Hierarchical Manifold Learning for Regional Image Analysis

Kanwal K. Bhatia*, Anil Rao, Anthony N. Price, Robin Wolz, Joseph V. Hajnal, Daniel Rueckert, and
THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

Abstract—We present a novel method of hierarchical manifold learning which aims to automatically discover regional properties of image datasets. While traditional manifold learning methods have become widely used for dimensionality reduction in medical imaging, they suffer from only being able to consider whole images as single data points. We extend conventional techniques by additionally examining local variations, in order to produce spatially-varying manifold embeddings that characterize a given dataset. This involves constructing manifolds in a hierarchy of image patches of increasing granularity, while ensuring consistency between hierarchy levels. We demonstrate the utility of our method in two very different settings: 1) to learn the regional correlations in motion within a sequence of time-resolved MR images of the thoracic cavity; 2) to find discriminative regions of 3-D brain MR images associated with neurodegenerative disease.

Index Terms—Disease classification, feature selection, manifold learning, motion analysis, multiscale analysis, regional manifold learning.

I. INTRODUCTION

THE HIGH dimensionality of medical images makes their interpretation and analysis a challenging task. This has led to the adoption of dimensionality reduction techniques in order to discard extraneous information or to improve interpretability of the data. Manifold learning is one such technique which has garnered interest in the medical imaging community in recent years. It aims to discover some underlying structure of the data, of much lower dimensionality, while still preserving the local structure of the data. Prior applications have included the analysis of sequences of single subjects [1], [2] as well as the investigation of whole populations [3]–[7].

In all these cases, images have been considered as a whole, with an entire image represented as a *single data point* in vector space. A single medical image, however, may consist of several anatomical structures, which may vary to different extents between subjects across a population or within an individual time sequence. Reducing the differences between two images to a single difference between two points therefore risks losing important information. In this paper, we address this issue by extending conventional manifold learning to account for regional variation within images, thereby enabling latent properties of the data to be localized. To do this, we propose a *hierarchical manifold learning (HML)* algorithm, with the aim of creating low-dimensional representations of datasets that vary spatially across an image.

A. Dimensionality Reduction

Medical images such as magnetic resonance images (MRI) or computed tomography (CT) are generally high dimensional. For example, each 2-D frame of size 100×100 pixels in a cardiac sequence can be viewed as a point in 10 000-dimensional space, while a typical 3-D brain image could be of the order of 1 000 000 dimensions. However, not all these dimensions may be useful for analysis, and may even confound interpretation: frames from the same cardiac sequence generally exhibit a high degree of similarity, while disease may only affect a subset of structures in the brain. Finding low-dimensional representations of data has therefore become an important area of research.

A variety of medical imaging applications has been shown to benefit from dimensionality reduction. Recent work has demonstrated improvements in brain image classification and regression by using feature selection of sparse sets of voxels [8]–[13]. Exploration of cardiac and respiratory motion has also utilized dimensionality reduction techniques such as principal component analysis (PCA) (e.g., for left ventricle segmentation [14], [15]) and independent component analysis (ICA) (e.g., for the extraction of features relevant to registration of cardiac perfusion images [16]). More recently, nonlinear techniques have gained popularity (see [17]) for a review), in motion estimation [1] and also in the classification of disease [18]. Manifold learning is one such technique which has shown promise in a variety of recent imaging applications.

B. Manifold Learning

An image containing n voxels can be represented as a single data point in n-dimensional space. However, it can be argued that a set of similar images is neither randomly nor uniformly distributed in this space, but instead lies on, or near, a sub-manifold of much lower dimension [4]. Manifold learning is a dimensionality reduction technique which aims to extract this *intrinsic dimensionality* by utilizing a graph representation of the set of

data points. Each data point is represented by a vertex of a graph and the weighted edges between vertices represents similarities between data points. The aim is then to find an *embedding*, $\mathbf{x}$, of this graph in m-dimensional Euclidean space, where $m \ll n$, while retaining the structure of the graph in some way.

Various manifold learning algorithms have been proposed, which differ in terms of how they preserve properties of the graph—and therefore the data-structure (see [19] for an overview). Local approaches, such as locally linear embedding (LLE) [20] and Laplacian eigenmaps (LE) [21], aim to maintain relationships between similar data points by preserving the distances between each point and its neighbors. Global methods, such as isomap [22] and semi-definite embedding (SDE) [23], instead aim to preserve both the local and global structure of the graph. In this paper, we base our work on LEs, which has shown good performance in learning manifold representations of medical images [1], [7], and also lends itself naturally to the extension to regional manifolds due to the optimization of an explicit cost function.

C. Medical Applications of Manifold Learning

Manifold learning has recently been shown to be useful in a variety of medical imaging tasks. It has been used to analyze dynamic sequences of images, such as the cardiac and respiratory cycles, as their nonlinear behavior has been shown to create challenges for linear techniques such as PCA [24], [25]. For example, in the segmentation of the left ventricle, Zhang *et al.* [2] apply ISOMAP [22] to cardiac MRI sequences to obtain an embedding more suited to level-set propagation. Ye *et al.* [18] also use ISOMAP, but as a feature selector for the classification of subjects with Tetralogy of Fallot. Duchateau *et al.* [26] use manifolds to model variation within a population of subjects with Septal Flash intra-ventricular dyssynchrony. Other work by Georg *et al.* [27] has applied manifold learning to MRI slices of the lung to assist in lung volume reconstruction. In ultrasound applications, LE have been used in both respiratory gating (Wachinger *et al.* [1]) and cardiac gating (Isguder *et al.* [28]).

Other applications have been in the analysis of populations of structural, in particular, brain images. Wasserman *et al.* [29] have applied manifold learning and clustering to cluster white matter fibre tracts from diffusion tensor MRI (DT-MRI). Gerber *et al.* [4] construct manifold representations of large sets of brain images, showing that the manifold model can be used as a statistically-significant descriptor of clinical measures associated with Alzheimer's disease. Wolz *et al.* [7] apply manifold learning to the hippocampus region of the brain in particular, in order to automatically classify subjects with Alzheimer's disease. Here, the translation of the data into a low-dimensional embedding is shown to improve classification accuracy. Aljabar *et al.* [3] describe how combining measures of shape and appearance in constructing manifold approximations of neonatal brain images can be used to create representations which are well correlated with clinical data. The manifold structure of a set of brain images has also been used in Hamm *et al.* [5] in order to improve nonlinear image registration between subjects.

In the above applications, whole images are considered as single data points when constructing the manifold representations, with a single measure of similarity used to characterize differences between the images. In some cases, this requires predefining a region of interest. For example, Wolz *et al.* [7] use a mask of the hippocampal area for Alzheimer's disease classification. In the absence of prior knowledge of which areas of an image are important, current manifold learning techniques may not work optimally. Furthermore, they do not allow regional variations within the data to be distinguished. This motivates the work in this paper on regional manifold learning, in order to explore local properties of images. We aim to construct manifold embeddings which vary smoothly across an image depending on the local properties of the dataset. To do this, we adopt a hierarchical model where images are repeatedly subdivided into smaller patches in order to enforce spatial consistency and to allow for multi-resolution analysis. This additionally enables regions of interest to be automatically determined for a given application.

D. Regional Manifold Learning

Two approaches that make use of the regional graph structure within a population have been recently proposed for registration and segmentation. In Ye *et al.* [30], registration between whole subjects is improved by registering along different population manifolds for distinct local regions within the image. However, there is no enforcement of spatial consistency within the manifold construction, with a constraint instead being applied to the overall deformations using Markov random fields. To do this, the regions where the manifolds are calculated need to be separated by some distance, and so do not cover the whole image. Cardoso *et al.* [31] use the local structure of populations for the segmentation of brain structures by label propagation. By constructing implicit local embeddings, information can be propagated across between neighboring nodes of the graph, enabling more accurate local segmentations.

We present more general algorithms for regional manifold learning which can be used to transform the data into a more interpretable space. A variety of different analyses can then be used on the resulting embeddings. In initial work [32], we suggested dividing an image into regular patches and finding an embedding for all patches simultaneously. This was shown to offer advantages over whole image, as well as independent patch embedding, in learning the cardiac and respiratory cycles in 2-D time sequences of the heart. One issue with this method is that the patch size needs to be chosen in some way, reflecting the size of the structures of the data in order to be useful. In addition, solving for all patches simultaneously makes its extension to large sets of 3-D data, or analysis at finer scales, difficult. In this paper, we continue the idea of patch embeddings; however, we take a multiscale approach to remove the need to choose patch size. This results in a hierarchy of low-dimensional embeddings created using large to successively smaller patch sizes. To ensure consistency of the embeddings across anatomy, we enforce similarity of the embeddings between hierarchy levels.

Recent work on multiscale clustering using adaptive multi-grids (AMGs) has been proposed for use on single images, for example, in the segmentation of brain tumors [33], [34]. These

take a bottom-up approach to clustering, iteratively coarsening a grid starting from a voxel level, in order to segment an individual image. AMGs have also been used to determine clusters along manifolds, such as in [35]. Similarly, diffusion wavelets [36] can be used to assess multiscale properties of a single manifold. In both these cases, data points are clustered according to their position in a manifold, however, each data point is still treated as a single entity. Our work differs conceptually from these as we aim to build regionally-varying manifolds across a set of images at different scales. The use of regular patches avoids the need to segment any images in the dataset, with each patch effectively acting as a region-of-interest over the population.

In [37], a coarse-to-fine classification framework uses graph embeddings for cancer classification. This incorporates class labels into the graph embedding with an additional sparse feature selection built into the optimization framework. While the goal of [37] is classification, we focus on more general approach to manifold learning, whereby spatially-varying manifolds are created across an image. These can then be used for different tasks including classification and regression.

The work of [38], which this paper extends, uses a simplistic hierarchical manifold alignment scheme where each new patch is aligned to a single parent. In this paper, we refer to the method as "Naïve manifold learning" and present the algorithm in more detail. A weakness of this method is that discontinuities can occur at patch boundaries, where a given patch may have a very different embedding to its immediate neighbor, due to being linked to different parents at an early stage. We refine our approach in this paper so that each patch embedding is influenced by more than one parent patch, to an extent dependent on its distance to these patches.

We use our proposed *hierarchical manifold learning* (HML) for the analysis of regional variations within the thoracic cavity. We demonstrate the utility of using the joint embedding over independent patch embeddings. Additionally, we apply it to a large population of 3-D brain images. Using this, we are able to automatically detect regions in the brain relevant to the classification of neurodegenerative disease.

II. Background

We apply our algorithm in two scenarios where manifold learning has been shown to be effective, but in which the ability to additionally learn *regional* properties of the data can open new avenues of investigation.

A. Cardiac Image Analysis

Being able to learn the cardiac cycle directly from images is of potential benefit in medical imaging. Previous work has used the manifold structure of the cardiac cycle as an aid to left ventricle segmentation [2]. Knowledge of the physical motion of the heart is also becoming of increasing interest in patient selection for cardiac resynchronization therapy (CRT), with *mechanical* dyssynchrony derived from image data shown to be a better predictor of outcome in certain patients [39], [40]. While manifold learning has been used to model populations of subjects with dyssynchrony for CRT [26], it has not been used currently to determine where within a single subject differences are occurring. Being able to automatically determine the phase differences between different regions of the heart directly from image data could therefore prove clinically useful.

Cardiac image analysis is, however, complicated by the presence of respiratory motion. This can also be analyzed using manifold learning and previous work has employed this for ultrasound gating [1] and for lung CT volume reconstruction [27]. Two techniques for measuring respiration are through respiratory bellows and free-breathing navigator-gated strategies. Bellows monitor the physical movement of the chest wall during breathing, while navigators measure the displacement of the diaphragm [41]. Both methods acquire surrogates for the respiratory motion at single physical locations which then need to be correlated with the motion of the heart. This motivates the investigation of how the physical movements at various locations within the abdominal/thoracic cavity are related, and how this can be learned directly from images.

