

Machine Learning in Computer–Aided Diagnosis: Medical Imaging Intelligence and Analysis

Kenji Suzuki
University of Chicago, USA

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Myla Harty
Acquisitions Editor: Erika Gallagher
Typesetters: Jennifer Romanchak
Print Coordinator: Jamie Snavelly
Cover Design: Nick Newcomer, Greg Snader

Published in the United States of America by
Medical Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

Copyright © 2012 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Machine learning in computer-aided diagnosis: medical imaging intelligence and analysis / Kenji Suzuki, editor.
p. cm.

Includes bibliographical references and index.

Summary: "This book provides a comprehensive overview of machine learning research and technology in medical decision-making based on medical images"--Provided by publisher.

ISBN 978-1-4666-0059-1 (hardcover) -- ISBN 978-1-4666-0060-7 (ebook) -- ISBN 978-1-4666-0061-4 (print & perpetual access) 1. Diagnostic imaging--Digital techniques. 2. Image analysis. 3. Machine learning. I. Suzuki, Kenji, 1968 Oct. 5-RC78.7.D53M33 2011
616.07'54--dc23

2011037853

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 18

Learning Manifolds: Design Analysis for Medical Applications

Diana Mateus

Technische Universität München, Germany & Helmholtz Zentrum München, Germany

Christian Wachinger

Technische Universität München, Germany

Selen Atasoy

Technische Universität München, Germany & Imperial College London, UK

Loren Schwarz

Technische Universität München, Germany

Nassir Navab

Technische Universität München, Germany

ABSTRACT

Computer aided diagnosis is often confronted with processing and analyzing high dimensional data. One alternative to deal with such data is dimensionality reduction. This chapter focuses on manifold learning methods to create low dimensional data representations adapted to a given application. From pairwise non-linear relations between neighboring data-points, manifold learning algorithms first approximate the low dimensional manifold where data lives with a graph; then, they find a non-linear map to embed this graph into a low dimensional space. Since the explicit pairwise relations and the neighborhood system can be designed according to the application, manifold learning methods are very flexible and allow easy incorporation of domain knowledge. The authors describe different assumptions and design elements that are crucial to building successful low dimensional data representations with manifold learning for a variety of applications. In particular, they discuss examples for visualization, clustering, classification, registration, and human-motion modeling.

INTRODUCTION

Computer-aided diagnosis often implies processing large and high dimensional datasets, for instance, high-resolution volumes containing millions of voxels, or 4D videos collecting motion information over time. Visualization and analysis of such data can be very time demanding for physicians but also very computationally expensive for machines assisting diagnosis tasks. Fortunately, in many cases the relevant information for an application can be represented in lower dimensional spaces. If appropriately chosen and designed, dimensionality reduction methods will not only decrease the processing time but also facilitate any posterior analysis. Therefore, they can be of great use to a variety of CAD (Computer Aided Diagnosis) applications, ranging from general problems such as classification and visualization, to more specific ones like multi-modal registration or motion compensation. Up to recent years, dimensionality reduction in CAD has relied mainly on *linear* methods, such as Principal Component Analysis (PCA). Linear methods are however not suitable for handling non-linear complex relationships among the data samples. Non-linear approaches based on *manifold learning* are a good alternative for dimensionality reduction in such cases (Lin & Zha, 2008; Pless & Souvenir, 2009).

Established manifold learning methods like Isomap (Tenenbaum, et al., 2000), Locally Linear Embedding (LLE) (Roweis & Saul, 2000) or Laplacian Eigenmaps (Belkin & Nigoyi, 2003) are widely used in different scientific communities for data representation, dimensionality reduction, visualization, and clustering. The name “manifold” learning comes from the assumption that data-points represented in a high dimensional space lie on a low dimensional manifold; it is this manifold that the different algorithms try to approximate and represent. Several properties make manifold learning approaches very attractive, for example, flexibility, simplicity, their capability to account for non-linear data relations and their

closed form solution. The flexibility is a result of representing the data points not by their coordinates in the high dimensional space, but instead by means of relational functions between pairs of data points. These pairwise relations are determined in terms of customized similarity measures and can be non-linear functions. The core of manifold learning algorithms is independent of these measures, handing to the designer the responsibility of determining the right similarities to capture the appropriate manifold structure. Properties of manifold learning such as the flexibility and non-linearity are relevant to computer aided diagnosis, given that medical datasets tend to be high dimensional and often represent complex non-linear phenomena. Moreover, many medical datasets often verify the assumption that the data lies close to a manifold structure. For instance, the contiguous frames of a video or the slices of a volume vary smoothly; also, the continuous deformation of an organ’s shape over time can be considered to form a manifold; finally, the variations of an organ over a population can also be expected to lie on a manifold. These facts have recently raised interest in using manifold learning methods for a variety of applications, including visualization, clustering, classification, statistical shape analysis, registration, and segmentation.

A direct application of manifold learning is *visualization*. (Lim, et al., 2003) presented an early work applying manifold learning for visualizing biomedical data, where differences between bone structures were displayed in 2D by means of Isomap. Several gene expression studies (Nilsson, et al., 2004; Bartenhagen, et al., 2010) also rely on manifold learning for visualization of microarray data. Visualization of cardiac Magnetic Resonance (MR) images using Isomap was explored in (Souvenir and Pless, 2007), where images in the same breathing phase were displayed as nearby points in the low dimensional space.

Using appropriate non-linear similarity functions and neighborhood systems, manifold learning can be designed to map complex cluster

patterns in the high dimensional space to simpler representations in a low dimensional space. Clustering in the reduced space is more efficient and usually performed with simple algorithms such as K -means. This property has also been explored for *clustering* biomedical data (Finn, et al., 2009; Atasoy, et al., 2010). (Finn, et al., 2009) show that a better visualization of high dimensional flow cytometry data is achieved when using a statistical manifold learning method, instead of the conventional linear projections. Phenotype patterns of leukemia that are challenging to distinguish by experts are easily clustered in the computed low dimensional embedding space. Another example of manifold learning applied to clustering was presented in (Atasoy, et al., 2010), where images of an endoscopic video, which reside in a high dimensional space, can be visualized and analyzed in a dimensionality-reduced space after learning the manifold. Such a representation is suitable for faster navigation and also for clustering similar scenes or similar imaging conditions.

The training of classifiers in the reduced space can also take advantage of the clustering property of manifold learning, as it is expected that non-linearly separable classes are better separated in the low dimensional representation. Therefore, manifold learning has also been used for *classification* of biomedical data (Wachinger, et al., 2010a; Lee, et al., 2008; Suzuki et al., 2010; Sparks & Madabhushi, 2010). (Wachinger, et al., 2010a) model the automatic patient position detection in MRI scanners as a classification problem. Laplacian Eigenmaps on low-resolution slices and a nearest-neighbors classifier are used to classify the position the new slices into body-parts. (Lee, et al., 2008) compare different linear and non-linear methods for dimensionality reduction applied to the classification of gene and protein expression. Cluster quality measures and the discriminant power of two supervised classifiers, namely Support Vector Machines (SVM) and decision trees, are used for the evaluation. (Suzuki, et al., 2010) address the detection of suspicious polyps in a

Computed Tomography Colonography (CTC) with a method that combines a non-linear dimensionality reduction method and Massive-Training Artificial Neural Networks (MTANN). The reduction method is inspired by the Laplacian eigenfunctions (Belkin & Nigogiyi, 2003) and relates to the Locality Preserving Projections algorithm (He, et al., 2005). The input to the method is the set of pairwise relations between sub-volumes of the CTC. The resultant low dimensional representation improves the time-consuming training stage of the MTANN. Finally, (Sparks & Madabhushi, 2010) reduce the dimensionality of morphological features by computing diffeomorphic similarities and using Laplacian Eigenmaps. The low dimensional representations are used in conjunction with a Support Vector Machine (SVM) to classify different grades of prostate cancer, and to differentiate between benign vs. malignant lesions of breast in DCE-MRI images.

More sophisticated classification schemes have also been explored. For instance, (Tiwari, et al., 2010) combine the power of multi-kernel learning with graph embedding (manifold learning) techniques to create a common representation that fuses multi-modal information from MRI and Spectroscopy. The method uses additionally partial supervision of labels (thus is a semi-supervised technique) to propagate labels and to discriminate benign from cancerous tissue and different grades of aggressive prostate cancer. For the same application, (Tiwari, et al., 2009) had previously investigated the use of spectral embedding in combination with probabilistic boosting trees. Also in (Tiwari, et al., 2008) a modified version of LLE to classify only the MRI images was proposed.

Manifold learning methods have also been used for *population studies* and *statistical analysis* of the brain under various imaging modalities (Verma, et al., 2007; Gerber, et al., 2010; Aljabar, et al., 2008; Aljabar, et al., 2010). (Verma, et al., 2007), rely on Isomap for creating a low dimensional representation of Diffusion Tensor MRI (DT-MRI)

of the brain. Diffusion tensors call for non-linear dimensional reduction methods, as they inherently lie on a non-linear sub-manifold of the \mathbf{R}^6 space, and given that plausible images are restricted by the structure of the brain. Isomap captures the manifold structure of the tensor images and embeds it into a linear space where it is easier to compute statistical measures and perform statistical tests. Following a similar reasoning, (Gerber, et al., 2010) argue that the space spanned by a set of MRI brain images can be approximated by a low-dimensional non-linear manifold. Relying on this assumption, Isomap is applied on brain databases containing T1-weighted MRI images acquired from Alzheimer patients and controls. An approximation to the diffeomorphic metric for small deformations is used for the pairwise similarities. The recovered manifold shows correlations to clinical parameters such as age, mini-mental-state examination and clinical dementia rating. The low dimensional representation is combined with regression to create mappings relating the reduced to the original space. Such a generative model is useful for retrieving brains with similar shapes and for creating priors and atlases. A related work (Aljabar, et al., 2008), addresses the diagnosis of dementia of brains according to morphological analysis of neuro-anatomical structures. Based on pairwise image similarities derived from the structural segmentation overlap, Laplacian Eigenmaps is used to discriminate clinical groups of elderly patients with Alzheimer's disease and controls. Later, (Aljabar, et al., 2010) investigate the use of multiple measures (of shape and appearance) to learn manifolds, and the combination of Laplacian Eigenmaps and Isomap into a single population representation. The method has been applied to neonatal brain MRI leading to a low dimensional representation that correlates well with clinical data and to morphological features. The representation can also be used to characterize the trajectories of brain development.

The analysis of organ deformation is an important medical application of manifold learning.

Images of organs under deformations produced by breathing, the heart beating, or similar deformations are clear examples of low dimensional data residing in high dimensional spaces. Low dimensional representations that learn the manifold of deformations can be used for segmentation (Zhang, et al., 2006; Etyngier et al. 2007; Kadoury & Paragios, 2010), reconstruction (Georg, et al., 2008, Wachinger, et al., 2010d) and registration (Hamm et al, 2010, Wolz, et al., 2010).

An application of learning deformation manifolds for *segmentation* was presented in (Zhang, et al., 2006), addressing the analysis of cardiopulmonary MR images for left-ventricle segmentation. Isomap is used to characterize the manifold of shapes of ungated, free-breathing cardiac MRI data. The method incorporates domain-specific constraints from the learnt manifold in a level-set segmentation algorithm, which behave better than the usual smoothness constraints. To learn the manifold, the pairwise similarities are computed from the phase difference of local Gabor filters. Resultant low dimensional coordinates of the manifold can be identified with changes in the breathing and the heartbeat phases. With a similar approach, (Etyngier et al. 2007) proposed to use manifold learning to create non-linear deformable shape priors of ventricle nuclei images. Authors employ the diffusion maps method (Lafon & Lee, 2006; Nadler, et al., 2006) to create a low dimensional representation, which serves as a non-linear shape prior attracting the shape to the manifold within a variational segmentation approach. The Nyström extension (Bengio, et al., 2003) is used to map new shapes to the manifold. In a recent work, (Kadoury & Paragios, 2010) address the segmentation of spinal column from CT data based on learning the manifold of shape variations of the spine. A metric is introduced that handles both small and large deformations. The dimensionality-reduced representation is used as a prior within a Markov Random Field (MRF) inference framework that balances the prior distribution with image data. Finally, in

(Wolz, et al., 2010), manifold learning is used for segmentation based on label propagation. A small number of manually labeled MRI brain atlases are progressively propagated to a large dataset of unlabeled images, starting from the atlas images to their neighbors until the entire dataset is segmented. The low dimensional representation is used to identify neighboring images in each propagation step, such that instead of registering very different images, a succession of local deformations between similar images is used for the propagation.

