Dimensionality Reduction Techniques for Statistical Analysis and Modeling in Medical Imaging

Jakob Wasza
Medical Image Registration Group
Pattern Recognition Lab, Friedrich-Alexander Universität Erlangen-Nürnberg

MIRC, 8/2/2012

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Agenda

- Introduction
  - Dimensionality Reduction
  - Statistical Analysis and Modeling
  - Nomenclature
- Linear Dimensionality Reduction
  - Principal Component Analysis (PCA)
  - Multidimensional Scaling (MDS)
  - Applications
- Non-linear Dimensionality Reduction
  - Kernel Principal Component Analysis (KPCA)
  - Manifold Learning and Graph-Based Methods
  - Applications

Motivation

- Dimensionality Reduction
- Statistical Analysis and Modeling
- Nomenclature

Dimensionality Reduction

- Curse of dimensionality
  - No statistical significance due to sparse sampling
  - Distance metrics not meaningful
  - Computational burden

- Visualization
  - Intuitive feel what the data looks like
  - Analysis of machine learning algorithms

- Underlying forces
  - How are samples created?
  - Important features and structures
  - Removal of misleading features

Statistical Analysis and Modeling

- Statistical analysis
  - **Describe**, **understand** and **predict** a population based on sample datasets (inference)
  - Characterize data in a short and compact form
  - *Out-of-sample* problem

- Statistical modeling
  - **Reconstruction** of a complex entity from a compact statistical description
  - *Pre-image* problem

Underlying statistical model?

Nomenclature

- High-dimensional (mean-centered) input points

$$\{\boldsymbol{x}_i\}_{i=1}^n \ , \ \boldsymbol{x}_i \in \mathbb{R}^D \ , \ \sum_i \boldsymbol{x}_i = \boldsymbol{0}$$

- Low-dimensional output points

$$\boldsymbol{y}_i \in \mathbb{R}^d \ , \ d << D$$

- Matrix notation

$$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{D \times n}$$

- Inner products

$$\boldsymbol{x}_i^\top \boldsymbol{x}_j = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j)$$

Nomenclature

- *Out-of-sample* problem
  - For an **unseen** sample $\widehat{\boldsymbol{x}} \in \mathbb{R}^D$, $\notin \{\boldsymbol{x}_i\}_{i=1}^n$ what is its corresponding low-dimensional representation $\widehat{\boldsymbol{y}} \in \mathbb{R}^d$?
  - This may be a non-trivial task...

- *Pre-image* problem
  - Given an **arbitrary** low-dimensional representation $\widehat{\boldsymbol{y}} \in \mathbb{R}^d$, what is the corresponding vector $\widehat{\boldsymbol{x}} \in \mathbb{R}^D$ in input space?
  - For many dimensionality reduction techniques, the exact pre-image simply does not exist!

Linear Dimensionality Reduction

- Principal Component Analysis (PCA)
- Multidimensional Scaling (MDS)
- Applications

Principal Component Analysis (PCA)

- Find the directions (i) along which the data has **maximum variance** and (ii) the **relative importance** of these directions

- Minimum reconstruction error:

$$\mathcal{J}_{\text{PCA}} = \sum_i \left\| \boldsymbol{x}_i - \sum_{j=1}^{d} (\boldsymbol{x}_i \cdot \boldsymbol{e}_j)\, \boldsymbol{e}_j \right\|^2 = \sum_i \|\boldsymbol{x}_i - \boldsymbol{E}\boldsymbol{y}_i\|^2$$

subject to $(\boldsymbol{e}_i \cdot \boldsymbol{e}_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

- Solution given by Eigen-decomposition of the covariance matrix

$$\boldsymbol{C} = \frac{1}{n} \sum_i \boldsymbol{x}_i \boldsymbol{x}_i^\top = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^\top \ , \ \boldsymbol{\Lambda} = diag\,(\lambda_1, \ldots, \lambda_D)$$

Principal Component Analysis (PCA)