We apply hierarchical manifold learning in order to learn the different motions occurring in a sequence of real-time cardiac MRI, obtaining spatially-varying respiratory and cardiac correlation maps.

B. Structural Correlates of Neurodegenerative Disease

Previous work [4], [5] has shown that the space spanned by a set of brain images can be usefully approximated by a nonlinear manifold of much lower dimensionality, and manifold learning has been used to build efficient representations of populations of brain images. This involves a transformation to a Euclidean coordinate system where operations such as classification can be done more successfully; reducing the dimension in this way has shown to improve classification rates in Alzheimer's disease [7], for example. Previous work, however, has solely focussed on whole images or specific segmented structures.

In our work, we use hierarchical manifold learning to automatically detect regions associated with Alzheimer's disease by investigating which areas discriminate best between healthy and diseased subjects. In addition, we show how the results can be combined to produce an overall classification of a test subject.

It should be noted that our work takes a very different approach to the techniques for sparse variable selection which have been proposed in recent literature [10], [12], [13]. In such methods, features with similar predictive power are not all guaranteed to be selected, since enforcing sparsity tends to favour solutions with few variables [42], [43]. While this may make little difference to overall classification accuracy, it may mean that some areas relevant to disease are not detected. Our method has no constraint on sparsity; instead we aim to transform the feature space to one more amenable to classification, while still respecting the local structure of the data. By encoding the local structure in the dimensionality reduction stage, we avoid the need to consider groups of voxels as classifiers. In this way, we can test image patches independently and efficiently, in order to produce spatially-varying maps showing the discriminative power of regions within an image.

III. METHODS

A. Manifold Learning

Manifold learning is a method for nonlinear dimensionality reduction which strives to find a faithful, low-dimensional characterization of an n-dimensional dataset. This is done by representing the structure of the data as a graph of d data points $G = (V, E)$, where each vertex V denotes a data point (image), and the edge weights E between vertices represent similarities between data points. The aim is then to find an embedding, $\mathbf{X} = \langle \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \rangle$, in m-dimensional Euclidean space (where $m \ll n$), which maintains the local graph structure.

1) Similarity Measures: In order to maintain the local structure of the data, the graph edge weights are set to correspond to the neighborhood (defined using k-nearest neighbors or ϵ-ball distances in the space of the original data) similarities between data points. These similarities can be encapsulated in a similarity matrix $\mathbf{W}$. One commonly-used metric is based on the ℓ_2 (Euclidean) norm distance $d_{ij} = \|\mathbf{u}_i - \mathbf{u}_j\|_2^2$ between the voxel-wise intensities $\mathbf{u}$ of each pair of images i and j. The final weights are then obtained using a Gaussian kernel function

$$W_{ij} = \begin{cases} e^{-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{2\sigma^2}}, & \text{if } i, j \text{ are neighbors;} \\ 0, & \text{otherwise.} \end{cases}$$

where σ indicates the variance of the Gaussian. Other similarity metrics used in medical imaging have included the correlation coefficient [1] as well as those based on Gabor filter responses [44].

2) Laplacian Eigenmaps: LE [21] is a particular manifold learning technique which has often been chosen for medical imaging applications [3], [1], [7]. The aim is to preserve structure in the data by ensuring that data points which are "close" in the high-dimensional space remain "close" in the low-dimensional embedding. This is done by minimizing the following cost function:

$$C(\mathbf{X}) = \sum_{ij} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) W_{ij} \qquad (1)$$

which minimizes the weighted Euclidean distance between the embedding coordinates $\mathbf{x}_i$ and $\mathbf{x}_j$ of data points i and j, respectively, in the low-dimensional embedding. The measure of closeness between points i and j is defined by the weight, W_{ij}, which indicates their similarity. For medical imaging applications, these points can represent whole images or, as will be described in this paper, image patches. One advantage of LE over other manifold learning techniques is its capacity to additionally handle nonmetric similarity measures such as information theoretic measures (such as mutual information [45] or normalized mutual information [46]). The solution to (1) is obtained using properties of the graph of the data. We define a diagonal matrix as the column (or row) sum of the weights for each vertex

$$D_{ii} = \sum_j W_{ij} \qquad (2)$$

and the graph Laplaician as an operator on the graph vertices

$$\mathbf{L} = \mathbf{D} - \mathbf{W}. \qquad (3)$$

Using these, we can rewrite the cost function as [21]

$$C(\mathbf{X}) = \sum_{ij} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) W_{ij} = 2\mathbf{X}^T \mathbf{L} \mathbf{X}. \qquad (4)$$

After adding the constraint $\mathbf{X}^T \mathbf{D} \mathbf{X} = 1$ to eliminate arbitrary scale factors in the solution, the m-dimensional solution to (4) is given by the eigenvectors $\mathbf{X}$ of the generalized eigenvalue problem

$$\mathbf{L} \mathbf{X} = \lambda \mathbf{D} \mathbf{X} \qquad (5)$$

corresponding to the m smallest nonzero eigenvalues (λ).

B. Regional Manifold Learning

Using the above methods of manifold learning requires the similarity to be defined between each data point. In medical imaging, these data points have traditionally represented whole images. However, two images may differ in more ways than can be captured by a single value. Additionally, it is not possible to identify where in the images the differences (or similarities) are occurring. It may be more informative to instead conduct manifold learning for images on a regional basis.

The idea of regional manifold learning is to construct manifold embeddings which vary spatially and smoothly through the image, respecting the local properties of the data. To avoid the need to predefine regions of interest, we partition the images into regular patches, finding a manifold embedding for each image patch. We argue that regions of an image which are physically close to one another should also possess similar low-dimensional representations, and so nearby patches should produce similar manifold embeddings. In order to encourage this, we leverage the idea of manifold alignment in LE [47], where two manifold embeddings may be solved to be close to each other in the low-dimensional space. We further extend this to solving for multiple manifolds.

1) Localized Manifold Learning: Our work on localized manifold learning [32] has shown that naïvely partitioning images into regular patches is insufficient: anatomical structures rarely fall conveniently within a regular grid. To overcome this, a weighting term between adjacent patches is added, in order to keep their low-dimensional embeddings similar, thus enforcing a degree of spatial consistency in the construction of the patch-based embeddings. One drawback of this approach is its scalability: all patches are solved for simultaneously, requiring the solution to a square matrix of side *number of patches ×number of images*. This makes it unsuitable for application to very large numbers of images or patches. Moreover, the lack of a systematic approach to selecting the patch size is also of concern. Different regions may prove to be more or less relevant at different scales and choosing a single scale may miss trends evident at other levels.
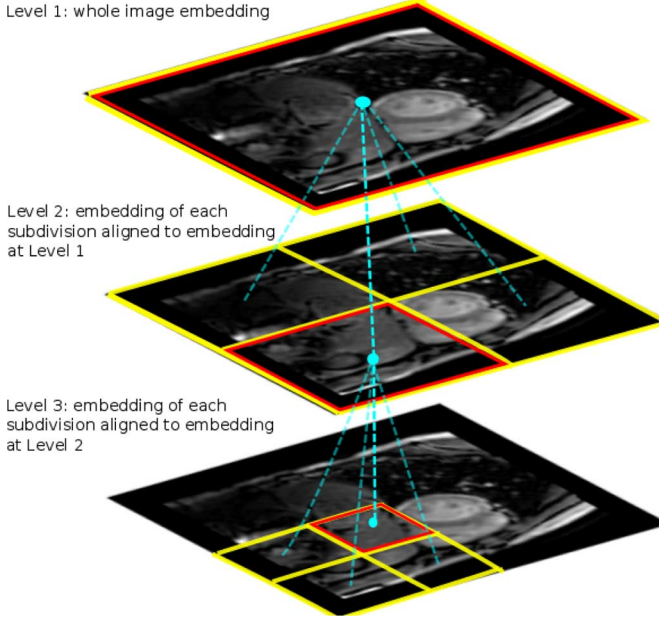
Fig. 1. Naïve hierarchical manifold learning embedding.

2) Naïve Hierarchical Manifold Learning (nHML): Hierarchical manifold learning aims to address these issues. Given the assumption that sets of whole images occupy a low-dimensional subspace, it can also be argued that corresponding areas contained within these images should also occupy similar subspaces, depending on the degree of overlap. Furthermore, neighboring areas within a single image should again lie in similar subspaces and therefore have similar embeddings. This drives the idea of hierarchical manifold learning.

The methodology is based on that of manifold alignment [47] and is illustrated in Fig. 1. Given a first level embedding of the full image in the standard way, we can then recursively subdivide the image into regular patches (for example, into four patches for a 2-D image and into eight for a 3-D image). As each subdivision is contained within the current patch, we would expect embeddings at successive levels to be close. In [38], we presented an initial version of the algorithm where we obtain the new embedding, $\mathbf{X}$, of one of these subparts by additionally aligning it to the embedding of the "parent" patch, $\hat{\mathbf{x}}$, from which it was obtained (see Fig. 1). The LE cost function is augmented to give

$$C(\mathbf{X}) = (1 - \mu) \sum_{ij} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) W_{ij}$$
$$+ \mu \sum_i (\mathbf{x}_i - \hat{\mathbf{x}}_i)^T (\mathbf{x}_i - \hat{\mathbf{x}}_i) \quad (6)$$

where $0 \leq \mu < 1$ is a weighting parameter to determine the influence of each term. The first term represents the standard LE cost function of an individual patch, while the second term constrains a given patch embedding to be close to the embedding of its parent patch. Low values for μ reduce the strength of the inter-manifold alignment and the embedding moves towards the embedding of individual patches. Higher values for
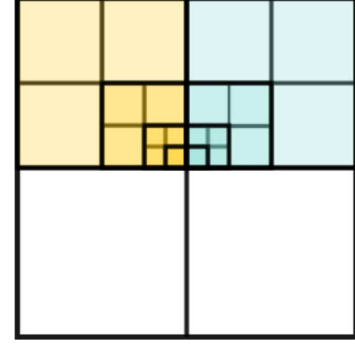

Fig. 2. Neighboring patches may have very different embeddings due to parents.

μ, in contrast, lead to closer aligned embeddings. From (4), we can rewrite (6) as

$$C(\mathbf{X}) = 2(1 - \mu)\mathbf{X}^T \mathbf{L} \mathbf{X} + \mu \sum_i (\mathbf{x}_i - \hat{\mathbf{x}}_i)^T (\mathbf{x}_i - \hat{\mathbf{x}}_i). \quad (7)$$

The optimal solution to this can be found by differentiating with respect to $\mathbf{x}$

$$C_{\mathbf{X}} = 4(1 - \mu)\mathbf{L}\mathbf{X} + 2\mu(\mathbf{X} - \hat{\mathbf{X}})$$
$$= [2\mu + 4(1 - \mu)\mathbf{L}]\mathbf{X} - 2\mu\hat{\mathbf{X}}.$$

Setting this to zero at the minimum, gives

$$2\mu(\mathbf{X} - \hat{\mathbf{X}}) + 4(1 - \mu)\mathbf{L}\mathbf{X} = 0.$$

For $\mu = 0$ this reduces to $4\mathbf{L}\mathbf{X} = 0$ which is the same as for the standard LE embedding. This is solved by adding the constraint described in [4]. For $\mu > 0$

$$2\mu\mathbf{X} - 2\mu\hat{\mathbf{X}} + 4(1 - \mu)\mathbf{L}\mathbf{X} = 0$$
$$\mu\mathbf{X} + 2(1 - \mu)\mathbf{L}\mathbf{X} = \mu\hat{\mathbf{X}}.$$

Giving

$$\mathbf{X} = (\mu\mathbf{I} + 2(1 - \mu)\mathbf{L})^{-1}\mu\hat{\mathbf{X}}. \quad (8)$$

The computational complexity therefore scales linearly with the number of patches, making the solution feasible even for very fine patch sizes. Note that since the Laplacian, by construction, is positive semi-definite, and μ and $(1 - \mu)$ are always positive, the inverse in (8) is guaranteed to exist.