Manifold learning can be useful for *reconstruction* of deforming organs from ungated images. (Wachinger, et al., 2010d) propose an image-based gating using Laplacian Eigenmaps. The low dimensional representation allows for identification of similar images, which are then used for reconstructing 4D ultrasound data from abdominal images affected by breathing motion. (Georg, et al., 2008) use manifold learning for building 4D reconstructions of the lung from CT images affected by cardiac and breathing motion. The low dimensional representation created with Isomap is used to determine an initial estimate of the pose of the current image sample in the space of deformation parameters. These parameters are related to the breathing and heartbeat phases. Pose estimation is then integrated within an optimization framework that iteratively reconstructs a static volume while updating the estimate of the poses in the low dimensional space.

In the case of *registration*, manifolds built based on distances computed from the deformations between two instances of a shape were used in (Hamm et al., 2010). The manifold can then be employed for registering pairs of shapes with large deformations by finding the path on the manifold and composing a global deformation from this path. This idea is similar to the one of (Wolz, et al., 2010) for segmentation. Finally, (Wachinger & Navab, 2010c) propose a different way of applying manifold learning to the multi-modal registration problem. This work relies on the principle that two

images of the same organ acquired with different modalities have almost identical intrinsic (self) similarities. Structural representations capturing these self-similarities are computed using Laplacian Eigenmaps independently on each image. The resultant representations can be then registered with mono-modal methods.

A last application considered here is the analysis of human motion (Schwarz, et al., 2009; Schwarz, et al., 2010) applied to diagnosis of neurological diseases. The method relies on manifold learning to create a series of low dimensional representations of activities of interest from motion capture data. Patients can then be equipped with a reduced number of portable inertial sensors that enable long-term analysis. During the test stage, the learnt motion models are used to detect the current activity and give an estimate of the pose (in terms of joint angles) only from the inertial sensor observations. As in (Kadoury & Paragios, 2010), the low dimensional representations are used to constrain the search space in the optimization. Also, similar to (Gerber, et al., 2010), mappings are learnt that relate the low and the high dimensional spaces to complete a generative model.

Chapter Organization: In the first part of the chapter we discuss the basic principles of manifold learning and recall the Laplacian Eigenmaps algorithm (Belkin & Nigoyi, 2003). A second part presents case studies from recent state of the art methods, where manifold learning is used for:

- Patient position detection by classification of full-body MRI slices (Wachinger, et al., 2010a).
- Visualization and clustering of endoscopic videos (Atasoy, et al., 2010).
- Multi-modal registration (Wachinger & Navab, 2010b).
- Image-based gating for 4D ultrasound reconstruction (Wachinger, et al., 2010d).
- Human Motion analysis (Schwarz, et al., 2010).

The parallel of these applications illustrates how to design the key elements of Laplacian Eigenmaps, namely the similarity measures and the neighborhood system, according to the application, and how to include domain knowledge in the dimensionality reduction process.

BACKGROUND: MANIFOLD LEARNING

Consider a dataset of N points $X = \{x_1, x_2, \dots, x_N\}$, with $\mathbf{x}_i \in \mathbb{R}^D \forall i \in \{1, \dots, N\}$, the goal dimensionality reduction is to find the best low dimensional representation $Y = \{y_1, y_2, \dots, y_N\}$, where $\mathbf{y}_i \in \mathbb{R}^d$ and $d \ll D$, that preserves the relevant information of X . The most common method to reduce the dimensionality is by means of Principal Component Analysis (PCA). After centering the data and stacking the points in a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$, PCA recovers a linear transformation \mathbf{T} that projects each point \mathbf{x}_i to a new coordinate system of smaller dimension, that is, $\mathbf{T} : \mathbf{x}_i \in \mathbb{R}^D \rightarrow \mathbf{y}_i \in \mathbb{R}^d$. The transformation is found from the eigenvectors of the covariance matrix $\mathbf{X}^T\mathbf{X}$. PCA is a very general method as the only assumptions made are that \mathbf{T} is a linear transformation, and that the low dimensional representation should preserve as much as possible the data variance. If it is more important to preserve the distances between points than their variance, then the also linear Multi-Dimensional Scaling (MDS) approach (Cox & Cox, 2001) reveals to be more appropriate. In MDS, the transformation is explicitly found by optimizing the distance preservation criteria. Notice that if the distances considered in MDS are Euclidean, the two approaches are equivalent.

Dimensionality reduction methods based on linear transformations are easy to compute and have the advantage of providing the transforma-

tion to project new incoming data-points into the reduced space. However, they are not capable of capturing complex non-linear relations among the data points. An academic example illustrating this problem is the Swiss-roll dataset (Roweis & Saul, 2000), a 2D manifold embedded in a 3D space. Linear methods fail to unfold the manifold. The Swiss-roll example can be related to image datasets. Individually each image is defined as a point in a high dimensional space (the dimension corresponding to the size of the image), however the set of images of interest (e.g. showing the changes of an organ under deformation due to breathing) lies in a sub-manifold of the image space and can be represented in a significantly lower dimensional space (e.g. a two dimensional space whose coordinates relate to the breathing motion). In such cases manifold learning approaches provide an alternative to approximate the manifold where data lives and “unfold” it to a low-dimensional space. Although other non-linear approaches (Sammon, 1969; Lawrence, 2005) exist, manifold learning methods are specifically designed for cases where the data follows the structure of a manifold. This *manifold assumption* states that data points are sampled from a low dimensional manifold living in a high dimensional space. This implies first, that the data should be indeed low dimensional, and second, that the data points change smoothly and continuously in the low dimensional space where they live, describing a metric that is locally linear.

In practice, the actual metric on the manifold is unknown and only a discrete set of data samples is available, so the actual manifold can only be approximated. Instead of considering data points individually, manifold learning tries to approximate the manifold locally by taking into account pairwise relations between the data points. The methods leave the design of these pairwise relations open to the user, such that they capture the relevant information for an application. When the manifold assumption is verified and a sufficient number of regularly sampled data points is available, manifold learning methods are

able to approximate the manifold and compute a meaningful low dimensional representation of the data samples.

A large variety of manifold learning methods exist. Examples are Isomap (Tenenbaum, et al., 2000), Locally Linear Embedding (Roweis & Saul, 2000), Laplacian Eigenmaps (Belkin & Nigoyi, 2003), Diffusion maps (Lafon & Lee, 2006; Nadler, et al., 2006) and Hessian Eigenmaps (Donoho & Grimes, 2003). A link of some of these algorithms to kernel methods, and in particular to Kernel PCA (Schölkopf, 1998), was established (Bengio, et al., 2004). Other methods and relations between them may be found in recent surveys (Lin & Zha, 2008, Pless & Souvenir, 2009; van der Maaten, et al., 2009) as well as in the further readings section.

The input to manifold learning methods is the dataset of N points $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ considered to be samples of a manifold M living in a space of high dimensionality \mathbf{R}^D . The goal is to find a non-linear function φ that maps the data samples x to a lower dimensional space ($d \ll D$):

$$\varphi : \mathbf{x} \in M \subset \mathbf{R}^D \rightarrow \mathbf{y} \in \mathbf{R}^d.$$

The problem of finding φ is highly unconstrained, so most methods do not explicitly find φ . Instead, only the low-dimensional coordinates of the mapped points $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ are recovered. The main differentiating feature among the aforementioned methods is the way to derive and constrain the mapping function.

In the following, we recall the Laplacian Eigenmaps method, introduced by Belkin and Nigoyi (2003), as an exemplary approach with an interesting mathematical derivation. Subsequently, we explore the capabilities of Laplacian Eigenmaps for different applications.

Laplacian Eigenmaps

To find the low dimensional representation of points $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, Laplacian Eigenmaps minimizes a cost function that enforces the preservation of the point's neighborhood structure, imposing data points that are neighbors in the original space to remain close after the mapping. The optimization is stated as finding the corresponding coordinates of the points $\mathbf{y}_i \in Y$ in a d -dimensional space:

$$\arg \min_{\mathbf{y}_i \in \mathbf{R}^d} \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \quad \forall i, j \in \{1, \dots, N\},$$

where $\mathbf{x}_i \sim \mathbf{x}_j$ denotes \mathbf{x}_i and \mathbf{x}_j are neighbors, $\|\cdot\|$ is the L2 norm, and w_{ij} are user defined weights usually giving higher penalties to pairs of points that are closer. To find a close form solution, the cost is reformulated in a matrix form that leads to an eigenvalue problem. First, the sum of squares is split as seen in Exhibit 1.

Let us define d_{ii} as the sum of the weights from a point \mathbf{x}_i to its neighbors $d_{ii} = \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij}$, and consider $w_{ij} = 0$ for non-neighboring points. Then, the first term is equivalent to Exhibit 2, and the cost can be rewritten as Exhibit 3.

Exhibit 1.

$$\sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i\|^2 + \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_j\|^2 - 2 \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \mathbf{y}_i^T \mathbf{y}_j.$$

Exhibit 2.

$$\sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i\|^2 = \sum_{\mathbf{x}_i} \sum_{\mathbf{x}_j} w_{ij} \|\mathbf{y}_i\|^2 = \sum_{\mathbf{x}_i} \|\mathbf{y}_i\|^2 \sum_{\mathbf{x}_j} w_{ij} = \sum_{\mathbf{x}_i} \|\mathbf{y}_i\|^2 d_{ii},$$

Exhibit 3.

$$\sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \sum_{\mathbf{x}_i} d_{ii} \|\mathbf{y}_i\|^2 + \sum_{\mathbf{x}_j} d_{jj} \|\mathbf{y}_j\|^2 - 2 \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \mathbf{y}_i^\top \mathbf{y}_j.$$

Let \mathbf{D} be a $N \times N$ diagonal matrix with entries d_{ii} and \mathbf{W} the $N \times N$ symmetric matrix containing weights w_{ij} . Then, define \mathbf{Y} as the $N \times d$ matrix containing in each row the sought low dimensional coordinates $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^\top$ and denote each dimension of \mathbf{Y} , with a superscripted $\mathbf{y}^{(l)}$, $l \in \{1, \dots, d\}$, such that matrix \mathbf{Y} can also be written as $\mathbf{Y} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(d)}]$. Using the relation¹ $\sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \mathbf{y}_i^\top \mathbf{y}_j = \text{trace}(\mathbf{Y}^\top \mathbf{W} \mathbf{Y})$, the cost function in matrix form is expressed as:

$$\sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = 2 \text{trace}(\mathbf{Y}^\top \mathbf{D} \mathbf{Y} - \mathbf{Y}^\top \mathbf{W} \mathbf{Y}).$$

Finally, let us define the positive semi-definite Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$. The optimization becomes:

$$\arg \min_{\mathbf{Y} \in \mathbf{R}^{N \times d}} \text{trace}(\mathbf{Y}^\top \mathbf{L} \mathbf{Y}) \quad \text{s.t.} \quad \mathbf{Y}^\top \mathbf{D} \mathbf{Y} = \mathbf{I} \quad \text{and} \quad \mathbf{y}^{(k)\top} \mathbf{D} \mathbf{1} = 0,$$

where the first constraint enforces the orthogonality between the different dimensions of the low dimensional representation and removes the arbitrary scaling factor, preventing the solution to collapse to a sub-space of lower dimension than d . The second constraint ensures the translation invariance. The sought low dimensional coordinates of the points \mathbf{Y} correspond to the first d

eigenvectors associated to the smallest non-zero eigenvalues, solution to the generalized eigenvalue problem $\mathbf{L} \mathbf{v} = \lambda \mathbf{D} \mathbf{v}$. Given \mathbf{L} is sparse this eigen-decomposition is computationally efficient.