- *Out-of-sample* problem
  - Principal axes $\boldsymbol{E}$ derived from the training set $\{\boldsymbol{x}_i\}_{i=1}^n$
  - For an unseen sample $\widehat{\boldsymbol{x}} \in \mathbb{R}^D$ its low-dimensional representation is:

$$\widehat{\boldsymbol{y}} = \boldsymbol{E}^\top \widehat{\boldsymbol{x}}$$

- *Pre-image* problem
  - For an arbitrary low-dimensional representation $\widehat{\boldsymbol{y}} \in \mathbb{R}^d$ the corresponding vector in input space is

$$\widehat{\boldsymbol{x}} = \boldsymbol{E}\widehat{\boldsymbol{y}}$$

  - This follows from the orthonormality of $\boldsymbol{E}$

This is straight-forward...

Principal Component Analysis (PCA)

- PCA requires the data to lie on a d-dimensional **linear subspace**
  - This will introduce some error in practice
  - Will definitely fail for curved manifolds

Second mode essentially for error compensation

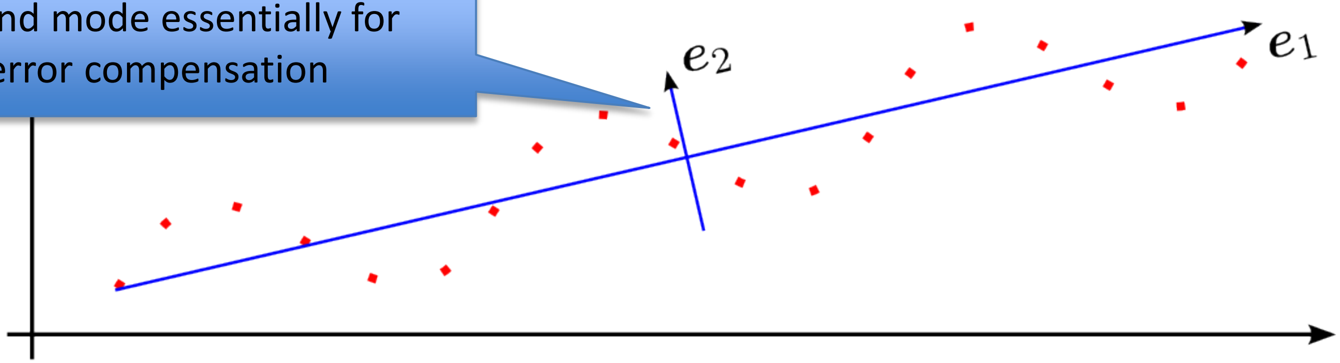


Figure: PCA applied to samples drawn from a sine curve.

Principal Component Analysis (PCA)

- PCA yields global modes of variation
  - Sparsity is sacrificed for the sake of variance maximization
  - Remedies: Factor-Rotations [1], Sparsity-Regularization [2,3]