3) Hierarchical Manifold Learning (HML): One problem with the previous method is that it may give rise to discontinuities at patch boundaries: after several subdivisions, two adjacent patches may have very different embeddings due to having different parents at an coarse level (Fig. 2). We therefore propose to align each patch embedding, not just to one parent manifold, but to several, with the strength of the aligning constraint dependent on the distance between patch centers (Fig. 3). We
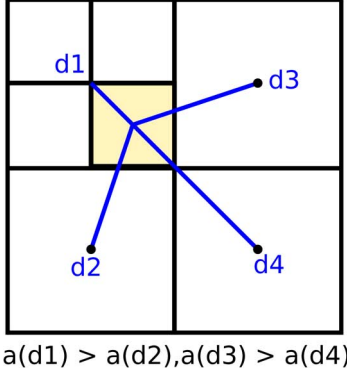
a(d1) > a(d2), a(d3) > a(d4)

Fig. 3. Hierarchical manifold learning: each patch aligned to four closest parent embeddings with weights a, where a is larger if the distance between patch centers is smaller.

use a_p to indicate the weight of alignment to existing coordinates $\hat{\mathbf{X}}_p$ of parent patch p. The aim is to minimize the cost function

$$C(\mathbf{X}) = (1-\mu) \sum_{ij} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) W_{ij}$$
$$+ \mu \sum_i \sum_p a_p (\mathbf{x}_i - \hat{\mathbf{x}}_p)^T (\mathbf{x}_i - \hat{\mathbf{x}}_p) \quad (9)$$

where the first term is the standard LE cost between patch i and patch j, and the second term aims to minimize the Euclidean distance between the new embedding, $\mathbf{X}$, of patch i, and the existing coordinates $\hat{\mathbf{X}}_p$ of each parent patch p. The weighting factor μ dictates the relative strength of each of these terms, and a_p determines the influence of each parent patch p where $\sum_p a_p = 1, a_p > 0$. Following from (4)

$$C(\mathbf{X}) = \mu \sum_p a_p (\mathbf{X} - \hat{\mathbf{X}}_p)^T (\mathbf{X} - \hat{\mathbf{X}}_p) + 2(1-\mu)\mathbf{X}^T \mathbf{L} \mathbf{X}$$
$$C_{\mathbf{X}} = 2\mu \sum_p a_p \mathbf{X} - 2\mu \sum_p a_p \hat{\mathbf{X}}_p + 4(1-\mu)\mathbf{L}\mathbf{X}. \quad (10)$$

At the minimum

$$C_{\mathbf{X}} = 2\mu \mathbf{X} \sum_p a_p - 2\mu \sum_p a_p \hat{\mathbf{X}}_p + 4(1-\mu)\mathbf{L}\mathbf{X} = 0. \quad (11)$$

If $\mu = 0$ this reduces to $4\mathbf{L}\mathbf{X} = 0$ which is solved using standard LE after adding a constraint as in [4]. If $\mu > 0$, we can write

$$\left(\mu \mathbf{I} \sum_p a_p + 2(1-\mu)\mathbf{L} \right) \mathbf{X} = \mu \sum_p a_p \hat{\mathbf{x}}_p.$$

Giving

$$\mathbf{X} = \left(\mu \mathbf{I} \sum_p a_p + 2(1-\mu)\mathbf{L} \right)^{-1} \mu \sum_j a_p \hat{\mathbf{X}}_p. \quad (12)$$

As $\sum_p a_p$ is always positive, the inverse is guaranteed to exist. Again, this is a linear equation which avoids the need to solve the eigenvalue problem for very large matrices as in [38]. We can therefore keep recursively subdividing to very fine image scales to produce sensitive spatially-varying manifold maps. By constraining each embedding to be close to those at the previous level, we avoid issues with smaller patches being dominated by noise. However if it is known that each patch contains completely discrete information, a value of $\mu = 0$ might be preferable.

C. Patch Division

When no prior information about structure size is available, hierarchically dividing the patches allows all patch levels to be examined, since manifolds are created at every level and can be analyzed independently. This can be used to discover which scales are most informative at different spatial locations in the data. However, in order to reduce computation, the divisions could be stopped when there is a likelihood of little change with further divisions, given by stability of a measure such as the entropy of the child patches.

D. Weighting Terms

Two parameters need to be chosen when calculating the embedding of a new patch. a_p dictates the relative strengths of alignment to each of the p parent patch embeddings. Additionally, the overall weighting term μ determines how strongly the alignment to the parent patches should be, relative to the independent embedding of the patch itself.

For a 2-D image, it would seem natural to constrain a new patch to be close to its four nearest (in terms of distance) parent patches. Similarly, eight parent patches could be used for a 3-D image. However, it could also be possible to determine the number of parents based on patch size, with more being used as the patch size becomes smaller (in order to enforce greater spatial consistency between smaller patches). For our experiments, we simply use four (2-D) and eight (3-D) parent patches, with the weighting term being given by $a_p = (1/d_p)/\sum_p(1/d_p)$, where d_p is the Euclidean distance between the center of the patch under consideration and its parent patch p.

The optimal value for μ for a particular application should be determined using cross-validation. However, an intuitive choice could be based on the fact that higher values for μ result in more uniform embeddings, whereas lower values give more discernibility between patches. If there is no shared information between patches for all data points, then setting $\mu = 0$ might also be appropriate. Again, the value chosen could also vary depending on the current patch size. Large patches may hold much information that we would like to be able to separate in subsequent embeddings. However, if a patch size is very small, it is likely to contain more similar information to its parent patch. Additionally, its intensities are likely to be influenced by noise to a greater extent—a higher value for μ could therefore be used to reduce unwanted effects of this.

E. Overlapping Patches

While the smoothing parameters reduce discontinuities between adjacent patch embeddings, it is also possible to use over-

lapping patches to achieve this. To do this, the values for α can still be chosen to be inversely proportional to the distance between patch centers and HML can be applied as described above. The manifold embeddings of the areas where patches overlap can be interpolated from the coordinates obtained from each overlapping patch, in order to maintain spatial consistency.

IV. RESULTS

We demonstrate properties of the algorithm using datasets of cardiac MRI sequences. Additionally, we show results on 3-D brain MR data from the Alzheimer's Neuroimaging Disease Initiative [48].

A. Cardiac Image Analysis

1) Materials: Healthy volunteers were imaged by real time free-breathing MR, using a balanced steady state free precession (SSFP) sequence with spatial resolution of $2 \times 2 \times 10$ mm and temporal resolution of 117 ms per frame ($FA/TE/TR = 20/1.2/2.4$ ms). Short axis, two-chamber and four-chamber cardiac views were acquired for 200 dynamics (the last 180 being used for analysis) in each acquisition. Each image frame is of size 128×128 pixels. Traces from ECG and respiratory bellows were recorded alongside sequence event markers (including ECG R-wave peak detection and image acquisition points) in order to align images to physiological motion.

2) Methods: We apply nHML and HML to a time-varying sequence of 2-D cardiac images. At each level, and at each patch, we construct a graph where each node represents part of a 2-D frame at one point in time, in a similar way to the conventional method of manifold learning. Since we expect similar intensity contrasts for all frames, we weight each pair of frames using their ℓ_2 (Euclidean) norm distance $d = \sum_{ij} \|\mathbf{u}_i - \mathbf{u}_j\|^2$ between the intensities $\mathbf{u}$ of each image. The final weight between the frames i and j is obtained by using a heat kernel: $W_{ij} = e^{-\sum_{ij} \|\mathbf{u}_i - \mathbf{u}_j\|^2 / 2\sigma^2}$, where σ was chosen empirically to be 2 in order to ensure a sufficient neighborhood size. To preserve only the local structure of the graph, we sparsify the graph, using a k-nearest neighbor approach by setting $k = \max(24, c)$, where c is the minimum number of nearest neighbors needed to maintain a fully-connected graph. Selecting a limited number of nearest neighbors is a standard step in manifold learning techniques. This is motivated by the aim of preserving only local structure of the data in order to achieve dimensionality reduction. In this respect, large Euclidean distances between data points of high dimensionality are unnecessary and may be misleading. As we are trying to recover both the cardiac and respiratory motions, we reduce the data to two dimensions.

To assess how well the manifold learning characterizes motion, we correlate the embedding coordinates for each patch with surrogate markers for the respiratory and cardiac cycles. The surrogate marker for the respiratory cycle is given directly by the bellows trace recorded alongside the acquisition. In order to define a proxy for the cardiac cycle, we manually segmented the left ventricle (LV) in each frame of the sequence. The area of the LV in a given frame was then used as a surrogate marker for its position in the cardiac cycle. In both cases, the magnitude of the
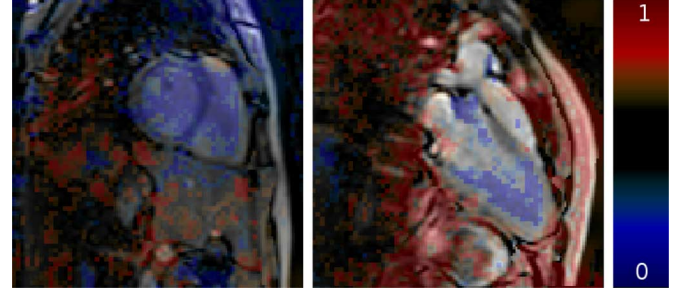

Fig. 4. Correlation magnitudes of HML embeddings with respiratory bellows for short-axis (left) and long-axis (right) acquisitions for patch size 2×2 pixels.
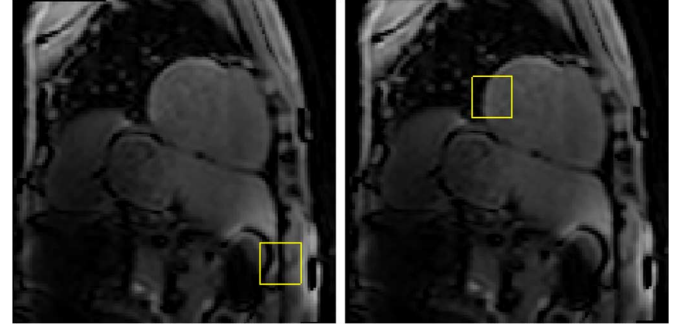

Fig. 5. Examples of noncardiac (L) and cardiac patches (R) analyzed in Figs. 6 and 7.

correlation coefficient between the embedding coordinates ($\mathbf{x}$) and the surrogate marker ($\mathbf{y}$) is given by

$$\rho = \left| \frac{\mathrm{Cov}\,(\mathbf{x}, \mathbf{y})}{\sigma_{\mathbf{x}} \sigma_{\mathbf{y}}} \right| \tag{13}$$

where $\mathrm{Cov}\,(\mathbf{x}, \mathbf{y})$ represents the covariance of $\mathbf{x}$ and $\mathbf{y}$, and $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{y}}$ denote the standard deviations of $\mathbf{x}$ and $\mathbf{y}$, respectively. Additionally, we calculate the correlation coefficients for individual patches, embedded independently using conventional LE, and also for the average intensity of each patch.