In practice, the Laplacian Eigenmaps algorithm consists of the following steps:

1. Define a pairwise similarity measure between elements of \mathbf{X} (that determines the weights w_{ij}).
2. Define a neighborhood system ($\mathbf{x}_i \sim \mathbf{x}_j$).
3. Build matrix \mathbf{W} from steps 1 and 2, using $w_{ij} = 0$ for non-neighboring points.
4. Compute the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$.
5. Solve the generalized eigenvalue problem $\mathbf{L} \mathbf{v} = \lambda \mathbf{D} \mathbf{v}$ (i.e. find the eigenvectors $\mathbf{v}^{(l)}$ corresponding to the d smallest non-zero eigenvalues $\lambda^{(l)}$, with $l \in \{1, \dots, d\}$).
6. The l -th eigenvectors contains the l -th coordinates of the points in the low dimensional space, $\mathbf{y}^{(l)} = \mathbf{v}^{(l)}$, and $\mathbf{Y} = [\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(d)}]$. As a result, every point $\mathbf{x}_i \in \mathbf{X}$, $\mathbf{x}_i \in \mathbf{R}^D$, is mapped to a low dimensional representation $\mathbf{y}_i \in \mathbf{Y}$, $\mathbf{y}_i \in \mathbf{R}^d$, with

$$\mathbf{y}_i = [v_i^{(1)}, v_i^{(2)}, \dots, v_i^{(d)}]^\top \quad \text{and} \quad v_i^{(l)}$$

denoting the i -th element of the l -th eigenvector $\mathbf{v}^{(l)}$.

The most common way to define the weights is to use a Gaussian kernel based on dissimilarity between data points:

$$w_{ij} = \exp\left(\frac{-\text{diss}(\mathbf{x}_i, \mathbf{x}_j)}{\sigma}\right),$$

with σ a parameter. Values for σ can be estimated from the standard deviation of the pairwise dissimilarities $\text{diss}(\mathbf{x}_i, \mathbf{x}_j)$. For building the neighborhood, k -nearest neighbors (k -nn) or ε -ball neighborhood systems are common choices. In k -nn, a distance or dissimilarity function (e.g. $\text{diss}(\mathbf{x}_i, \mathbf{x}_j)$) is used from each point \mathbf{x}_i to all other points in the set X , the k points that are the closest to \mathbf{x}_i are then chosen to be \mathbf{x}_i 's neighbors. In the ε -ball system, all the points in X that lie within a distance (or dissimilarity) of ε to \mathbf{x}_i are its neighbors, e.g. $\mathbf{x}_i \sim \mathbf{x}_j$ if $\text{diss}(\mathbf{x}_i, \mathbf{x}_j) < \varepsilon$.

The Laplacian Eigenmaps method receives its name from graph theory, as it models the dimensionality reduction problem as the non-linear embedding of graph into a space of a low dimension. Consider a weighted graph with each vertex representing one of the data-samples $\mathbf{x}_i \in \mathbf{R}^D$, and weighted edges between neighboring nodes $\mathbf{x}_i \sim \mathbf{x}_j$. The matrix \mathbf{W} containing the weights w_{ij} defines the weighted adjacency matrix of the graph, while the d_{ii} values measure the connectivity of each vertex and give to matrix \mathbf{D} the name of degree matrix. Finally, the matrix \mathbf{L} plays the role of an operator on functions defined on the vertices of the graph, and can be thought of as the discrete equivalent of the continuous Laplace-Beltrami operator (Elmoataz, et al., 2008), from where the ‘‘Laplacian’’ term comes from. Notice the definition of $\mathbf{L}=\mathbf{D}-\mathbf{W}$ is only one possibility among other discretizations of the Laplace-Beltrami operator (Hein, et al., 2007).

From the graph embedding point of view, the graph created from the samples in X is an approximation of the manifold M . Thus, the success

of finding a meaningful low dimensional representation depends on the data points reasonably sampling the manifold, and on the definition of appropriate neighborhood systems and weights according to domain knowledge.

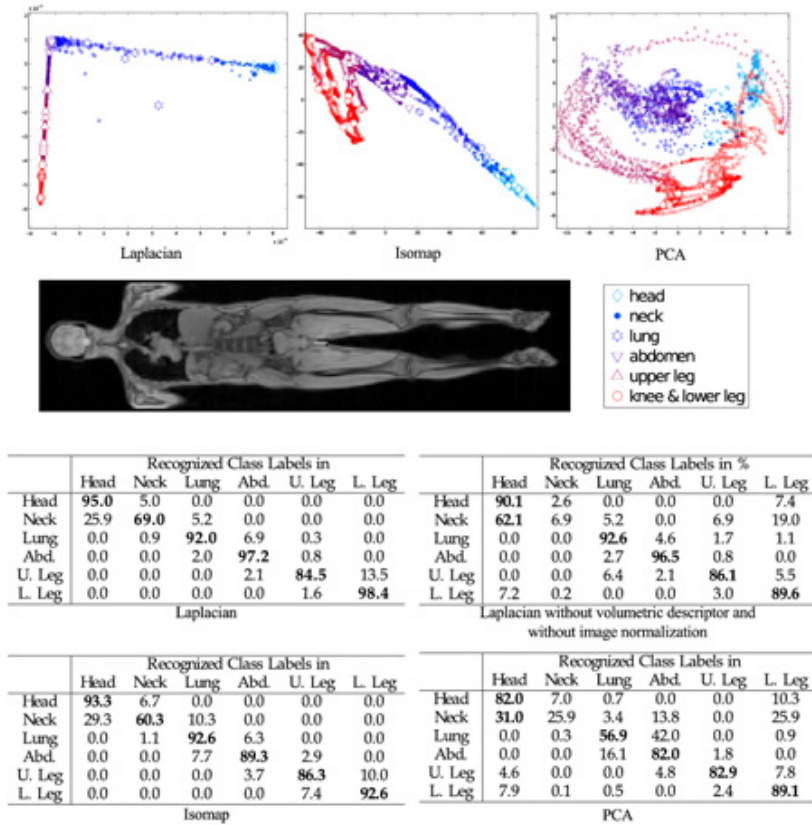
DESIGNING MANIFOLD LEARNING METHODS FOR MEDICAL APPLICATIONS

In this section we discuss a series of applications for which a solution involving manifold learning, and in particular the Laplacian Eigenmaps, has been proposed.

Patient Position Detection by Classification of Full-Body MRI Slices

Although Magnetic Resonance Imaging (MRI) does not require ionizing radiation, the applied radio frequency power produces heating. The amount of heating depends on the body part being imaged. (Wachinger, et al., 2010a) propose to detect the currently observed body part to allow better monitoring of the absorbed power as well as the optimization of the image acquisition. To preserve the MRI acquisition protocol, authors use a sequence of low-resolution images acquired during the initial placement of the patient in the scanner. During this acquisition, the bed moves with a relatively high but constant speed, which results in low-resolution slices ($\sim 64 \times 64$ pixels) with a slice spacing of 7.5 mm to 15 mm (see coronal a view in Figure 1). The detection of the patient position is modeled as the classification of the current slice into one of the specified regions of the body, e.g. head, abdomen, etc. The proposed solution uses Laplacian Eigenmaps during a training phase to create a low dimensional representation of the whole-body low-resolution slices. In the test phase, a simple classifier is deployed in the low dimensional representation to assign the

Figure 1. (Top) 2D representations obtained with Laplacian Eigenmaps, Isomap and PCA, using the normalization and the slice volumetric descriptors. Different markers show the ground-truth classification of the MRI slices. (Middle) A coronal view of the low resolution MRI slices. (Bottom) Quantitative Validation of the classification of MRI slices for patient position detection.



region of the body to a new acquired slice. Due to the controlled motion of the patient, the content of the slices varies smoothly and thus complies with the conditions for manifold learning.

Prior to classification, a low dimensional representation from N available low-resolution slices is built using Laplacian Eigenmaps. A graph is constructed with nodes representing the images $\{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_N\}$ and edges determined with a k -nn system ($k=40$). To compute the graph's edge weights, images are first normalized and converted to vectors. Additionally, a descriptor \mathbf{J}_i^T is defined for each slice \mathbf{I}_i collecting the $2s+1$ contiguous slices, $\mathbf{J}_i^T = [\mathbf{I}_{i-s}^T, \dots, \mathbf{I}_i^T, \dots, \mathbf{I}_{i+s}^T]$. The descriptor captures the volumetric context of \mathbf{I}_i . Finally, the weights are computed with a Gauss-

ian kernel based on the L2 norm between pairs of slice descriptors:

$$\text{diss}_{MRI}(\mathbf{J}_i, \mathbf{J}_j) = \|\mathbf{J}_i - \mathbf{J}_j\|^2.$$

The combination of the volumetric slice descriptors with the simple choices of weights and neighborhood system lead to low-dimensional representations that are appropriated for the classification of the slices.

The behavior of different (2D) representations obtained with Laplacian Eigenmaps, Isomap and PCA, using the normalization and the slice volumetric descriptors is illustrated in Figure 1, with different markers showing the ground-truth classification. Classification in the low dimensional

space produced by PCA is difficult. Isomap keeps the different classes grouped, but boundaries are not well defined. Laplacian Eigenmaps reflects most faithfully the continuous change from the head to the feet of the slices.

Quantitative classification results using a simple nearest-neighbor classifier also favor Laplacian Eigenmaps over PCA and Isomap, as tables in Figure 1 show. 13 whole body datasets were acquired. Manual labeling under supervision of a medical expert assigned each slice to one of 6 classes: head, neck, lung, abdomen, upper leg, and lower leg. A cross-validation with a leave-one-out strategy was performed.

From the tables in Figure 1 we see that dimensionality reduction with Laplacian Eigenmaps significantly facilitates classification leading to the highest overall correct classification rate of 94.0% (Compared to 93.3% for Isomap and 80.0% for PCA). Results also show that domain knowledge integrated in the form of image normalization and volumetric slices improves the classification results.

Visualization and Clustering of Endoscopic Videos

Cancer monitoring in Gastrointestinal Endoscopic (GI) videos involves long inspection of the video frames by an expert. Low dimensional representations of the endoscopic videos based on a manifold learning can facilitate the visualization of the videos and assist the expert. (Atasoy, et al., 2010) describe the adaptation of the Laplacian Eigenmaps for three tasks:

- Clustering of good and uninformative frames,
- Clustering of similar scenes,
- Video segmentation for summarization.

The approach followed in the three cases consists of first, creating an Endoscopic Video Manifold (EVM) by embedding the GI video in

a low dimensional space using Laplacian Eigenmaps; and second, deploying a K -means algorithm find clusters in the embedding space.

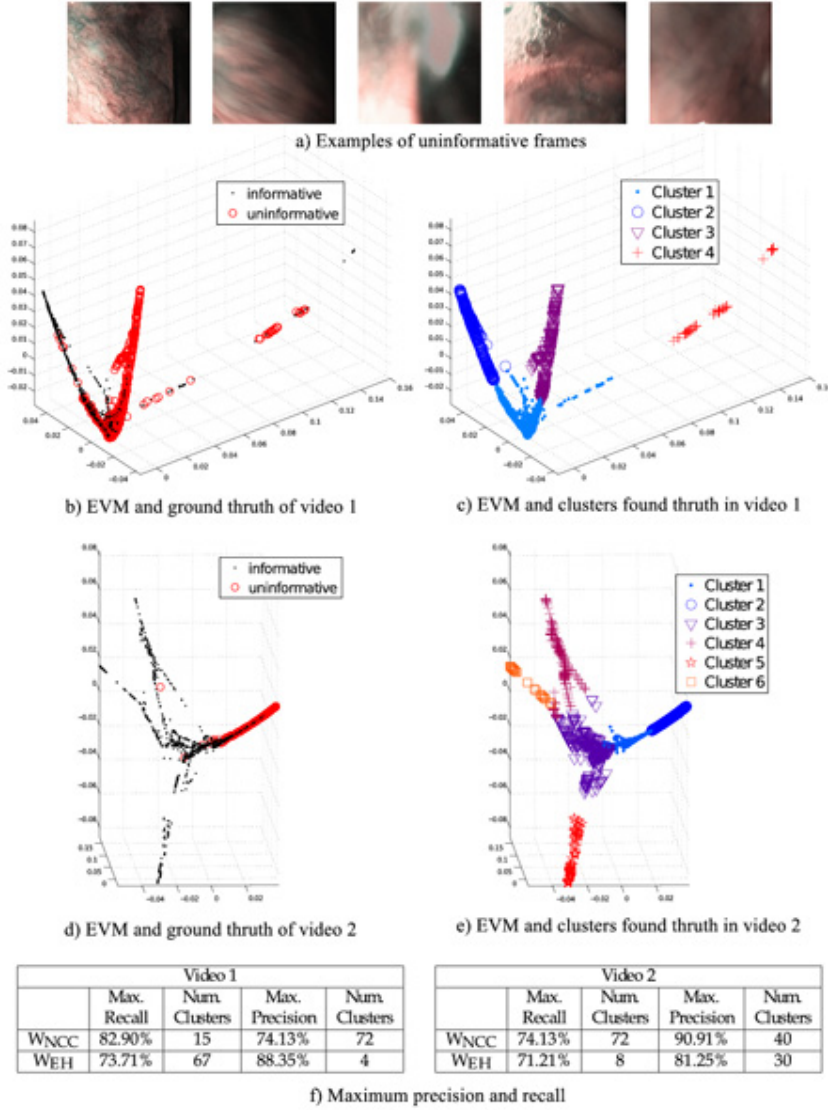
The first application of EVMs is the clustering of good and uninformative frames. Uninformative frames arise from motion or out-of-focus blur, specular highlights, and artifacts caused by turbid fluid in the GI. Example of such frames can be seen in Figure 2-a. As it is difficult to systematically detect these frames and it is not admissible to remove frames from the video, (Atasoy, et al., 2010) propose to cluster them. Because clustering is unsupervised, no prior labeling of example frames is required, while the particularities of each video (e.g. the patient, the optics, the modality) are automatically taken into account. To capture the difference between good and uninformative frames in the design of the Laplacian Eigenmaps graph, authors rely on the power spectrum of the images. Intuitively, good quality images are sharp and have large amounts of high-frequency content. In contrast, the content of blurred images and images with fluid remains usually in the low frequencies. Therefore, authors propose to use the difference between the power spectrum of pairs of images to define the graph weights.