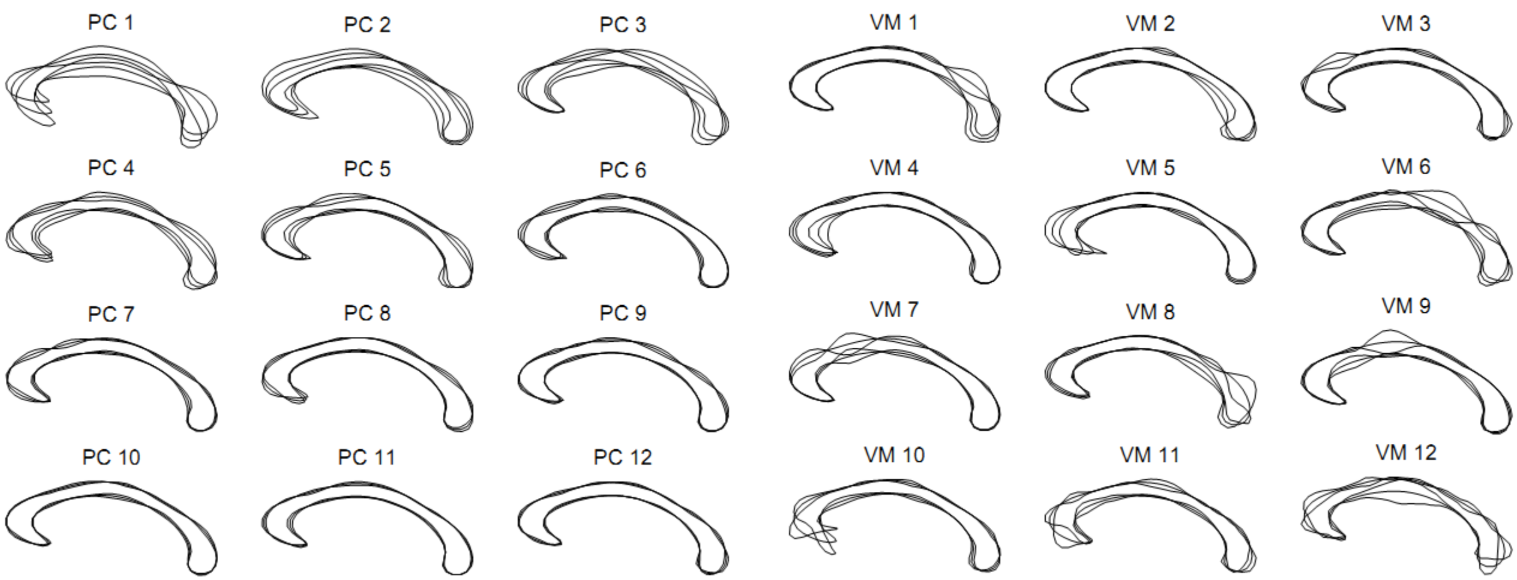


Figure: PCA and varimax rotations applied to corpus callosum annotations [1].

Multidimensional Scaling (MDS)

- Compute low-dimensional representation that **preserves mutual inner products**
  - Pairwise distances (equals PCA)
  - Mutual angles
  - Generalized metrics

- Objective function

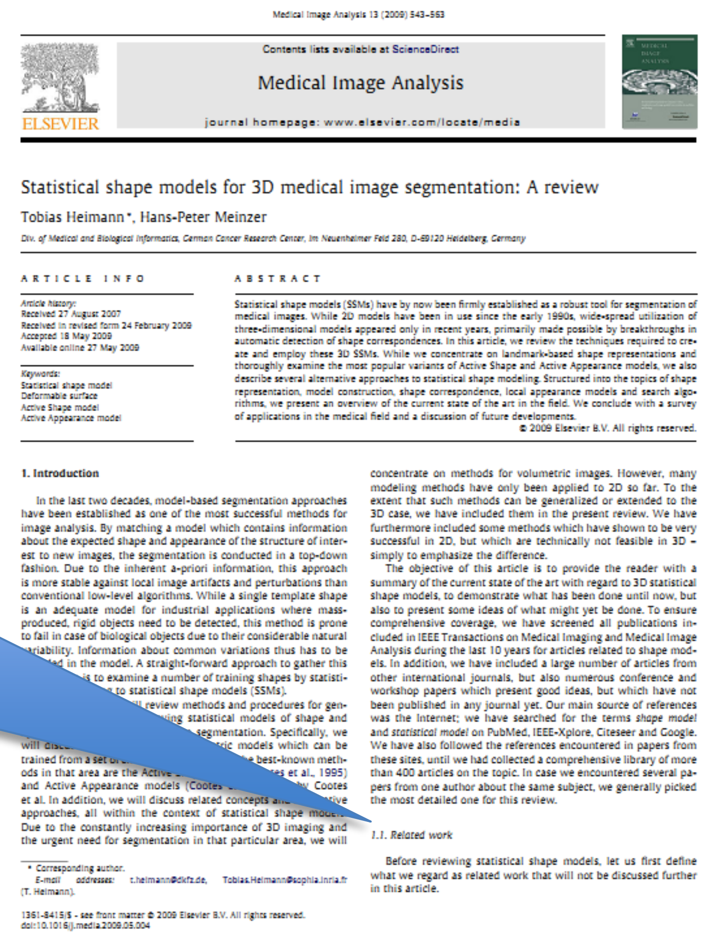
$$\mathcal{J}_{\text{MDS}} = \sum_{i,j} \left(\boldsymbol{x}_i \cdot \boldsymbol{x}_j - \boldsymbol{y}_i \cdot \boldsymbol{y}_j \right)^2$$