3) Correlation Maps: Fig. 4 shows maps of the magnitude of the correlation of each patch (of size 2×2 pixels) embedding in two sample acquisition views with the acquired bellows trace in each instance, obtained using HML. In these cases, the value of μ was varied depending on patch size, from 0.1 at the first level embedding (patch size 64×64 pixels) to 0.8 at the final embedding (patch size 2×2 pixels).

4) Comparison of Embeddings for a Given Cardiac and Non-cardiac Patch: By using a manifold embedding dimensionality of two, each patch in each time frame is represented by a 2-D coordinate. We can compare the coordinate in each dimension separately, with the respiratory cycle (as obtained from the bellows) and the cardiac cycle (as given by the segmented area of the left ventricle) for each patch. Fig. 5 shows two example patches of size 16×16 pixels, one in the region of the heart, which can be expected to demonstrate both respiratory and cardiac motions, and one outside the heart which is primarily subject to respiratory motion only. The noncardiac patch shown is chosen to be close to the surface of the body in order to be representative of the movement detected by the respiratory bellows. Plots of the
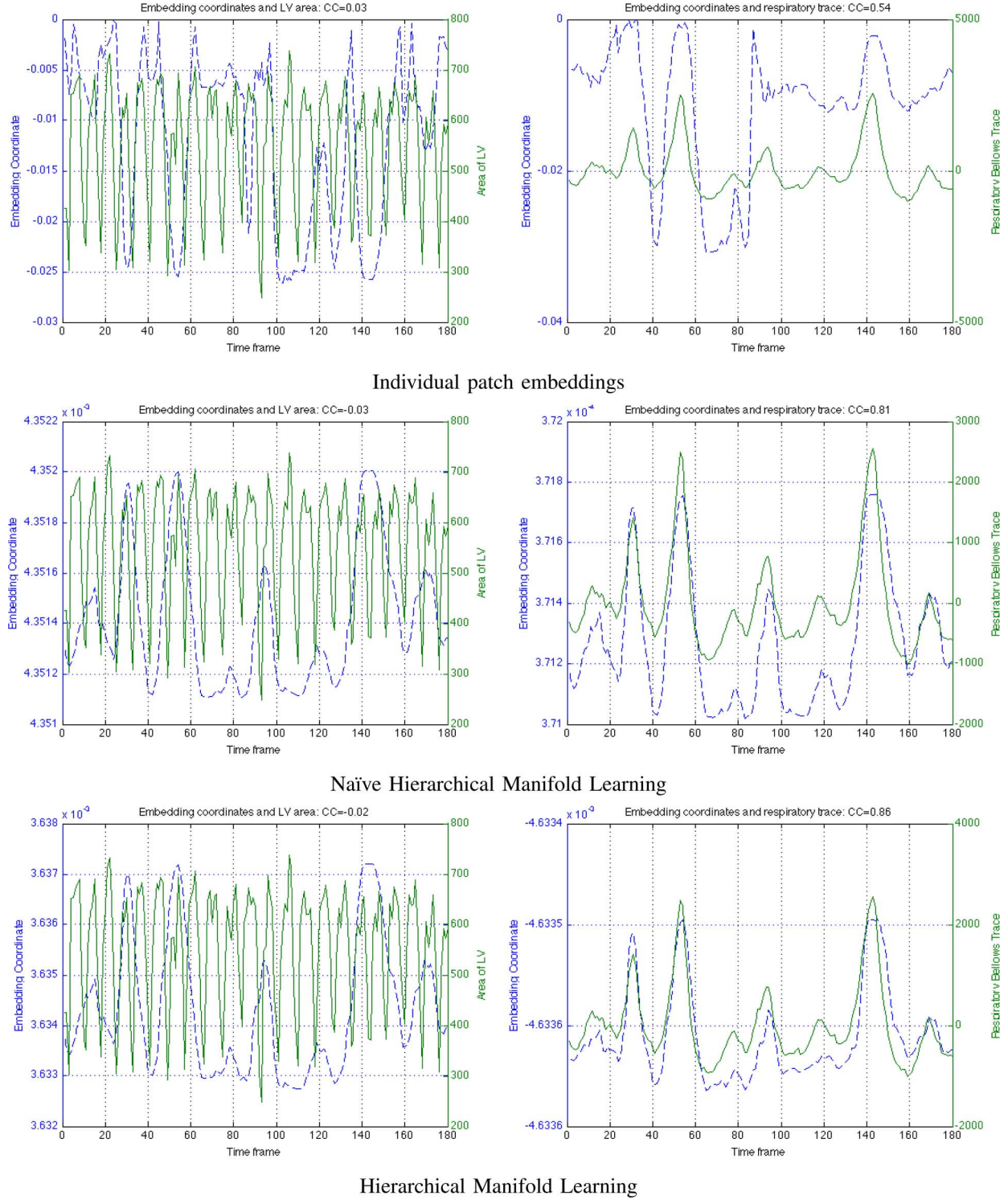
Fig. 6. Each dimension of 2-D manifold embeddings (dashed lines) for patch located outside cardiac region compared to respiratory (R) and cardiac (L) traces (solid lines), together with the magnitude of the correlation coefficients, for independent embeddings (top row), nHML (middle row), and HML (bottom row).

coordinates obtained for each time frame, alongside their respiratory and cardiac surrogate traces can be seen in Figs. 6 and 7, for the noncardiac and cardiac patches, respectively. In these examples, the value for μ is varied according to the hierarchy level: (0.1 at 64 pixels, 0.2 at 32 pixels, and 0.4 at 16 pixels). It can be seen that, for these patches, an individual patch embedding fails to capture the respiratory motion. Using the spa-

tial information in naïve hierarchical and weighted hierarchical manifold learning schemes allows more consistent recovery of the respiratory motion for both patches.

5) Comparison of NHML and HML and Independent Embeddings: Fig. 8 shows maps of correlation for each patch in an image sequence with the respiratory bellows trace for patch sizes of side 32, 16, and 4 pixels. For the sake of clarity, μ has
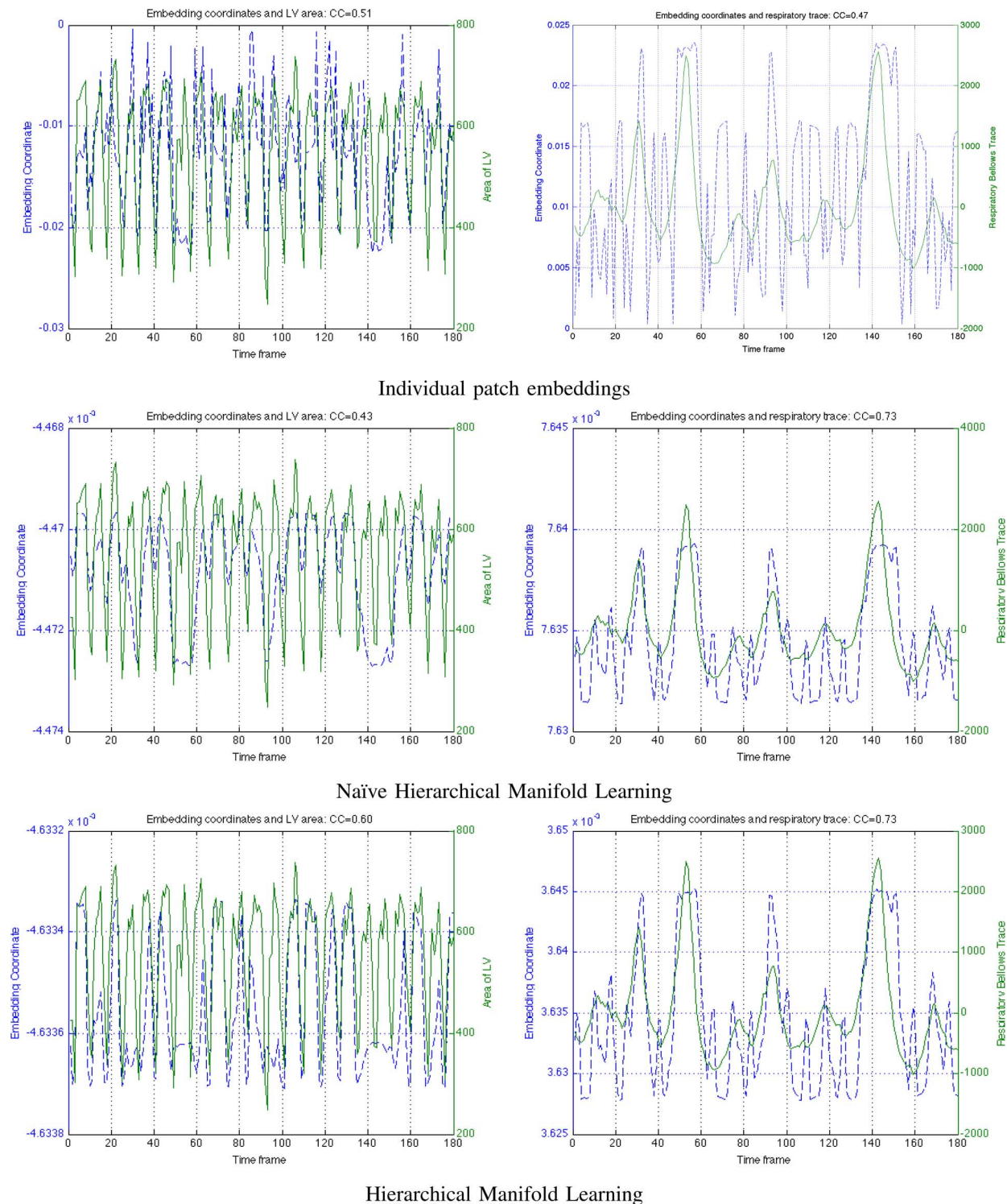
Fig. 7. Each dimension of 2-D manifold embeddings (dashed lines) for patch located within cardiac region compared to respiratory and cardiac traces (solid lines), together with the magnitude of the correlation coefficients, for independent embeddings (top row), nHML (middle row), and HML (bottom row).