Formally, consider a video is a set of N images $\{\mathbf{I}_1, \dots, \mathbf{I}_p, \dots, \mathbf{I}_N\}$, with $\mathbf{I}_i \in \mathbf{R}^D$. First, we compute the power spectrum of each frame \mathbf{I}_i using a discrete Fourier transform and log-polar coordinates. This results in a power spectrum $F_i(f, \theta)$, defined for a discrete set of frequencies $f \in \{f_1, f_2, \dots, f_{\max}\}$ and orientations $\theta = \{\theta_1, \theta_2, \dots, \theta_{\max}\}$. In order to make the image representation independent of its orientation authors compute a rotation-invariant power spectrum; first, summing the contributions of all orientations to each frequency, $F_i(f) = \sum_{\theta} F_i(f, \theta)$, and then, creating an Energy Histogram (EH) with B bins ($B=30$ in the experiments)

$\mathbf{h}_i = [h_i^1, h_i^2, \dots, h_i^B] = \text{hist}(F_i(f), B)$. Finally, the weights are computed with:

Learning Manifolds

Figure 2. Clustering of non-informative frames using the low dimensional representation of a Gastro-Intestinal (GI) endoscopic video. (a) Examples of non-informative frames. (b-e) 3D representations obtained with Laplacian Eigenmaps and the w_{EH} weights. (d) Quantitative evaluation of the clusters.



$$w_{ij} = w_{EH}(\mathbf{I}_i, \mathbf{I}_j) = \pi - \arccos \left(\frac{\langle \mathbf{h}_i, \mathbf{h}_j \rangle}{\|\mathbf{h}_i\| \cdot \|\mathbf{h}_j\|} \right).$$

Authors also investigate the use of an intensity-based similarity, the Normalized Cross-Correlation (NCC):

$$w_{ij} = w_{NCC}(\mathbf{I}_i, \mathbf{I}_j) = \text{ncc}(\mathbf{I}_i, \mathbf{I}_j).$$

Resultant low dimensional representations (EVMs) of the video using w_{EH} and a k -nn system are shown in Figure 2-(b,d) for two videos; the ground truth uninformative frames are overlaid. The clusters obtained with K -means on the EVMs are shown in Figure 2-(c,e). Note the separation of the uninformative frames in the EVMs and

the cluster correlation with the ground truth. For quantitative evaluation, clusters that contain more than 50% of uninformative frames are labeled as uninformative and K is varied from 1 to 80. Results of best precision and recall are shown in Figure 2-f.

(Atasoy, et al., 2010) also consider assisting endoscopic video examinations by clustering frames of the same scene, or by segmenting the video for summarization. In the first case, authors generate the EVMs using Laplacian Eigenmaps with k -nn and w_{NCC} . In the second, the approach is adapted to create segments by detecting big changes in the imaging conditions of the video. Ordinary changes are produced by the motion of the camera (endoscope) and are smooth. The smoothness of camera motion field is measured by computing the divergence of the optical flow. If the optical flow field $\mathbf{V}(\mathbf{I}_i, \mathbf{I}_j)$ from frame \mathbf{I}_i to frame \mathbf{I}_j is smooth, then the divergence at each location will be close to zero leading to a high similarity. Additionally, the measure leads to high similarity only if the scene and the imaging conditions (blur, specular highlights) are the same. The flow divergence contributions are summed up over the image to define the divergence of optical flow fields (doff):

$$\text{doff}(\mathbf{I}_i, \mathbf{I}_j) = \sum_{x=1}^D \left| \nabla \cdot \mathbf{V}_x(\mathbf{I}_i, \mathbf{I}_j) \right|,$$

where $\nabla \cdot$ is the divergence operator. Finally, the similarity is computed as:

$$w_{ij} = w_{\text{DOFF}}(\mathbf{I}_i, \mathbf{I}_j) = 1 - \left(\frac{\text{doff}(\mathbf{I}_i, \mathbf{I}_j)}{\max \text{doff}(\mathbf{I}_i, \mathbf{I}_j)} \right).$$

In the case of the optical flow similarities, the k -nn of a frame \mathbf{I}_i are searched only within a temporal window around \mathbf{I}_i . The method of Black

and Anandan (Black & Anandan, 1993) is used to compute the optical flow.

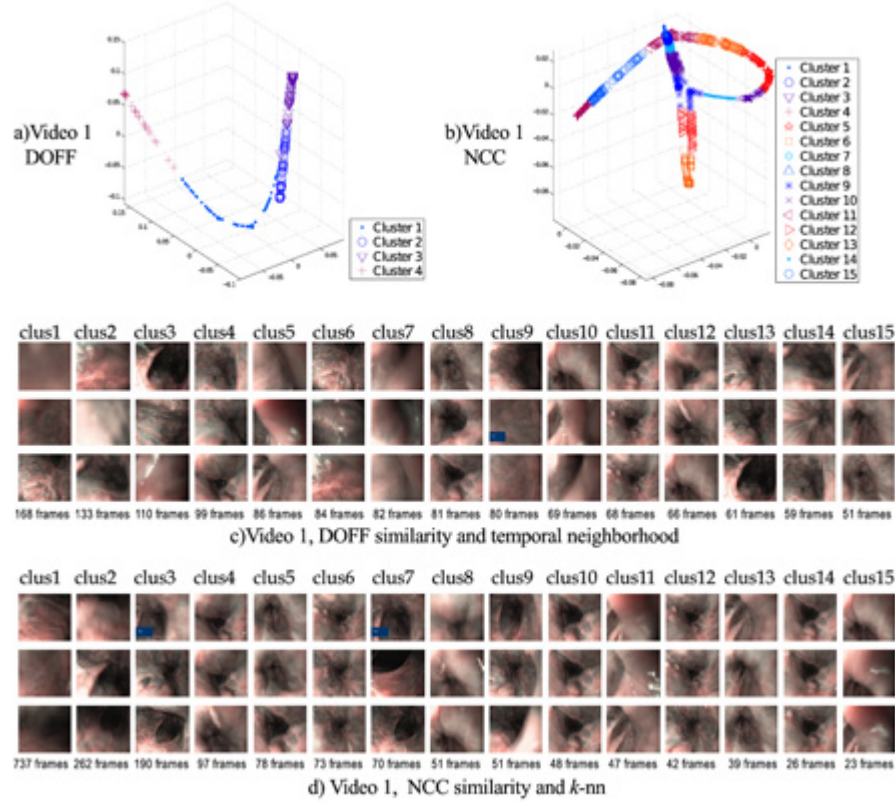
Results of the EVMs created with w_{NCC} and w_{DOFF} and K -means clustering for one video are shown in Figure 3-(a,b). Only the EVM of the largest connected component is shown. The first, mean and last frames of each cluster are shown in the rows of Figure 3-(c,d). The w_{NCC} similarity leads to clusters of similar scenes, independent of whether they are temporally close. They can be used to match videos of the same patient taken at different times. Results using w_{DOFF} and the temporally constrained neighborhoods are rather suitable for video summarizing.

Multi-Modal Registration

Multi-modal registration consists in finding a map between images of the same scene acquired with different imaging modalities. The standard approach to multi-modal registration is to use sophisticated similarity metrics such as mutual information to compare the images. An alternative approach is to create structural representations of the images (Wachinger & Navab, 2010b). These representations are built to capture structure shared by the two images, and thus can be registered as mono-modal images using the L1 or L2 norm. This avoids the need of considering the intensity transformations. As a result, direct application of L1 and L2 norm suffices to align the structural images. In (Wachinger & Navab, 2010b), structural representations are created using Laplacian Eigenmaps as explained below.

Good structural representations are similar across modalities. This requirement is verified when the intrinsic similarities are comparable (i.e. if a pair of patches is [intrinsically] similar in one image, the corresponding pair of patches in the second image should also be [intrinsically] similar). The problem reduces to finding a representation that preserves local (similar) relationships. As this is exactly what Laplacian

Figure 3. Clustering scenes and segmentation of GI endoscopic videos based on intensity and flow-based similarities and Laplacian Eigenmaps



Eigenmaps optimizes for, the algorithm is suitable for creating structural representations.

A structural representation is modeled as the result of a map f that takes a pixel x from image $\mathbf{I} \in \mathbf{R}^D$, where $x \in \{1, \dots, D\}$, and maps it to a descriptor $\mathbf{y}_x \in \mathbf{R}^d$. The descriptor captures the structural information of a patch $\mathbf{p}_x \subset \mathbf{I}$ around x , that is, $f: \mathbf{p}_x \rightarrow \mathbf{y}_x \in \mathbf{R}^d$. An ideal structural representation should exhibit the following properties:

Locality Preservation: two similar patches in the same image $\mathbf{p}_i, \mathbf{p}_j \subset \mathbf{I}$ should be mapped to similar structural descriptors:

$$\|\mathbf{p}_i - \mathbf{p}_j\| \leq \gamma \quad \Rightarrow \quad \|f(\mathbf{p}_i) - f(\mathbf{p}_j)\| \leq \gamma',$$

where both γ and γ' are small numbers. This property is important for both the robustness to noise and the capture range of the registration.

Structural Equivalence: A pair of descriptors obtained for patches from different modalities (1,2), $\mathbf{p}_a \in \mathbf{I}_1$ and $\mathbf{p}_b \in \mathbf{I}_2$, is considered to be structurally equivalent if and only if:

$$f_1(\mathbf{p}_a) = f_2(\mathbf{p}_b),$$

where f_1 and f_2 are the maps associated to modalities 1 and 2, respectively. This enforces the desired representation for multi-modal registration, improves the discrimination, and avoids trivial solutions such as mappings to constant values.

