment of systolic dyssynchrony for cardiac resynchronization therapy is not clinically useful. Circulation 123, 656–662.

Szulik, M., Tillekaerts, M., Vangeel, V., Ganame, J., Willems, R., Lenarczyk, R., Rademakers, F., Kalarus, Z., Kukulski, T., Voigt, J.U., 2010. Assessment of apical rocking: a new, integrative approach for selection of candidates for cardiac resynchronization therapy. European Journal of Echocardiography 11, 863–869.

Tenenbaum, J.B., De Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.

Tenenbaum, J.B., Langford, J.C., 2002. Response to letter "The Isomap Algorithm and Topological Stability" from Balasubramanian, M. and Schwartz, E.L. Science 295, 7.

Toga, A.W., Thompson, P.M., Mori, S., Amunts, K., Zilles, K., 2006. Towards multimodal atlases of the human brain. Nature Reviews Neuroscience 7, 952–966.

Young, A.A., Frangi, A.F., 2009. Computational cardiac atlases: from patient to population and back. Experimental Physiology 94, 578–596.

Zhang, Z., Wang, J., Zha, H., 2012. Adaptive manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 253–265.

Zheng, W.S., Lai, J.H., Yuen, P.C., 2010. Penalized preimage learning in kernel principal component analysis. IEEE Transactions on Neural Networks 21, 551–570.

Vitæ

Nicolas Duchateau received his Engineering degree in Optics from the Institut d'Optique, Palaiseau, France, in 2007, and his MSc degree in Mathematics, Vision and Machine-learning from the Ecole Normale Supérieure de Cachan, France, in 2008. He recently received his PhD degree from the Universitat Pompeu Fabra (UPF), Barcelona, Spain, working within the Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB) under the supervision of Mathieu De Craene and Alejandro Frangi. His main research interests target the use of image registration and statistical atlases for the quantification of heart motion and deformation. He now works with the Fundació Clínic, Barcelona, Spain, within the Arrhythmia, Resynchronization and Cardiac Imaging unit, under the supervision of Dr. Marta Sitges.

Mathieu De Craene received his PhD degree from the Université catholique de Louvain, Belgium, in 2005. His thesis focused on developing automatic registration methods for medical images. He has been a visiting student at the Computational Radiology Laboratory, Boston, IL. He joined the CISTIB at the UPF in 2006, where he worked under the supervision of Alejandro Frangi. He now works at Philips Research, Medisys, Suresnes, France as research engineer. His main research interests are in the development of registration methods for the quantification of heart motion and deformation.

Gemma Piella received her MSc degree in Telecommunication Engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, and her PhD degree from the University of Amsterdam, The Netherlands, in 2003. From 2003 to 2004, she was a visiting professor at the UPC. She then stayed at the Ecole Nationale des Telecommunications, Paris, France, as a postdoctoral fellow. She joined the UPF in 2005, now working at the CISTIB. Her main research interests are in the design of image registration techniques and their application to cardiac imaging.

Alejandro F. Frangi received his MSc degree from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 1996, and his PhD degree from the Image Sciences Institute, University Medical Center, Utrecht, NL, in 2001. He has been visiting researcher at Imperial College, London, UK, and in Philips Medical Systems BV, The Netherlands. He is currently Associate Professor at the UPF, and Professor of Biomedical Image Computing at the University of Sheffield, UK. He is Senior Member of IEEE and Associate Editor of IEEE Transactions on Medical Imaging, Medical Image Analysis, the International Journal for Computational Vision and Biomechanics and Recent Patents in Biomedical Engineering journals.