been held fixed at 0.2 throughout. As can be seen, simply taking the average intensity of each patch creates very noisy results. The independent manifold embedding of each patch individually also produces noisy correlation maps, particularly when using smaller patch sizes. It is particularly difficult to obtain strong correlations in the area around the heart as this area is subjected to both cardiac and respiratory motions, thus compli-

cating analysis. However, by incorporating spatial information, the nHML and HML schemes produce smoother embeddings and more realistic correlations in the region of the heart. The correlation between the patch embeddings and the area of the left ventricle is shown in Fig. 9. It can be seen that using the weighted HML scheme is less susceptible to unrealistic borders at patch boundaries (evident in the third row of Fig. 9) when
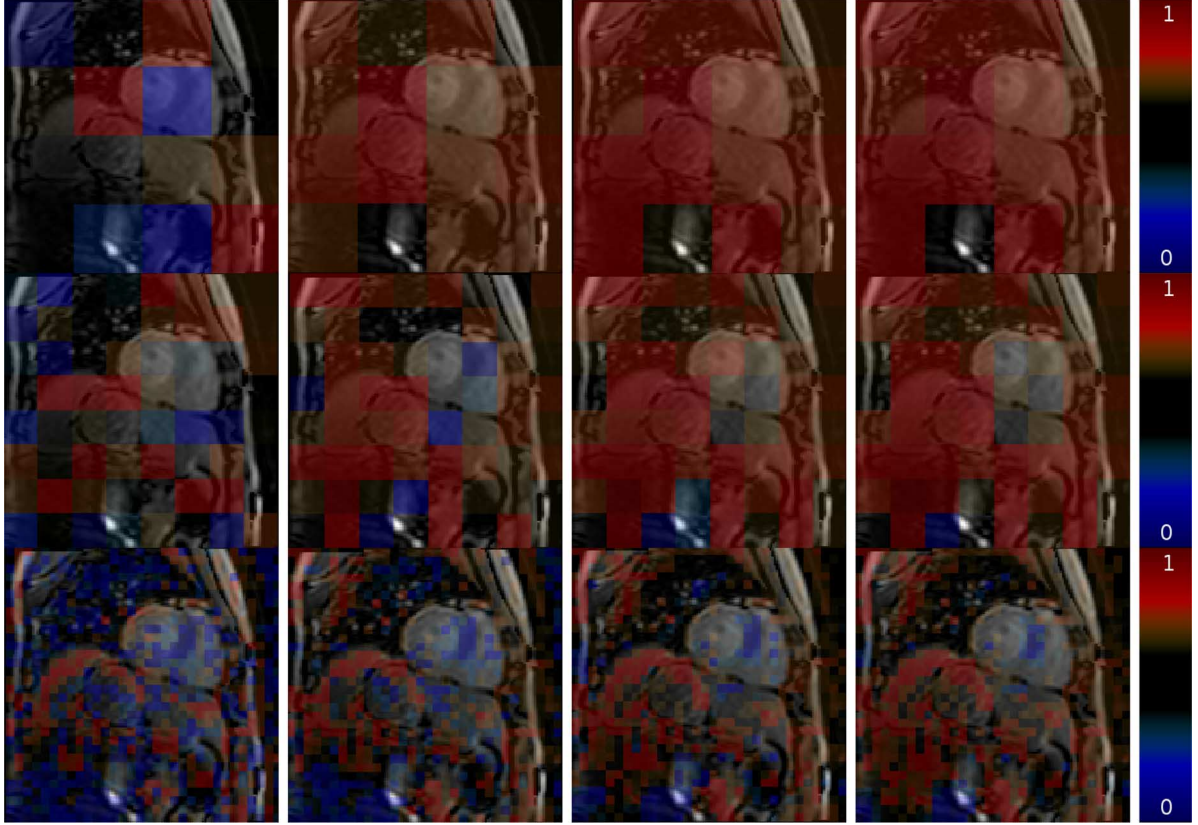
Fig. 8.   Correlation with respiratory bellows trace at various manifold resolution levels for $\mu = 0.2$. L-R: patch intensities, individual patch embedding, nHML, HML.

using a fine patch size. Note that the area in the bottom left corner showing strong correlation with the cardiac cycle contains an artery which changes intensity according in line with the cardiac cycle.

Fig. 10 shows the correlation of each patch with the area of the right ventricle (RV). The volume of the right ventricle is known to vary with respiration, hence the correlation with areas outside the heart such as the diaphragm.

Patch correlation with features obtained using PCA are shown in Fig. 11. Correlation is poorer than with manifold learning due to the nonlinear motion involved. Additionally, when very small patches often do not contain enough information to lead to a good correlation. This is overcome with the alignment term in HML.

6) Effect of Varying Parameter μ: We next show how the weighting parameter μ affects the results of nHML and HML. From (6) and (9), it can be seen that a higher value for μ should make the embeddings smoother, while a lower value shifts the embeddings closer towards independent patch embedding. Figs. 12 and 13 show this in the resulting correlations with the respiratory and cardiac cycles, respectively. It could be argued from this that if the feature to be extracted is likely to be highly localized, a smaller value for μ would be preferable.

7) Computational Aspects: Table I compares the computational time required for MATLAB implementations of nHML, HML, and the simultaneous embedding of [32], for varying patch sizes, on a 2.4-GHz Intel Core i5 processor. Both nHML

and HML allow the embeddings of much smaller patch sizes to be efficiently computed.

B. Investigating Structural Correlates of Disease

To demonstrate how the algorithm can be used for large sets of 3-D data, we have applied HML to the Alzheimer's Disease Neuroimaging Initiative (ADNI [48]) dataset of brain images. More details about the ADNI study are described in the Appendix.

The goal is to automatically discover structures in the brain associated with Alzheimer's disease. The dataset includes 429 3-D MRI subjects comprising 231 normal controls (NC) and 198 subjects with Alzheimer's disease (AD).

1) Preprocessing: We use images of size $160 \times 192 \times 160$ voxels. The data are skull-stripped and intensity normalized to a chosen member of the subject group using a linear regression model. Alignment to MNI space is performed using affine and nonlinear components to account for global differences in the population. The nonlinear transformation model adopted is a free-form deformation model based on B-splines [49], with a relatively large control point spacing of 10 mm, in order to preserve detailed differences in the images.

2) HML Parameters: In choosing the weighting between manifold levels, μ, we need to balance distinguishability between voxels with maintaining spatial consistency. We vary the value of μ depending on the size of patch, under the intuition that as patches get smaller, their embeddings should stay more
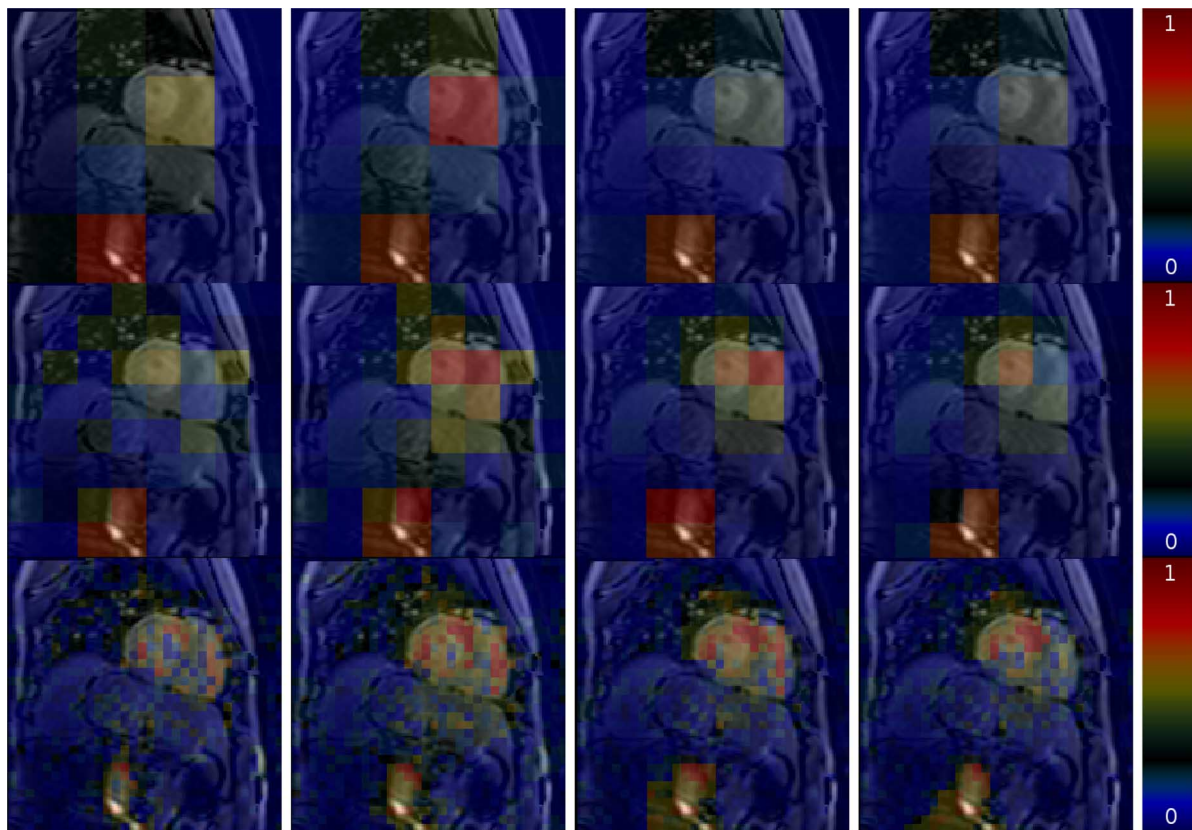
Fig. 9. Correlation with LV area at various manifold resolution levels for $\mu = 0.2$. L-R: patch intensities, individual patch embedding, nHML, HML.

similar to the previous level. For patch sizes of side 80, 40, 20, 10, and 5 voxels, we choose $\mu = 0.05, 0.1, 0.15, 0.20, 0.25$, respectively.

The similarity between data points used is a heat kernel function of the Euclidean distance between each patch of the aligned data, as defined in Section III-A1, with $\sigma = 1$. Other potential measures could be based on the magnitude of the deformation used in the alignment, or a combination of deformation and intensities as in [7]. The graph neighborhood is selected to be as many neighbors as needed to keep the graph fully connected. We construct the manifolds using the 429 AD and NC subjects, reducing to five dimensions to maintain comparability to other applications of manifold learning to brain imaging [7].

3) Classification: To investigate the structural correlates of disease, we assess how well each patch can discriminate between AD and NC subjects. For each patch individually, we apply a support vector machine (SVM) classifier [50] to its manifold coordinates. For this, we use svmtrain from MATLAB, using a linear (dot product) kernel and default parameters, as in previous work [7]. We use a ten-fold cross-validation scheme (where the same training and test subjects are used for each patch). Fig. 14 shows the atlas of the population in MNI space, overlaid with the $5 \times 6 \times 5$ voxel patches found to have classification accuracies of over 70%. The most discriminative patches can be seen to be concentrated around the medial temporal lobe, in keeping with established results [51], [52].

In order to obtain a single, overall classification of a test subject, we take a modal vote of the classification results for each patch, weighted by its accuracy in classifying the training data.

In Table II, we show the effect of including only those patches which have produced accuracies above a certain threshold in training in the final vote. No result is shown in cases where any of the training sets did not include any patches above the threshold. For comparison, a standard LE of the set of whole images results in a 65% classification accuracy using the same ten-fold cross-validation with SVM classification. For comparison, we also include in Fig. 15 the result of voxel-based morphometry (VBM) using SPM8 [53]. This required the images to be passed through the standard VBM preprocessing pipeline: firstly, grey matter (GM) was segmented from the images and the segmentations were coarsely registered to the MNI template [54]. Volume preservation, i.e., "modulation," was used throughout to ensure that the resulting images have the same volume of GM as the native segmentations. Finally, the images were smoothed using an isotropic Gaussian 8-mm FWHM kernel to account for residual misalignment of the images. Correction for multiple comparisons was done using family-wise error correction. Note however, that VBM only highlights areas showing differences between groups and does not give a classification for each subject.