Theoretically, the embedding obtained with Laplacian Eigenmaps gives an optimal approximation to the locality preservation criteria for an

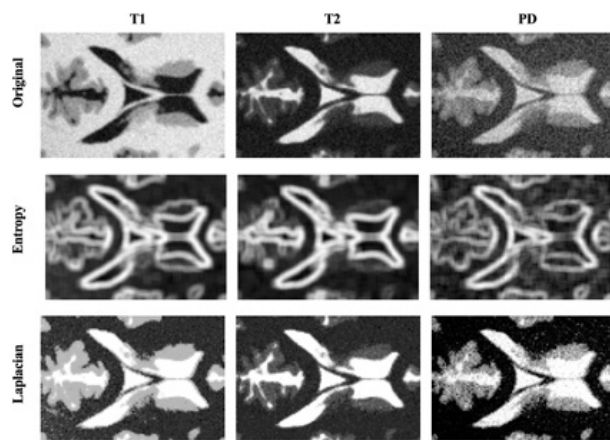
ideal structural description. Indeed, Laplacian Eigenmaps searches for similar patches in the high dimensional patch space and embeds them in a low dimensional by imposing the preservation of locality. For each modality, the embedding of image \mathbf{I} is obtained by representing patches of the image $\{\mathbf{p}_1, \dots, \mathbf{p}_N\}$ with the node of a graph and using a k -nn strategy for the graph connectivity. Weights are found with the Gaussian kernel and the L2 norm, this is equivalent to using the Sum of Squared Differences (SSD) between the patches:

$$\text{diss}_{SSD} = \text{ssd}(\mathbf{p}_i, \mathbf{p}_j) \quad \mathbf{p}_i, \mathbf{p}_j \in \mathbf{I}.$$

Finally, Laplacian Eigenmaps is employed to embed the graph in to a 1D space. The structural image is created using the 1D embedding as the new intensities

Wachinger & Navab (2010b) performed experiments on T1, T2, and PD-weighted MR images from the BrainWeb database1. Images contain 3% noise and 20% intensity non-uniformity for realistic results. Figure 4 shows the original images, the structural images based on Laplacian Eigenmaps and a different type of structural images based on entropy (Wachinger & Navab, 2010c) for comparison. The assumption of comparable internal similarities in the images is justified, because the

Figure 4. Multi-modal Registration with structural Representations. (Top) Examples of Structural representations: a structural representation based on entropy (Wachinger & Navab, 2010c), and the Laplacian based structural representation. (Bottom) Errors for translation (in mm) and rotation (in degrees) for the registration of 3 datasets based on L2 (mono-modal), Mutual Information (MI), and the entropy and Laplacian structural representations.



| Datasets | Similarity | Rotation | Trans X | Trans Y | RMS |
|----------|------------|----------|---------|---------|--------------|
| T1-T2 | L2 | 4.879 | 9.019 | 6.471 | 7.000 |
| | MI | 2.325 | 3.768 | 5.226 | 3.954 |
| | Entropy | 2.084 | 4.539 | 5.231 | 4.180 |
| | Laplacian | 2.584 | 2.061 | 2.1680 | 2.271 |
| T1-PD | L2 | 2.760 | 6.422 | 5.755 | 5.227 |
| | MI | 2.304 | 4.138 | 4.907 | 3.937 |
| | Entropy | 2.283 | 4.782 | 4.750 | 4.108 |
| | Laplacian | 1.750 | 3.007 | 1.929 | 2.297 |
| T2-PD | L2 | 1.784 | 2.947 | 2.916 | 3.942 |
| | MI | 2.161 | 4.628 | 3.812 | 3.680 |
| | Entropy | 1.723 | 4.296 | 3.780 | 3.450 |
| | Laplacian | 1.171 | 2.350 | 1.984 | 1.900 |

appearance of the Laplacian structural images across the modalities is similar. Note the different nature of the entropy and Laplacian images.

To quantify the potential of the Laplacian images for registration, surface plots of the similarity measures for translation from Wachinger & Navab (2010b) are shown in Figure 5 (a-l). A comparative study is done using the L2-norm on the original images, MI on the original images, and L2-norm on entropy and Laplacian images for all combinations of multi-modal alignment. Maxima indicate the best alignment. MI shows a very sharp peak at the correct position, but seems to have a limited capture range. Entropy images also indicate the correct position, but the cost functions contain several local maxima. The cost function based on Laplacian images has the largest capture range. Figure 5-m illustrates similarity plots for rotation. Note the limited capture range for MI and the local maxima for entropy images. Laplacian images lead to a wide and smooth peak, as desired.

Finally, a registration study was performed for all multi-modal image combinations. The random starting position deviates up to 15 mm in translation and 10 in rotation from the correct pose. Figure 4-m (bottom) shows the average absolute error for translation and rotation, together with the overall root mean squared error (RMS), for 100 registration runs for each configuration. 1 mm is weighted equal to 1 degree to quantify translational and angular displacement from the ground truth in one single value. Results confirm the Laplacian structural images are appropriate for multi-modal registration. The proposed method has a significantly lower error in comparison to MI and entropy images, whose performance is comparable.

Ultrasound Image-Based Gating for 4D Volumetric Reconstruction

Breathing motion leads to a significant displacement and deformation of organs in the abdominal

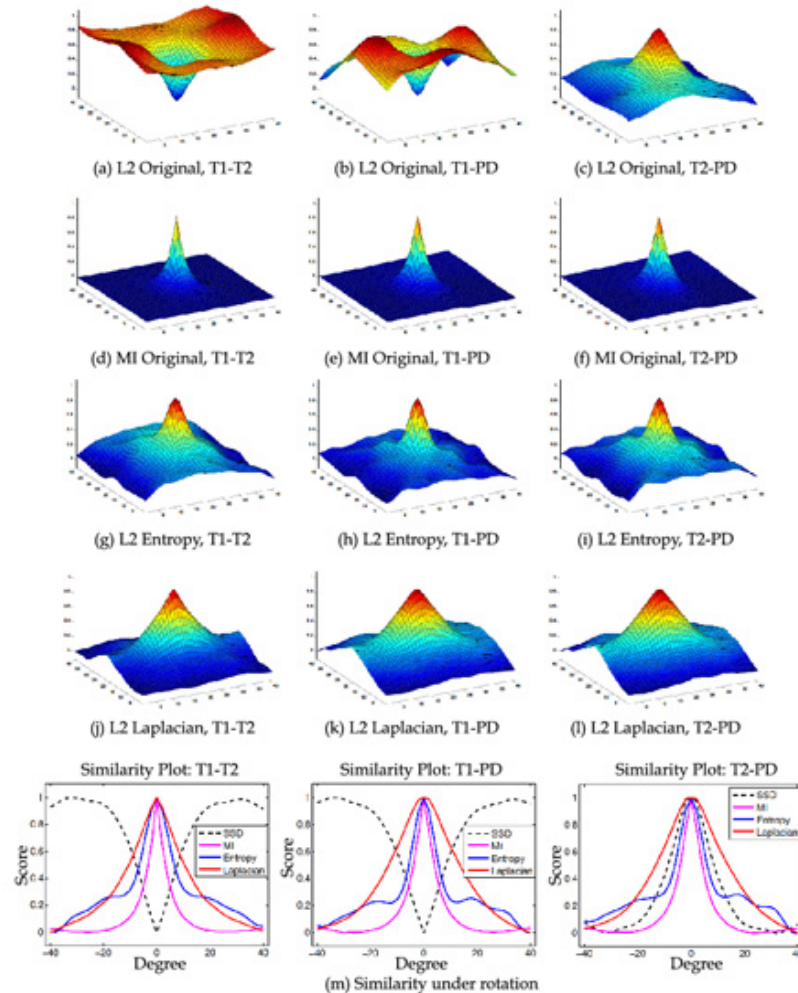
region. This makes the detection of the breathing phase for numerous applications necessary. Usually the assignment of an image to a breathing phase is achieved with external gating devices. These devices have long setup times, prolong the overall acquisition, are costly, and consequently, rarely used in practice. Moreover, the synchronization of image data and breathing signal is not trivial. As an alternative, a purely image-based respiratory gating approach for ultrasound was recently proposed in (Wachinger, et al., 2010d) based on manifold learning. The method is automatic, and does not require prior information or training data.

The method is applied to 2D and 3D mosaicing and to the acquisition of breathing affected 4D ultrasound with a mechanically steered transducer (wobbler). For the later, the image-based gating is used independently to images acquired from the same wobbler angle. Once the respiratory signals are estimated for each angle, the resultant curves are aligned and a globally consistent respiratory signal is computed. The recovered signal allows the reconstruction of volumes for specific breathing stages.

The image-based gating relies on the assumption that the ultrasound images lie on a low dimensional manifold in the ambient space. This suggestion is justified because variations between neighboring slices are smooth, and slices from the same respiratory phase but different acquisition times are similar. The Laplacian Eigenmaps algorithm assigns each ultrasound frame a coordinate in low dimensional space, mapping images of similar breathing stages nearby. The new low-dimensional space is thereby well suited for estimating the breathing cycle.

Consider the N ultrasound images $\{\mathbf{I}_1, \dots, \mathbf{I}_l, \dots, \mathbf{I}_N\}$ that are acquired over several breathing cycles. The idea in (Wachinger, et al., 2010d) is to map these images to a low 1-dimensional space. If the main image changes come from breathing the 1D representation is expected to correlate with the breathing signal. A graph is constructed where each vertex stands for an image \mathbf{I}_i . k -nn are

Figure 5. (a-l) Similarity measures with respect to translation in x and y directions. Maxima indicate best alignment. (m) Plot of similarity measures with respect to rotation.



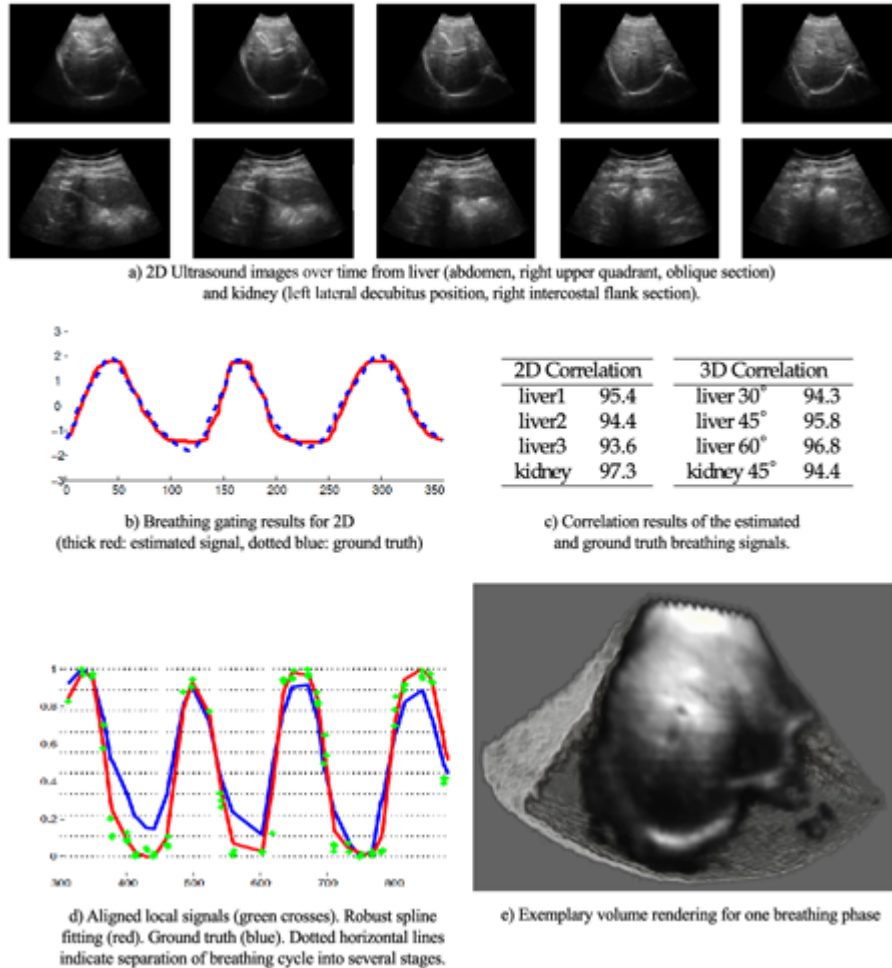
selected to determine the edges based on the SSD image similarity ($k=14$). A Gaussian kernel also based on the SSD is used to assign the weights to the edges. The one-dimensional representation is recovered with Laplacian Eigenmaps. For 4D ultrasound videos, the manifold learning is applied to the images of each angle independently. Then, the resultant curves are aligned, and a robust spline-curve fitting is used to create a globally consistent respiratory signal.