4) Comparative Results on Medial Temporal Lobe: To investigate further, we examine in more detail an area around the medial temporal lobe as shown in Fig. 16. We continue subdividing the manifolds to reach patch sizes of $2 \times 3 \times 2$ voxels. To demonstrate the benefits of the spatial consistency term, we compare the results of our algorithm to those obtained when using standard LE manifold learning on each patch independently (IND), as well as on the whole image. In all cases, we
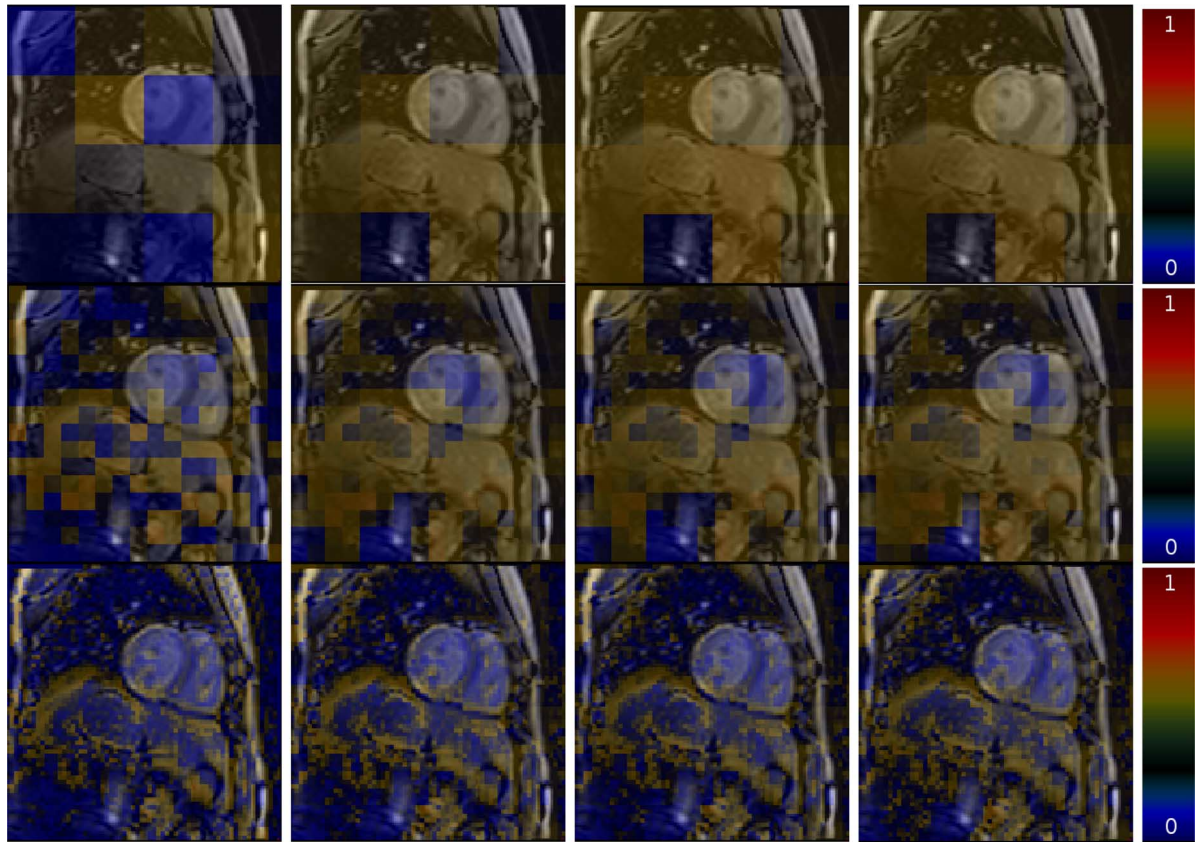
Fig. 10.   Correlation with RV area at various manifold resolution levels for $\mu = 0.2$. L-R: patch intensities, individual patch embedding, nHML, HML.

use five embedding dimensions and a distance function based on the Euclidean norm, as for the whole brain image embedding. For the HML implementation, we increase the value of μ by increments of 0.05, from 0.1 at $32 \times 48 \times 32$ voxels, to 0.25 at $2 \times 3 \times 2$ voxels.

The classification is performed using the same methods as in the case of the whole brain. Ten-fold cross-validation is again used, with each training set used to score the $2 \times 3 \times 2$ voxels patches in terms of classification accuracy. The top 1%, 2%, and 5% of these patches are then used to classify the test data, with a weighted modal vote giving the final classification results shown in Table III. The best accuracy was found to be 84%; in comparison, the standard LE embedding of the whole region gives a classification accuracy of 74% using the same train and test subsets. Regional manifold learning can therefore lead to a significant improvement over whole image embedding, even when the image is preselected to contain highly discriminative features.

Although the classification rates obtained using HML are comparable to those of independent patch manifold learning, HML is more reliable at determining regions of interest associated with disease. Fig. 17 shows $2 \times 3 \times 2$ voxels patches producing classification accuracy rates of over 70% at different example locations including the amygdala, hippocampus, and entorhinal cortex, for both HML and independent manifold learning on each patch. It can be seen that more patches are discovered with HML. We believe that this is more desirable

as the aim of our work is not solely in classification, but also in determining what the structural correlates of disease are: we therefore want to determine which anatomical structures are affected by disease and to what extent.

We can see that incorporating the spatial information in building the manifolds enables finer structures relevant to Alzheimer's disease to be discovered. As we would expect, the amygdala, hippocampus and entorhinal cortex are shown to be important in the classification of AD and therefore likely to be affected by the disease. It can be seen that the HML displays more clarity in determining anatomical regions of interest, particularly in narrow, elongated structures, where naïve classification fails to detect differences in the populations.

V. DISCUSSION

We have presented Hierarchical Manifold Learning algorithms to explore regional variations within image data. Our algorithm does not require preselecting regions of interest or patch size, and can be used to obtain embeddings even at very fine scales. By incorporating localized spatial relationships of the data into the construction of the manifold for each patch, we are able to treat each patch embedding independently in subsequent analysis. We have found that this gives smoother results than if the manifold embeddings for each patch are obtained without any spatial constraint.

We have shown how HML can be used to determine local correlations within the cardiac and respiratory cycles in image
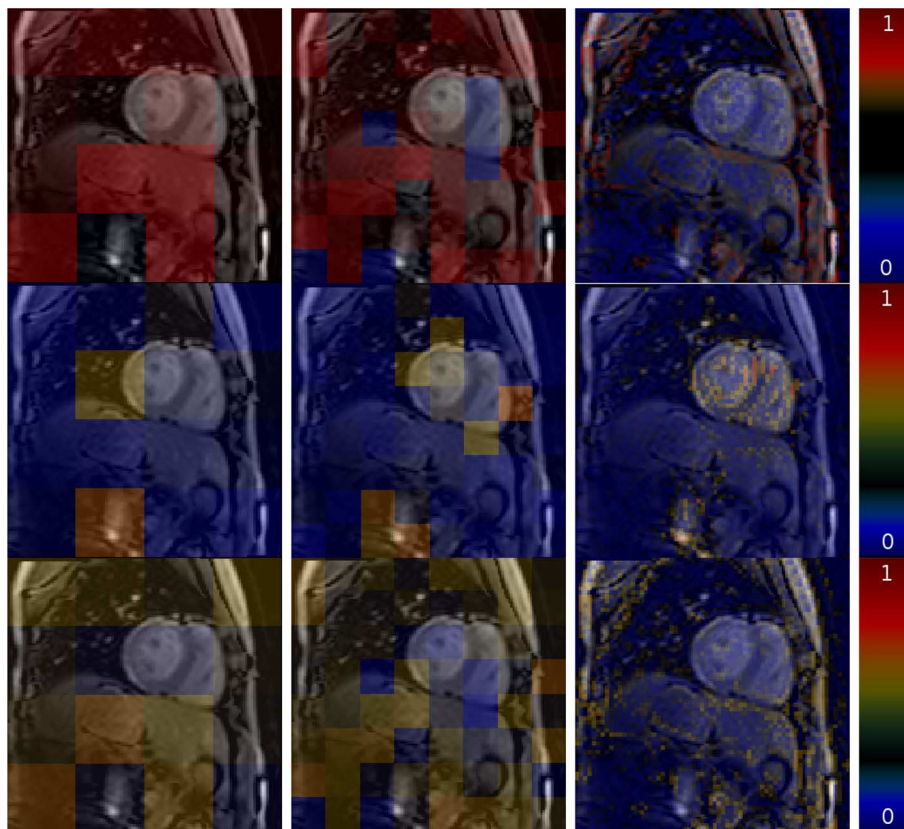
Fig. 11. Patch-based correlation using PCA. Top row: correlation with respiratory bellows trace; middle row: correlation with segmented area of the left ventricle; bottom row: correlation with segmented area of right ventricle.

sequences of the thoracic cavity. Additionally, we have demonstrated the scalability of the algorithm by applying it to a population of 3-D whole brain image datasets, in order to identify anatomical areas of interest associated with Alzheimer's disease. Patches found to be particularly predictive (as obtained using training data) can then be used as features to classify new test subjects. It should be noted that, although our method identifies a small number of areas of the image relevant to classification, there is no requirement or bias towards sparse results. One issue with sparsity-inducing methods is that if several areas have a similar predictive performance, it is not guaranteed that all of these areas would be selected. In the application of finding structures correlated with disease, we would like to uncover all relevant regions.

Our method is a general one, and not tailored to a particular application such as classification or regression. Instead, the objective is to transform the data into a space where classification and regression techniques (or other analysis) can be more reliably applied, while spatial relationships between patches are maintained. Of course, it is possible to build the knowledge of the task into constructing the manifolds, and this is likely to lead to improved results for that particular goal. For example, in the task of disease classification, knowledge of the labels of the training data could be incorporated into the graph weights, to ensure that same pathologies are kept close together in the manifold.

A. Parameters

The key parameter that needs to be tuned in our algorithms is the weighting factor between successive levels of the hierarchy, μ. We have described intuitive factors which motivate the choice of μ, namely, that higher values should be used if more smoothness is desirable (and therefore higher values for smaller patches), and lower values when greater distinction between patches is needed. However, these could be estimated in cross-validation of the particular task under consideration, as should any parameters used in classification schemes (e.g., for thresholding the number of discriminative patches).