Experiments are performed on multiple patient datasets acquired from different positions, focus-

ing on the liver and kidney. The performance of the image-based gating is quantified by against measurements from an external gating system. Figure 6-a illustrates excerpts of two datasets. The number of images for manifold learning varies between 100 and 300. Figure 6-b shows the result of the respiratory gating for one of the 2D datasets together with the ground truth signal. The estimated signal closely resembles the ground truth, and peaks of the signals coincide. The correlation coefficient computed for multiple 2D datasets is in the range of 95%, confirming

Learning Manifolds

Figure 6. (a) Excerpts of two ultrasound datasets; (b) Estimated breathing phase for a 2D dataset; (c) Correlation of the estimated and ground truth breathing signals; (d) Estimated breathing phase for a 3D dataset; (e) 3D compounding using the breathing signal in (d) for one phase.



the visual similarity of the graphs. Correlation coefficients for four 2D datasets are reported in Figure 6-c-left.

For the 4D experiments, the result of a fitted curve is shown in Figure 6-d. The correlation coefficient between the fitted curves and ground truth for four 3-D datasets is calculated and reported in Figure 6-c-right. The breathing signal is split into 9 different breathing stages, and a 3D volume is compounded for each of the stages. A volume rendering of one of the volumes is shown in Figure 6-e.

Human Motion Analysis

Analysis of human motion is useful for diagnosis and life-quality studies of neurological diseases involving human motion disorders such as Parkinson, multiple sclerosis, or epilepsy (Schwarz, et al., 2009). The main difficulty for a meaningful analysis is the need of capturing motion in everyday environments for extended periods of time. This is only possible with portable sensors. However, human motion is complex and recovering poses from only sparse, noisy and ambiguous

sensor readings is challenging. Schwarz, et al. (2010) propose a learning-based solution to the problem, relying both on a reduced set of inertial orientation sensors, and on prior motion models built during the learning stage. The method allows recovering the human pose and activity over time from the portable sensors. A set of predefined activities is considered, which is in accordance to protocols that physicians employ for diagnosis (e.g. jump, run, walk, etc).

The approach starts with the creation of multiple person-specific *motion models* in the *learning phase*. Synchronous full-body poses and sensor readings of a person performing the activities of interest are captured during this phase. The motion models are created from the low dimensional representation of the full-body poses computed using Laplacian Eigenmaps. Additionally, regression is used to learn a mapping from the low-dimensional representations to both the joint-angle space and to the space of sensor readings. During the *test phase*, the motion models are embedded in a Bayesian tracking framework to determine the pose and activity of the subject. The motion models constrain the high dimensional space of full-body poses and, at the same time, help dealing with the ambiguous information provided by the sensors. Using the learnt motion models, tracking of full-body poses can be done in everyday environments, as only the sensors need to be worn. An overview of the method is shown in Figure 7-a.

Formally, for given a set of M activities of interest identified by an index $\alpha \in \{1, \dots, M\}$, training data consisting of synchronously acquired full-body poses and sensor readings is used to build M motion models. In practice, N^α full-body poses $\mathbf{X}^\alpha = \{\mathbf{x}_1^\alpha, \mathbf{x}_2^\alpha, \dots, \mathbf{x}_{N^\alpha}^\alpha\}$ are recorded with a motion-capture system for each activity, where each vector $\mathbf{x}_i^\alpha \in \mathbf{R}^{D_x}$ describes a human pose represented by D_x joint angles of a skeleton model (here $D_x=35$). Along with the full-body poses inertial sensor observations from a reduced number (4 to 6) of sensors $\mathbf{S}^\alpha = \{\mathbf{s}_1^\alpha, \mathbf{s}_2^\alpha, \dots, \mathbf{s}_{N^\alpha}^\alpha\}$

are also acquired, where each observation $\mathbf{s}_i^\alpha \in \mathbf{R}^{D_s}$ contains a 2D global orientation measurement for all the sensors.

To build the motion models, first, the low dimensional representations are obtained for each activity α . Following the Laplacian Eigenmaps algorithm, a graph is built with the nodes representing each pose \mathbf{x}_i^α . A k -nn system is employed and weights are determined with a Gaussian kernel based on the Euclidean distance:

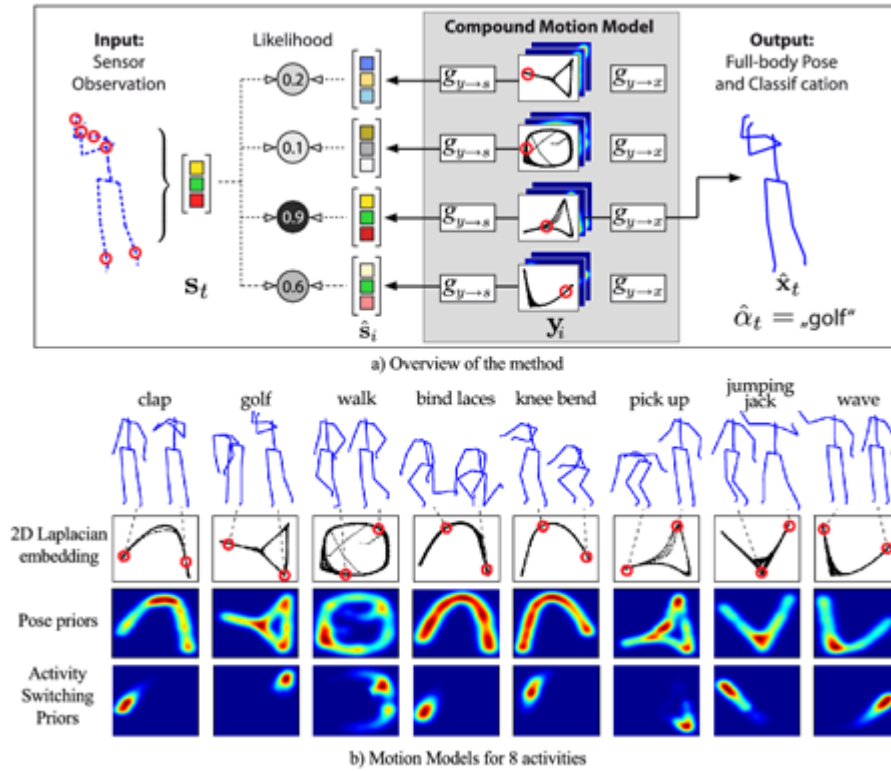
$$\text{diss}_{\text{joint}}(\mathbf{x}_i^\alpha, \mathbf{x}_j^\alpha) = \|\mathbf{x}_i^\alpha - \mathbf{x}_j^\alpha\|^2.$$

As the result of applying the Laplacian Eigenmaps algorithm to the poses of each activity, a series of low dimensional representations $\{\mathbf{Y}^1, \dots, \mathbf{Y}^\alpha, \dots, \mathbf{Y}^M\}$ is obtained, one per activity. Each low dimensional representation contains the coordinates of the mapped points $\mathbf{Y}^\alpha = [\mathbf{y}_1^\alpha, \mathbf{y}_2^\alpha, \dots, \mathbf{y}_{N^\alpha}^\alpha]^\top$ with $\mathbf{y}_i^\alpha \in \mathbf{R}^d \forall i \in \{1, \dots, N^\alpha\}$, and with $d \ll D_x$.

The low dimensional representations for each activity are accompanied with three other elements to build a complete motion model (c.f. Figure 7-b):

- **Predictive Mappings:** are learnt from the training data using non-linear kernel regression, in order to relate poses in low dimensional embedding space to sensor measurements ($g_{\mathbf{y} \rightarrow \mathbf{s}}^\alpha(\mathbf{y})$) and to full-body poses ($g_{\mathbf{y} \rightarrow \mathbf{x}}^\alpha(\mathbf{y})$), with $\mathbf{y} \in \mathbf{R}^d$.
- **Pose Likelihood Prior:** Using kernel density estimation and the training data for each activity, authors derive the likelihood prior for arbitrary poses in low dimensional embedding space $p_{\text{pose}}^\alpha(\mathbf{y})$. Intuitively, poses $\mathbf{y} \in \mathbf{R}^d$ that are close to points \mathbf{y}_i^α in training data should have higher likelihood.

Figure 7. (a) Overview of the human body pose tracking and activity recognition method; (b) Motion models based on a low-dimensional representation obtained with Laplacian Eigenmaps for different activities.



- **Activity Switching Prior:** Distribution $p_{\text{switch}}^{\alpha}(\mathbf{y})$ is defined for every motion model that describes how likely a switch of activity is, given a pose in the low dimensional space $\mathbf{y} \in \mathbf{R}^d$.

Bayesian Tracking Using Multiple Motion Models

The testing phase of the method consists of tracking the pose in low dimensional embedding space. We seek for the most likely pose \mathbf{y}_t and activity α_t , given the observations up to \mathbf{s}_t . Following a generative model and a standard Bayesian tracking formulation, the solution is given by the maximum of the posterior probability at time t , shown in Exhibit 4.

The *observation model* $p(\mathbf{s}_t | \mathbf{y}_t, \alpha_t)$ relates observations to the learned motion model, and is defined as a product of three terms: first, the *prediction likelihood* measuring the likelihood of observation \mathbf{s}_t given the predicted sensor reading $g_{y \rightarrow s}^{\alpha_t}(\mathbf{y}_t)$ from pose \mathbf{y}_t ; second, a *temporal smoothness term*, and third, the *pose likelihood prior* $p_{\text{pose}}^{\alpha_t}(\mathbf{y})$ encouraging poses close to the training data.

The *dynamics model* $p(\mathbf{y}_t, \alpha_t | \mathbf{y}_{t-1}, \alpha_{t-1})$ determines how pose estimates are updated from one time step to the next. Following (Isard & Blake, 1998), a model

$$p(\mathbf{y}_t, \alpha_t | \mathbf{y}_{t-1}, \alpha_{t-1}) = p(\mathbf{y}_t | \mathbf{y}_{t-1}, \alpha_t, \alpha_{t-1})p(\alpha_t | \mathbf{y}_{t-1}, \alpha_{t-1})$$

Exhibit 4.

$$\begin{aligned}
 p(\mathbf{y}_t, \alpha_t | \mathbf{s}_t) &= p(\mathbf{s}_t | \mathbf{y}_t, \alpha_t) p(\mathbf{y}_t, \alpha_t | \mathbf{s}_{t-1}) \\
 &= p(\mathbf{s}_t | \mathbf{y}_t, \alpha_t) \iint p(\mathbf{y}_t, \alpha_t | \mathbf{y}_{t-1}, \alpha_{t-1}) p(\mathbf{y}_{t-1}, \alpha_{t-1} | \mathbf{s}_{t-1}) d\mathbf{y}_{t-1} d\alpha_{t-1}.
 \end{aligned}$$

is used that factors the activity dynamics from the dynamics of the pose:

- Assuming that all sequences of consecutive activities are equally likely, the probability of switching from a given activity α_{t-1} to any other activity only depends on the previous pose \mathbf{y}_{t-1} . Thus, the *activity dynamics model* is based only on the activity-switching prior

$$p(\alpha_t | \mathbf{y}_{t-1}, \alpha_{t-1}) = p_{\text{switch}}^{\alpha_{t-1}}(\mathbf{y}_{t-1}).$$
- The *pose dynamics model* governs the evolution of poses in embedding space. When there is no switch of activity ($\alpha_t = \alpha_{t-1}$), dynamics are governed by a random walk. In the case of activity switching ($\alpha_t \neq \alpha_{t-1}$), the dynamics model follows the pose likelihood prior $p_{\text{pose}}^{\alpha_t}(\mathbf{y})$ of activity α_t .

A particle filter (Isard & Blake, 1998) adapted to use multiple motion models is used to sample the posterior density. This allows simultaneously evaluating pose hypotheses in different motion models and selecting the most appropriate model. The current activity $\hat{\alpha}_t$ is estimated as the most frequent activity among the highest-weight particles. The pose estimate $\hat{\mathbf{y}}_t$ in low dimensional space is computed as a convex combination of the positions of the highest-weight particles with activity $\hat{\alpha}_t$. The full-body pose at time t is obtained as $\hat{\mathbf{x}}_t = g_{\mathbf{y} \rightarrow \mathbf{x}}^{\hat{\alpha}_t}(\hat{\mathbf{y}}_t)$. Figure 8-a illustrates the particle filter algorithm for a testing sequence switching from the “waving” to the “golfing” activity.