Another parameter that needs to be chosen is the number of dimensions to which to reduce the data. This could also be done using cross-validation for the particular task. Note, however, that the embedding coordinates of the existing dimensions remain the same as new dimensions are added, so the cross-validation can be performed with a single dimensionality-reduction step. The parameters in this paper have been chosen for the particular tasks of motion estimation and in order to maintain comparability to previous work in Alzheimer's disease detection. However, parameters could also be optimized such that they produce embeddings consistent with preserving local structure of the dataset. Measures such as the residual variance, which assess the correlation between distances between points in high- and low-dimensional space, could be used to find the intrinsic dimensionality of a dataset.
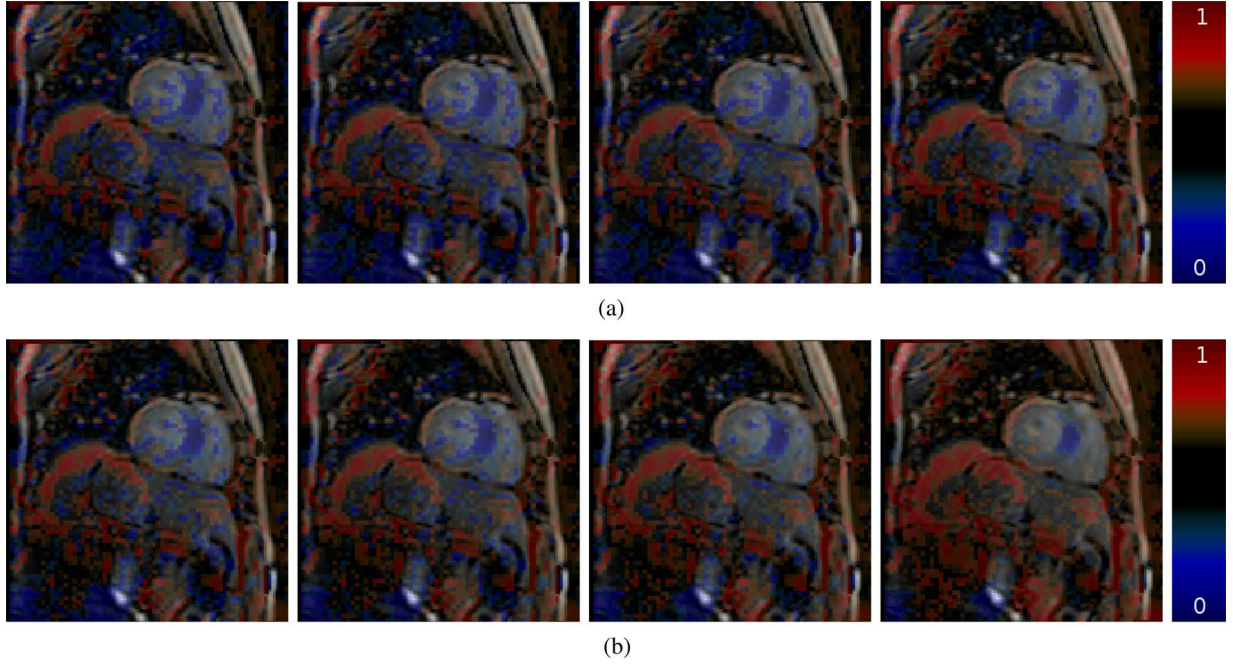
Fig. 12. How embeddings of nHML (first row) and HML (second row) vary with weighting parameter μ: correlation with bellows trace for (L-R) $\mu = 0.2, 0.4, 0.6, 0.8$. (a) nHML. (b) HML.
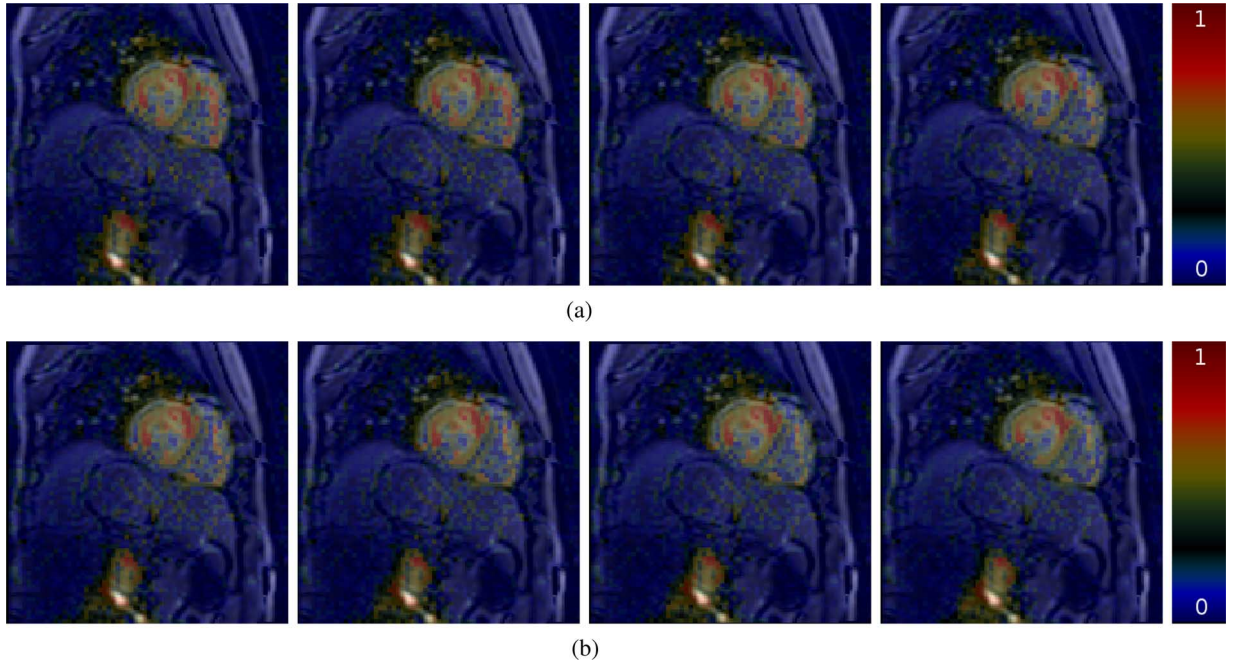

Fig. 13. How embeddings vary with parameter μ for nHML (first row) and HML (second row): correlation with LV area for (L-R) $\mu = 0.2, 0.4, 0.6, 0.8$. (a) nHML. (b) HML.

B. Classification

We have outlined a very simple strategy for classification and there are many areas which we believe can be improved. For the similarity measure used to determine the graph weights between subjects, we have currently only used the Euclidean norm distance of image intensities. Although the images have been intensity-normalized, the presence of disease is likely to alter the sizes of structures and therefore distort their intensity histograms, potentially leading to unreliable normalization. A preferable—but more computationally intensive—method would be to combine intensity differences with deformations required to warp one subject to another as in [47], [7]. This would require pairwise nonlinear deformations to be computed between each pair of subjects.

A variety of classifiers could be applied in the resulting manifold space (SVM, linear discriminant analysis, random forests) and a variety of different ensemble methods could be used to

TABLE I
TIME IN SECONDS TO RUN EACH EMBEDDING FOR
VARIOUS PATCH SIZES

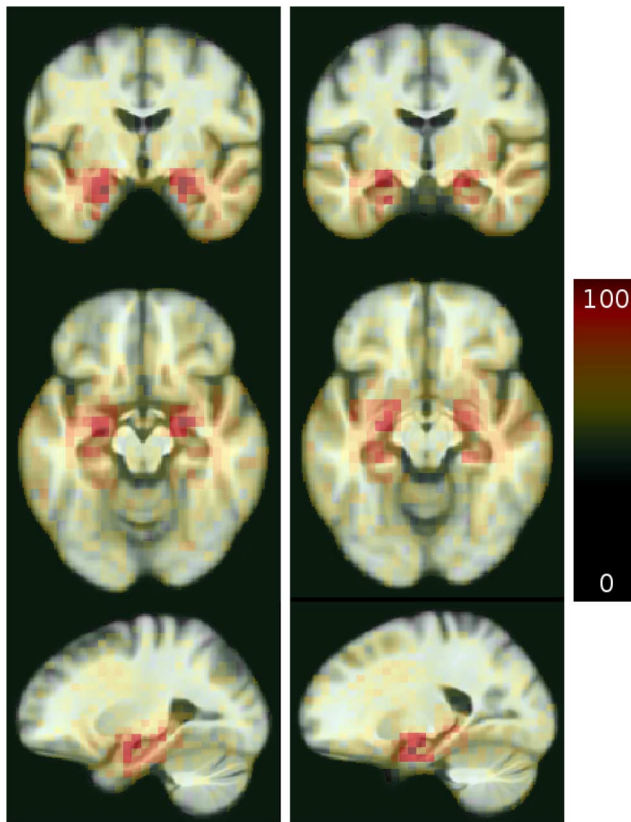| Patch size (pixels)	HML	nHML	SML [32]
64	1.18	1.14	0.22
32	1.66	1.72	16.1
16	2.80	2.74	4872

Fig. 14. Classification accuracy of patches of size $5 \times 6 \times 5$ voxels in discriminating between ADs and NCs.

Fig. 15. Voxel based morphometry showing differences between normal controls and ADs thresholded at a significance level of $p < 0.05$.

Fig. 16. Region of interest around medial temporal lobe.

TABLE II
WEIGHTED MODAL CLASSIFICATION ACCURACY (%) USING PATCHES
OVER VARYING THRESHOLDS OF ACCURACY IN TRAINING

Patch side length (voxels)	160	80	40	20	10	5
> 75%	-	-	-	-	-	81
> 70%	-	-	-	72	75	76
> 65%	-	-	69	69	68	71

TABLE III
WEIGHTED MODAL CLASSIFICATION ACCURACIES (%) ON TEST DATA, USING
PATCHES OF SIZE $2 \times 3 \times 2$ VOXELS. ONLY A PERCENTAGE OF THE PATCHES
WHICH CLASSIFY TRAINING DATA MOST ACCURATELY ARE INCLUDED

Percentage of patches used (%)	5	2	1
HML	81	84	84
IND	78	81	83

enhance the prediction power of the patches. In the final classification we have used a majority vote of patches deemed to be relevant to classification, weighted by how well those patches classify training data. A more interesting approach would be to combine the predictions of patches by weighting them based on how well they classify badly-classified data, with similar motivation to that of boosting strategies [55].

C. Computational Aspects

To demonstrate how our algorithm can be applied to large sets of 3-D data, we have applied it to a population of 429 3-D brain images of size $160 \times 192 \times 160$ voxels from the ADNI cohort. The greatest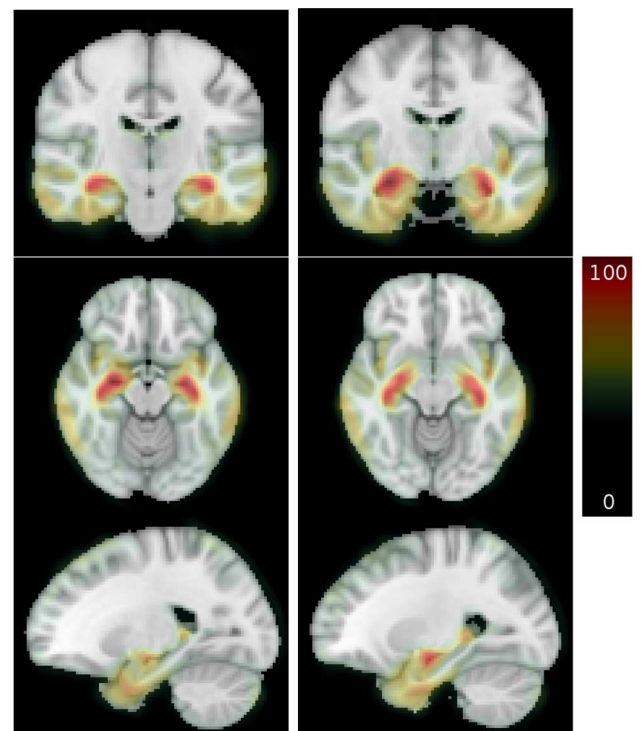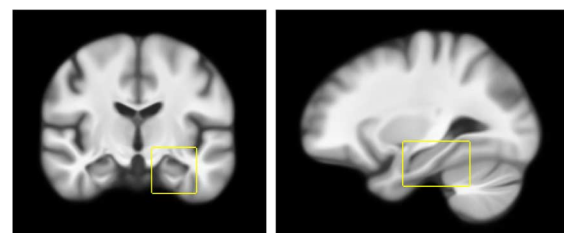 time expense in this pipeline was found to be in the SVM classification of each patch. While constructing the manifold representations is comparatively quick, the main computational issue is the large memory requirement. The size of manifolds to be stored at each level of the hierarchy is of the order of (number of patches $\times$ number of dimensions $\times$ number of images). To reduce these demands, it would be beneficial to reduce the number of patches which need to be calculated and stored. This could be done by using results from early levels of the hierarchy to preselect regions of interest. Alternatively, the decision to continue dividing the patches could be based on some image-based quality, such as the entropy of the images within that patch. For example, it would make little sense to continue the process if it is clear that all subsequent smaller patches only consisted of background.
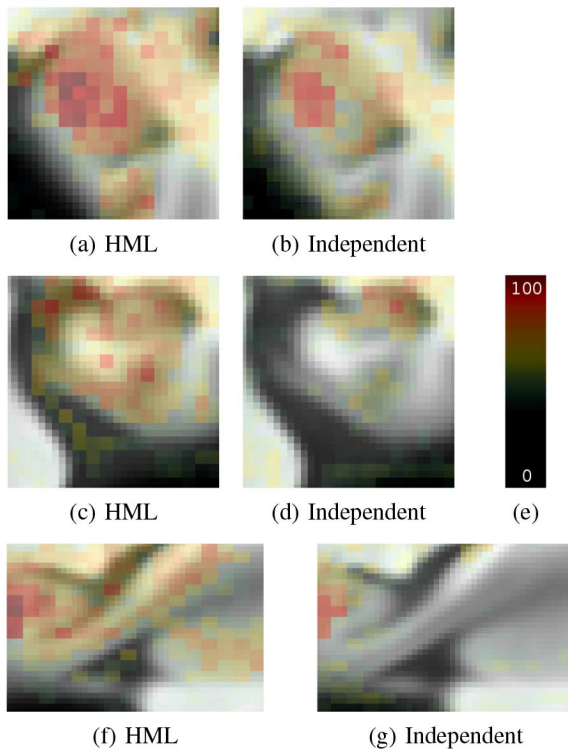
Fig. 17. Classification accuracy of patches of size $2 \times 3 \times 2$ voxels. Images (a), (c), and (f) show the results of using HML while (b), (d), and (g) show the results on the same section of anatomy using independent patch embeddings. It can be seen that HML detects more regions associated with Alzheimer's disease.