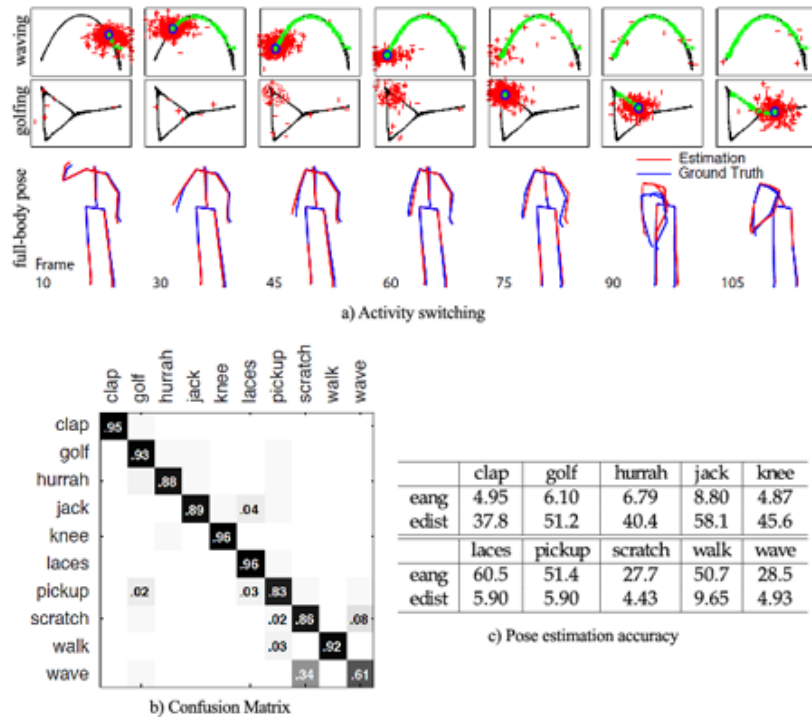
For the experiments, a synchronized dataset of full-body poses \mathbf{X}^α and sensor values \mathbf{S}^α , $\alpha \in \{1, \dots, M\}$ is obtained using a motion capture system and six wearable inertial orientation sensors placed on wrists, upper arms and shin-bones. An observation $\mathbf{s}_i^\alpha \in \mathbf{R}^{D_s}$ has $D_s = 12$. In the training phase, a 2D manifold embedding \mathbf{Y}^α is learned for each activity. Authors consider $M = 10$ activities: clapping, golfing, arms up, jumping jack, knee bends, binding laces, picking up, scratching head, walking and waving. Each of the movements was recorded 6 times with 9 persons. Every movement has 600 frames. The testing data consists of 5 sequences per person containing all activities (~2000 frames each). For tracking, only the inertial sensor values were used, the motion capture data served as ground truth. All experiments were performed in a cross-validation scheme.

Activity Classification. The number of particles per activity manifold is an indicator of activity class membership. The confusion matrix in Figure 8-b giving the classification rates for all activities over all testing sequences is almost diagonal. Misclassifications mainly occur at the beginning and end of activities, which is normal, as the person stands idle. On average, authors report a correct classification rate of 89% for all non-idle frames.

Pose Estimation Accuracy. To measure how precisely the poses estimated match the ground truth authors use two metrics. The angular error e_{ang} gives the deviation from the ground truth in terms of joint angles. The distance error e_{dist} is the difference in 3D space between predicted

Learning Manifolds

Figure 8. (a) Particle filter-based activity switching: Two manifold embeddings over several frames with particles (red crosses), particles used for predicting full-body pose (dark circles), and trace of previous frames (green crosses). Shown below the corresponding predicted and ground-truth body poses; (b) Activity classification. Top: Number of sampled particles per frame and per activity. Bottom: Ground truth classification and predicted activities for each frame of the sequence; (c) Confusion matrix computed from the classification results for all testing sequences; (d) Pose estimation accuracy for all activities.



joint locations and the ground truth. Averaged over all frames of the testing sequences, the algorithm achieved $e_{ang} = 6.23^\circ$ per joint and $e_{dist} = 45.2\text{mm}$. As shown in Figure 8-c, the deviation from the ground truth only increases for fast movements with a large variability, such as jumping jack or walking. Results are comparable to other state-of-the-art methods that use visual observations (Vlasic, et al., 2007).

DISCUSSION

In the previous sections we have presented methods that use manifold learning to deal with different medical applications. The medical field

seems to be well adapted for such approaches, as many datasets comply with the manifold assumption. In such cases, manifold learning methods recover non-linear low dimensional maps that capture data relationships better than the linear ones. We have focused on Laplacian Eigenmaps because it is mathematically well supported and preserves local neighborhoods. Other algorithms rely on different preservation criteria and give different results. When choosing an algorithm it is important to be aware of the cost function being optimized and the constraints imposed on the mapping (from the high to the low dimensional space), and verify they are consistent with the goals of the application. After choosing the algorithm there are design tools (e.g. the neigh-

neighborhood system and pairwise similarities) which are determinant to the outcome of the map. If the manifold assumption is broken, the sampling of the manifold is not regular, or the chosen method and design tools are not well adapted to the application, it is expected that manifold learning does not behave better than linear reduction methods (van der Maaten, et al., 2009). Also it should be noticed that when prior information on the nature of the manifold is available or when data samples are generated in a controlled manner, it may be more appropriate to approximate the manifold's metric or to fit a parameterized manifold instead of learning it (Baloch & Davatzikos, 2009; Xie, et al., 2010).

Concerning the design tools, we have described the basic k -nn and ϵ -ball neighborhood systems, and discussed the possibility to define the neighbors to account for temporal constraints. Besides a grid search, there is currently no systematic way of computing the parameters k and ϵ . More complex systems (e.g. adaptive neighborhoods [Wang, et al., 2005]) can be useful for compensating non-uniform samplings. The neighborhood system can also be used to incorporate partial prior knowledge leading to a semi-supervised embedding of the graph (LeCun, et al., 2006; Tiwari, et al., 2010).

Concerning the design of the weights, it is most common to use the Gaussian kernel with an L2 norm. Studies of different image similarities are presented in Souvenir & Pless (2007), Atasoy et al. (2010), including SSD, NCC, the power spectrum and the divergence of the optical flow. Other ways to compute the weights include the use of neural networks (Lim, et al., 2003), phase difference of local complex Gabor filters (Zhang, et al., 2006), shape similarities based on the Sobolev norm and signed distance functions (Etyngier, et al., 2007), approximations to the diffeomorphic metric (Gerber, et al., 2010), diffeomorphic similarities (Sparks & Madabhushi, 2010), structural segmentation overlaps (Aljabar, et al., 2008) and the norm of the displacement field after registration (Aljabar, et al., 2010). Finally, Tiwari, et al.

(2010) explore the computation of the weights using multi-kernel to integrate heterogeneous information from independent channels.

One of the difficulties of using manifold learning is that, as opposed to linear maps, the algorithms do not inherently provide a function to map new data points to the low-dimensional space. This problem has been approached with the approximation of locally linear projections (He, et al., 2005) or by means of kernel-regression (Bengio, et al., 2003, Etyngier, et al., 2007; Gerber, et al., 2010). Such maps are required when the low-dimensional representation is used within a generative model (Schwarz, et al., 2010). There also exist non-linear reduction methods that directly learn mappings from and to the low dimensional representation (Lawrence, 2005; Carreira-Perpiñan & Lu, 2008). A second difficulty of manifold learning is the complexity of the eigenanalysis. When the systems are sparse, as is the case in Laplacian Eigenmaps, efficient solutions exist. However, very large datasets may still require approximate solutions (Talwalkar, et al., 2008). The last difficulty concerns the dimension of the low-dimensional space, which is not automatically determined by the algorithms. As opposed to PCA, eigenvalues of spectral manifold methods do not provide information on the dimension. In cases of simple movements (e.g. respiratory motion) a small number of dimensions generally suffices, however without a-priori knowledge the problem of determining a suitable number of dimensions for the embeddings remains unsolved.

CONCLUSION

In this chapter we have discussed the design of manifold learning algorithms for medical applications. A review of the state of the art in the field is provided showing the increasing interest of researchers in the domain and a wide range of applications where these methods can be applied. We recalled the theory and assumptions for

manifold learning and described in details one exemplary algorithm (Laplacian Eigenmaps). Five case studies were presented showing a parallel on the adaption of this method to very different applications and illustrating the key design elements in each case. The last case study additionally shows how to integrate low dimensional representations created with manifold learning in a generative framework. Although manifold learning is attractive because of its flexibility and non-linearity, several difficulties may arise in practice; in particular, the lack of a function to relate the high and low dimensional spaces, and of some means to determine the parameters. Most important to ensure a successful application of manifold learning techniques are:

- verifying that the data complies with the manifold assumption and that it is approximately regularly sampled,
- choosing an algorithm whose cost function and constraints preserve the relevant information for the application,
- and designing pairwise relationships that incorporate the domain knowledge to best approximate the manifold.

REFERENCES

- Aljabar, P., Rueckert, D., & Crum, W. (2008). Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. *NeuroImage*, *43*(2), 225–235. doi:10.1016/j.neuroimage.2008.07.055
- Aljabar, P., & Wolz, R. Srinivasan, L., Counsell, S., Boardman, J.P., Murgasova, M., ... Rueckert, D. (2010). *Combining morphological information in a manifold learning framework: Application to neonatal MRI*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Atasoy, S., Mateus, D., Lallemand, J., Meining, A., Yang, G. Z., & Navab, N. (2010) *Endoscopic video manifolds (EVMs)*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Baloch, S., & Davatzikos, C. (2009). Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis. *NeuroImage*, *45*(1), S73–S85. doi:10.1016/j.neuroimage.2008.10.048
- Bartenhagen, C., Klein, H. U., Ruckert, C., Jiang, X., & Dugas, M. (2010). Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. *BMC Bioinformatics*, *11*, 567. doi:10.1186/1471-2105-11-567
- Belkin, M., & Nigoyi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373–1396. doi:10.1162/089976603321780317
- Bengio, Y., Delalleau, O., Le Roux, N., Païement, J.-F., Vincent, P., & Ouimet, M. (2004). Learning eigenfunctions links spectral embedding and Kernel PCA. *Neural Computation*, *16*(10), 2197–2219. doi:10.1162/0899766041732396
- Bengio, Y., Païement, J. F., Vincent, P., Delalleau, O., Le Roux, N., & Ouimet, M. (2003). Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering. [MIT Press.]. *Advances in Neural Information Processing Systems*, 177–184.
- Black, M., & Anandan, P. (1993). A framework for the robust estimation of optical flow. In *Proceedings on International Conference on Computer Vision*, (pp. 231-236). IEEE Press.
- Carreira-Perpiñan, M., & Lu, Z. (2008). *Dimensionality reduction by unsupervised regression*. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition.

- Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional Scaling*. London, UK: Chapman and Hall.
- Donoho, D.L. & Grimes, C. (2005). Hessian eigenmaps: New locally linear embedding techniques for high dimensional data. *National Academy of Sciences*, 102(21), 7426–7431.
- Elmoataz, A., Lezoray, O., & Bougleux, S. (2008). Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing. *IEEE Transactions on Image Processing*, 17(7), 1047–1060. doi:10.1109/TIP.2008.924284
- Etyngier, P., Ségonne, F., & Keriven, R. (2007). Active-contour-based image segmentation using machine learning techniques. In N. Ayache, S. Ourselin, & A. Maeder (Eds.), *International Conference on Medical Image Computing and Computer Assisted Intervention*. Berlin, Germany: Springer.
- Finn, W.G., Carter, K.M., Raich, R., Stoolman, L.M., & Hero, A.O. (2009). Analysis of clinical flow cytometric immunophenotyping data by clustering on statistical manifolds: Treating flow cytometry data as high dimensional objects. *Cytometry*, 76(B), 1–7.
- Georg, M., Souvenir, R., Hope, A., & Pless, R. (2008). *Simultaneous data volume reconstruction and pose estimation from slice samples*. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition.
- Gerber, S., Tasdizen, T., Fletcher, T. P., & Whitaker, R. (2010). Manifold modeling for brain population analysis. *Medical Image Analysis*, 14(5), 643–653. doi:10.1016/j.media.2010.05.008
- Hamm, J., Ye, D. H., Verma, R., & Davatzikos, C. (2010). GRAM: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*, 14(5), 633–642. doi:10.1016/j.media.2010.06.001
- He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H.-J. (2005). Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3), 328–340. doi:10.1109/TPAMI.2005.55
- Hein, M., Audibert, J. Y., & von Luxburg, U. (2007). Graph laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8, 1325–1370.
- Isard, M., & Blake, A. (1998). A mixed-state condensation tracker with automatic model switching. In *Proceedings on International Conference on Computer Vision*, (pp. 107-112). IEEE Press.
- Kadoury, S., & Paragios, N. (2010). *Nonlinear embedding towards articulated spine shape inference using higher-order MRFs*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Lafon, S., & Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and dataset parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9), 1393–1403. doi:10.1109/TPAMI.2006.184
- Lawrence, N. D. (2005). Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Journal of Machine Learning Research*, 6, 1783–1816.
- LeCun, Y., Hadsell, R., & Chopra, S. (2006). *Dimensionality reduction by learning an invariant mapping*. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition.
- Lee, G., Rodriguez, C., & Madabhushi, A. (2008). Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein-expression studies. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(3), 368–384. doi:10.1109/TCBB.2008.36

- Lim, I., Ciechomski, P., Sarni, S., & Thalmann, D. (2003). Planar arrangement of high dimensional biomedical datasets by isomap coordinates. In *Proceedings on IEEE Symposium on Computer-Based Medical Systems*, (pp. 50–55). IEEE Press.
- Lin, T., & Zha, H. (2008). Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 796–809. doi:10.1109/TPAMI.2007.70735
- Nilsson, J., Fioretos, T., Höglund, M., & Fontes, M. (2004). Approximate geodesic distances reveal biologically relevant structures in microarray data. *Bioinformatics (Oxford, England)*, 20(6), 874–880. doi:10.1093/bioinformatics/btg496
- Pless, R., & Souvenir, R. (2009). A survey of manifold learning for images. *IPSJ Transactions on Computer Vision and Applications*, 1, 83–94. doi:10.2197/ipsjtcva.1.83
- Roweis, S., & Saul, L. (2000)... *Science*, 290(5500), 2323–2326. doi:10.1126/science.290.5500.2323
- Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5), 401–409. doi:10.1109/T-C.1969.222678
- Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299–1319. doi:10.1162/089976698300017467
- Schwarz, L., Mateus, D., & Navab, N. (2009). *Discriminative human full-body pose estimation from wearable inertial sensor data*. Paper presented at International Workshop on 3D Physiological Human. Zermatt, Switzerland.
- Schwarz, L., Mateus, D., & Navab, N. (2010). Multiple-activity human body tracking in unconstrained environments. In F.J. Perales & R.B. Fisher (Eds.), *Conference on Articulated Motion and Deformable Objects*. Berlin, Germany: Springer.
- Souvenir, R., & Pless, R. (2007). Image distance functions for manifold learning. *Image and Vision Computing*, 25(3), 365–373. doi:10.1016/j.imavis.2006.01.016
- Sparks, R., & Madabhushi, A. (2010). *Novel morphometric based classification via diffeomorphic based shape representation using manifold learning*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Suzuki, K., Zhang, J., & Xu, J. (2010). Massive-training artificial neural network coupled with Laplacian-eigenfunction-based dimensionality reduction for computer-aided detection of polyps in CT colonography. *IEEE Transactions on Medical Imaging*, 29, 1907–1917. doi:10.1109/TMI.2010.2053213
- Talwalkar, A., Kumar, S., & Rowley, H. A. (2008). Large-scale manifold learning. In *Proceedings on IEEE Computer Vision and Pattern Recognition*. IEEE Press.
- Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. doi:10.1126/science.290.5500.2319
- Tiwari, P., Kurhanewicz, J., Rosen, M., & Madabhushi, A. (2010). *Semi supervised multi-kernel (SeSMiK) graph embedding: Identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Tiwari, P., Rosen, M., & Madabhushi, A. (2008). Consensus-locally linear embedding (C-LLE): Application to prostate cancer detection on magnetic resonance spectroscopy. In D. Metaxas, L. Axel, G. Fichtinger, & G. Székely (Eds.), *International Conference on Medical Image Computing and Computer Assisted Intervention*, vol 2, (pp. 330-338). IEEE Press.

- Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., & Madabhushi, A. (2009). Spectral embedding based probabilistic boosting tree (ScEPTre): Classifying high dimensional heterogeneous biomedical data. In G.Z. Yang, D. Hawkes, D. Rueckert, A. Noble, & C. Taylor (Eds.), *International Conference on Medical Image Computing and Computer Assisted Intervention*. vol 5762, (pp. 844–851). Berlin, Germany: Springer.
- van der Maaten, L.J.P., Postma, E.O., & van den Herik, H.J. (2009). Dimensionality reduction: A comparative review. *Tilburg University Technical Report*. TiCC-TR 2009-005.
- Verma, R., Khurd, P., & Davatzikos, C. (2007). On analyzing diffusion tensor images by identifying manifold structure using isomaps. *IEEE Transactions on Medical Imaging*, 26(6), 772–778. doi:10.1109/TMI.2006.891484
- Vlasic, D., Adelsberger, R., Vannucci, G., & Barnwell, J. (2007). *Practical motion capture in everyday surroundings*. *ACM Transactions on Graphics*. ACM Press.
- Wachinger, C., Mateus, D., Keil, A., & Navab, N. (2010a). Manifold learning for patient position detection in MRI. In W. Niessen & E. Meijering (Eds.). *International Symposium on Biomedical Imaging*. IEEE Press.
- Wachinger, C., & Navab, N. (2010b). Manifold learning for multi-modal image registration. In F. Labrosse, R. Zwigelaar, Y. Liu, & B. Tiddeman (Eds.), *British Machine Vision Conference*. BMVA Press.
- Wachinger, C., & Navab, N. (2010c). *Structural image representation for image registration*. Paper presented at IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis.
- Wachinger, C., Yigitsoy, M., & Navab, N. (2010d). *Manifold learning for image-based breathing gating with application to 4D ultrasound*. Paper presented at International Conference on Medical Image Computing and Computer Assisted Intervention.
- Wang, J., Zhang, Z., & Zha, H. (2005). Adaptive manifold learning. [Cambridge, MA: MIT Press.]. *Advances in Neural Information Processing Systems*, 17, 1473–1480.
- Wolz, R., Aljabar, P., Hajnal, J. V., Hammers, A., & Rueckert, D. (2010). LEAP: Learning embeddings for atlas propagation. *NeuroImage*, 49(2), 1316–1325. doi:10.1016/j.neuroimage.2009.09.069
- Xie, Y., Ho, J., & Vemuri, B. C. (2010). Image atlas construction via intrinsic averaging on the manifold of images. In *Proceedings on IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 2933-2939). IEEE Press.
- Zhang, Q., Souvenir, R., & Pless, R. (2006). On manifold structure of cardiac MRI data: Application to segmentation. In *Proceedings on IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 1092-1098). IEEE Press.

ADDITIONAL READING

- Baraniuk, R. G., & Wakin, M. B. (2009). Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1), 51–77. doi:10.1007/s10208-007-9011-z
- Bartenhagen, C., Klein, H. U., Ruckert, C., Jiang, X., & Dugas, M. (2010). Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. *BMC Bioinformatics*, 11, 567. doi:10.1186/1471-2105-11-567

- Bishop, C., Svensen, M., & Williams, C. (1998). GTM: The generative topographic mapping. *Neural Computation*, *10*(1), 215–234. doi:10.1162/089976698300017953
- Burges, C. J. C. (2005). Geometric methods for feature selection and dimensional reduction: A guided tour. In *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers.
- de Silva, V., & Tenenbaum, J. B. (2003). Global versus local methods in nonlinear dimensionality reduction. [Cambridge, MA: MIT Press.]. *Advances in Neural Information Processing Systems*, *15*, 721–728.
- Gerber, S., Tasdizen, T., & Whitaker, R. (2009). *Dimensionality reduction and principal surfaces via kernel map manifolds*. Paper presented at International Conference on Computer Vision.
- Hamm, J., Lee, D., Mika, S., & Schölkopf, B. A. (2003). Kernel view of the dimensionality reduction of manifolds. *Technical Report TR-110*. Max Planck Institute for Biological Cybernetics.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. doi:10.1126/science.1127647
- Lafon, S., Keller, Y., & Coifman, R. R. (2006). Data fusion and multicue data matching by diffusion maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(11), 1784–1797. doi:10.1109/TPAMI.2006.223
- Langs, G., & Paragios, N. (2008). *Modeling the structure of multivariate manifolds: Shape maps*. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition.
- Lee, J. A., & Verleysen, M. (2007). *Nonlinear dimensionality reduction*. New York, NY: Springer. doi:10.1007/978-0-387-39351-3
- Li, H., Teng, L., Chen, W., & Shen, I.-F. (2005). *Lecture Notes on Computer Science: Vol. 3496. Supervised learning on local tangent space* (pp. 546–551). Berlin, Germany: Springer Verlag.
- Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., & Boyer, E. (2008). *Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration*. Paper presented at IEEE Computer Vision and Pattern Recognition. Anchorage, AK.
- Pless, R. (2003). Using isomap to explore video sequences. In *Proceedings on International Conference on Computer Vision*, (pp. 1433-1440). IEEE Press.
- Saul, L. K., Weinberger, K. Q., Ham, J. H., Sha, F., & Lee, D. D. (2006). Spectral methods for dimensionality reduction. In *Semisupervised Learning*. Cambridge, MA: MIT Press.
- Schölkopf, B., Herbrich, R., & Smola, A. (2001). A generalized representer theorem. In D. P. Helmbold & B. Williamson (Eds.), *Annual Conference on Computational Learning Theory and European Conference on Computational Learning Theory*, (pp. 416-426). London, UK: Springer-Verlag.
- Shi, J., & Malik, J. (2000). *Normalized cuts and image segmentation*. Paper presented at IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Tsai, F. S. (2010). Comparative study of dimensionality reduction techniques for data visualization. *Journal of Artificial Intelligence*, *3*, 119–134. doi:10.3923/jai.2010.119.134
- van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. doi:10.1007/s11222-007-9033-z

Yang, X., Fu, H., Zha, H., & Barlow, J. (2006). Semi-supervised nonlinear dimensionality reduction. In *Proceedings on International Conference on Machine Learning*, (pp. 1065-1072). New York, NY: ACM Press.

Zhang, Z., & Zha, H. (2004). Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1), 313–338. doi:10.1137/S1064827502419154

ENDNOTE

- ¹ Let \mathbf{w}_j be a column (or row) of \mathbf{W} , with $j \in \{1, \dots, N\}$, and denote the j -th element of vector $\mathbf{y}^{(l)}$ as $y_j^{(l)}$, then (see Exhibit 5).

Exhibit 5.

$$\begin{aligned}
 \sum_{\mathbf{x}_i \sim \mathbf{x}_j} w_{ij} \mathbf{y}_i^\top \mathbf{y}_j &= \sum_{l=1}^d \sum_{j=1}^N \sum_{i=1}^N w_{ij} y_i^{(l)} y_j^{(l)} = \sum_{l=1}^d \sum_{j=1}^N y_j^{(l)} \sum_{i=1}^N w_{ij} y_i^{(l)} = \sum_{l=1}^d \sum_{j=1}^N w_{ij} y_j^{(l)} \mathbf{y}^{(l)\top} \mathbf{w}_j \\
 &= \sum_{l=1}^d \mathbf{y}^{(l)\top} \sum_{j=1}^N \mathbf{w}_j y_j^{(l)} = \sum_{l=1}^d \mathbf{y}^{(l)\top} \mathbf{W} \mathbf{y}^{(l)} \\
 &= \text{trace} \left(\begin{bmatrix} \mathbf{y}^{(1)\top} \\ \vdots \\ \mathbf{y}^{(d)\top} \end{bmatrix} \mathbf{W} [\mathbf{y}^{(1)} \quad \dots \quad \mathbf{y}^{(d)}] \right) = \text{trace} \left([\mathbf{y}_1 \quad \dots \quad \mathbf{y}_N] \mathbf{W} \begin{bmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_N^\top \end{bmatrix} \right) \\
 &= \text{trace}(\mathbf{Y}^\top \mathbf{W} \mathbf{Y})
 \end{aligned}$$