VI. Conclusion

We have proposed an algorithm for HML—an extension to current manifold learning techniques enabling regional image analysis. We have shown its flexibility in applying it in two different scenarios: 1) the regional analysis of the respiratory and cardiac cycles of an individual; 2) the automatic determination of discriminative areas in the classification of Alzheimer's disease. We believe HML may be useful in learning local variations of datasets, where regions of interest are not known *a priori*. Our algorithm can be applied on large sets of 3-D images to obtain embeddings at fine scales.

Appendix

Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a $60 million, five-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55–90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

References

[1] C. Wachinger, M. Yigitsov, and N. Navab, "Manifold learning for image-based breathing gating with application to 4d ultrasound," in *Proc. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2010, pp. (II)26–33.

[2] Q. Zhang, R. Souvenir, and R. Pless, "On manifold structure of cardiac MRI data: Application to segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. (1)1092–1098.

[3] P. Aljabar, R. Wolz, L. Srinivasan, S. J. Counsell, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "A combined manifold learning analysis of shape and appearance to characterize neonatal brain development," *IEEE Trans. Med. Imag.*, vol. 30, no. 12, pp. 2072–2086, Dec. 2011.

[4] S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, and R. Whitaker, A. D. N. I., "Manifold modeling for brain population analysis," *Med. Image Anal.*, vol. 14, no. 5, pp. 643–653, 2010.

[5] J. Hamm, C. Davatzikos, and R. Verma, "Efficient large deformation registration via geodesics on a learned manifold of images," in *Proc. 12th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2009, pp. 680–687.

[6] R. Sparks and A. Madabhushi, "Novel morphometric based classification via diffeomorphic based shape representation using manifold learning," in *Proc. Conf. Medical Image Comput. Comput.-Assist. Intervent.*, 2010, pp. 658–655.

[7] R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D. P. Zhang, D. Rueckert, J. Soininen, and J. Lotjonen, "Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease," *PLoS One*, vol. 6, no. 10, p. e25446, 2011.

[8] N. K. Batmanghelich, B. Taskar, and C. Davatzikos, "Generative-discriminative basis learning for medical imaging," *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 51–69, Jan. 2012.

[9] M. Liu, D. Zhang, and D. Shen, A. D. N. I., "Ensemble sparse classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 2, pp. 1106–1116, 2012.

[10] A. Rao, Y. Lee, A. Gass, and A. Monsch, "Classification of Alzheimer's disease from structural MRI using sparse logistic regression with optional spatial regularisation," in *Proc. Eng. Med. Biol. Soc.*, 2001, pp. 4499–4502.

[11] M. Sabuncu and K. van Leemput, "The relevance voxel machine (RVOXM): A self-tuning bayesian model for informative image-based prediction," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2290–2306, Dec. 2012.

[12] G. Varoquaux, A. Gramfort, B. Thirion, R. Jenatton, G. Obozinski, and F. Bach, "Sparse structured dictionary learning for brain resting-state activity modeling," in *Proc. NIPS Workshop Practical Appl. Sparse Model.: Open Issues New Directions*, 2010, pp. 552–565.

[13] M. Vounou, E. Janousova, R. Wolz, J. L. Stein, P. M. Thompson, D. Rueckert, and G. Montana, "Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease," *NeuroImage*, vol. 60, pp. 700–716, 2012.

[14] W. Sun, M. Cetin, R. Chan, V. Reddy, G. Holmvang, V. Chandar, and A. Willsky, "Segmenting and tracking the left ventricle by learning the dynamics in cardiac images," in *Proc. 19th Int. Conf. Inf. Process. Med. Imag.*, 2005, pp. 552–565.

[15] C. Petitjean and J. N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Med. Image Anal.*, vol. 15, no. 2, pp. 169–184, Apr. 2011.

[16] J. Milles, R. J. van der Geest, M. Jerosch-Herold, J. H. Reiber, and B. P. Lelieveldt, "Fully automated motion correction in first-pass myocardial perfusion MR image sequences," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1611–1621, Nov. 2008.

[17] L. van der Maaten, E. Postma, and J. van der Herik, Dimensionality reduction: A comparative review TiCC, Tilburg Univ., Tilburg, The Netherlands, Tech. Rep., 2007.

[18] D. H. Ye, H. Litt, C. Davatzikos, and K. M. Pohl, "Morphological classification: Application to cardiac MRI of tetralogy of fallot," in *Proc. Funct. Imag. Model. Heart*, 2011, pp. 180–187.

[19] L. Cayton, Algorithms for manifold learning UCSD, La Jolla, CA, Tech. Rep., 2005.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[21] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computat.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[22] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, p. 2319, 2000.

[23] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. (2)988–995.

[24] A. Schumacher, "Linear and nonlinear approaches to the analysis of R-R interval variability," *Biol. Res. Nursing*, vol. 5, no. 3, pp. 211–221, 2004.

[25] S. C. Malpas, "Neural influences on cardiovascular variability: Possibilities and pitfalls," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 282, no. 1, pp. H6–H20, 2002.

[26] N. Duchateau, M. D. Craene, G. Piella, and A. F. Frangi, "Constrained manifold learning for the characterization of pathological deviations from normality," *Med. Image Anal.*, vol. 16, no. 8, pp. 1532–1549, 2012.

[27] M. Georg and R. Souvenir, "Manifold learning for 4d CT reconstruction of the lung," in *IEEE Workshop Math. Methods Biomed. Image Anal.*, Anchorage, AK, 2008, pp. 1–8.

[28] G. Isguder, G. Unal, M. Groher, N. Navab, A. K. Kalkan, M. Degertekin, H. Hetterich, and J. Rieber, "Manifold learning for image-based gating of intravascular ultrasound (IVUS) pullback sequences," in *Proc. 5th Int. Workshop Med. Imag. Augment. Reality*, 2010, pp. 139–148.

[29] D. Wasserman and R. Deriche, "Simultaneous manifold learning and clustering: Grouping white matter fiber tracts using a volumetric white matter atlas," in *MIDAS J.—Manifolds Med. Imag.: Metrics—Learn. Beyond, MICCAI 2008 Workshop*, 2008 [Online]. Available: http://www.midasjournal.org/browse/publication/636

[30] D. H. Ye, J. Hamm, D. Kwon, C. Davatzikos, and K. M. Pohl, "Regional manifold learning for deformable registration of brain MR images," in *Proc. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2012, pp. (3)131–-138.

[31] M. J. Cardoso, R. Wolz, M. Modat, N. C. Fox, D. Rueckert, and S. Ourselin, "Geodesic information flows," in *Prod. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2012, pp. 262–270.

[32] K. K. Bhatia, A. N. Price, J. V. Hajnal, and D. Rueckert, "Localised manifold learning for cardiac image analysis," in *Proc. SPIE Med. Imag. 2012: Image Process.*, San Diego, CA, Feb. 2012, p. 83140H.

[33] J. J. Corso, E. Sharon, and A. L. Yuille, "Multilevel segmentation and integrated Bayesian model classification with an application to brain tumor segmentation," in *Proc. 9th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2006, pp. 790–798.

[34] J. J. Corso, E. Sharon, S. El-Saden, U. Sinha, and A. L. Yuille, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 629–640, May 2008.

[35] D. Kushnir, M. Galun, and A. Brandt, "Fast multiscale clustering and manifold identification," *Pattern Recognit.*, vol. 39, no. 10, pp. 1876–1891, 2006.

[36] M. Maggioni, A. Szlam, R. R. Coifman, and J. C. Bremer, "Diffusion-driven multiscale analysis on manifolds and graphs: Top-down and bottom-up constructions," in *Proc. SPIE Conf.*, Aug. 2005, vol. 5914, pp. 445–455.

[37] M. Liu, L. Lu, X. Ye, S. Yu, and H. Huang, "Coarse-to-fine classification via parametric and nonparametric models for computer-aided diagnosis," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Management*, New York, NY, 2011, pp. 2509–2512.

[38] K. K. Bhatia, A. Rao, A. N. Price, R. Wolz, J. V. Hajnal, and D. Rueckert, "Hierarchical manifold learning," in *Proc. MICCAI*, 2012, pp. 512–519.

[39] K. C. Bilchick, V. Dimaano, K. C. Wu, R. H. Helm, R. G. Weiss, J. A. Lima, R. D. Berger, G. F. Tomaselli, D. A. Bluemke, H. R. Halperin, T. Abraham, D. A. Kass, and A. C. Lardo, "Cardiac magnetic resonance assessment of dyssynchrony and myocardial scar predicts function class improvement following cardiac resynchronization therapy," *JACC Cardiovasc. Imag.*, vol. 1, pp. 561–568, Sep. 2008.

[40] V. Delgado and J. J. Bax, "Assessment of systolic dyssynchrony for cardiac resynchronization therapy is clinically useful," *Circulation*, vol. 123, pp. 640–655, 2011.

[41] F. Savill, T. Schaeffter, and A. P. King, "Assessment of input signal position for cardiac respiratory motion models during different breathing patterns," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, Chicago, IL, Mar. 2011, pp. 1698–1701, 2011.

[42] Y. Qi, T. P. Minka, R. W. Picard, and Z. Ghahramani, "Predictive automatic relevance determination by expectation propagation," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 671–678.

[43] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc.: Ser. B*, vol. 67, no. 2, pp. 301–320, 2005.

[44] R. Souvenir and R. Pless, "Image distance functions for manifold learning," *Image Vis. Comput.*, vol. 25, no. 3, pp. 365–373, 2007.

[45] W. W. III, P. Viola, H. Astumi, and S. Nakajima, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[46] C. Studholme, D. L. G. Hill, and D. G. Hawkes, "An overlap invariant entropy measure of 3-D medical image alignment," *Pattern Recognit.*, vol. 32, pp. 71–86, 1999.

[47] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds," in *Proc. 10th Int. Workshop Artif. Intell. Stat.*, 2005, pp. 120–127.

[48] Alzheimer's Disease Neuroimaging Initiative [Online]. Available: http://www.loni.ucla.edu/ADNI

[49] D. Rueckert, L. I. Sononda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[50] V. N. Vapnik*, Statistical Learning Theory*.   New York: Wiley-Blackwell, 1998.

[51] A. Convit, J. de Asis, M. J. de Leon, C. Y. Tarshish, S. D. Santi, and H. Rusinek, "Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to alzheimer's diesase," *Neurobiol. Aging*, vol. 21, no. 1, pp. 19–26, 2000.

[52] J. A. Kaye, T. Swihard, D. Howeison, A. Dame, M. M. Moore, T. Karnos, R. Camicolo, M. Ball, B. Oken, and G. Sexton, "Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia," *Neurology*, vol. 48, no. 5, pp. 1297–1304, 1997.

[53] Statistical Parametric Mapping Ver. 8 [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/software/spm8

[54] J. Ashburner and K. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.

[55] Y. Freund and R. E. Schapire, "Game theory, on-line prediction and boosting," in *Proc. 9th Annu. Conf. Computat. Learn. Theory*, New York, 1996, pp. 325–332.