- Solution given by spectral decomposition of the Gram matrix

$$G_{i,j} = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$$

Applications

- Statistical shape models for 3D medical image segmentation


Medical Image Analysis 13 (2009) 543–563

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Statistical shape models for 3D medical image segmentation: A review

Tobias Heimann *, Hans-Peter Meinzer

Div. of Medical and Biological Informatics, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 27 August 2007
Received in revised form 24 February 2009
Accepted 18 May 2009
Available online 27 May 2009

Keywords:
Statistical shape model
Deformable surface
Active Shape model
Active Appearance model

ABSTRACT

Statistical shape models (SSMs) have by now been firmly established as a robust tool for segmentation of medical images. While 2D models have been in use since the early 1990s, wide-spread utilization of three-dimensional models appeared only in recent years, primarily made possible by breakthroughs in automatic detection of shape correspondences. In this article, we review the techniques required to create and employ these 3D SSMs. While we concentrate on landmark-based shape representations and thoroughly examine the most popular variants of Active Shape and Active Appearance models, we also describe several alternative approaches to statistical shape modeling. Structured into the topics of shape representation, model construction, shape correspondence, local appearance models and search algorithms, we present an overview of the current state of the art in the field. We conclude with a survey of applications in the medical field and a discussion of future developments.

1. Introduction

In the last two decades, model-based segmentation approaches have been established as one of the most successful methods for image analysis. By matching a model which contains information about the expected shape and appearance of the structure of interest to new images, the segmentation is conducted in a top-down fashion. Due to the inherent a-priori information, this approach is more stable against local image artifacts and perturbations than conventional low-level algorithms. While a single template shape is an adequate model for industrial applications where mass-produced, rigid objects need to be detected, this method is prone to fail in case of biological objects due to their considerable natural variability. Information about common variations thus has to be ... in the model. A straight-forward approach to gather this ... is to examine a number of training shapes by statistical ... to statistical shape models (SSMs). ... review methods and procedures for gen... statistical models of shape and ... segmentation. Specifically, we will disc... models which can be trained from a set ... best-known methods in that area are the Active ... et al., 1995) and Active Appearance models (Cootes ... Cootes et al. In addition, we will discuss related concepts ... approaches, all within the context of statistical shape models. Due to the constantly increasing importance of 3D imaging and the urgent need for segmentation in that particular area, we will concentrate on methods for volumetric images. However, many modeling methods have only been applied to 2D so far. To the extent that such methods can be generalized or extended to the 3D case, we have included them in the present review. We have furthermore included some methods which have shown to be very successful in 2D, but which are technically not feasible in 3D – simply to emphasize the difference.

The objective of this article is to provide the reader with a summary of the current state of the art with regard to 3D statistical shape models, to demonstrate what has been done until now, but also to present some ideas of what might yet be done. To ensure comprehensive coverage, we have screened all publications included in IEEE Transactions on Medical Imaging and Medical Image Analysis during the last 10 years for articles related to shape models. In addition, we have included a large number of articles from other international journals, but also numerous conference and workshop papers which present good ideas, but which have not been published in any journal yet. Our main source of references was the Internet; we have searched for the terms *shape model* and *statistical model* on PubMed, IEEE-Xplore, Citeseer and Google. We have also followed the references encountered in papers from these sites, until we had collected a comprehensive library of more than 400 articles on the topic. In case we encountered several papers from one author about the same subject, we generally picked the most detailed one for this review.

1.1. Related work

Before reviewing statistical shape models, let us first define what we regard as related work that will not be discussed further in this article.

* Corresponding author.
E-mail addresses: t.heimann@dkfz.de, Tobias.Heimann@sophia.inria.fr (T. Heimann).

1361-8415/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2009.05.004

On the one hand, the PCA approach employed by the vast majority of studies unquestionably has its weaknesses: A number of modeled shapes will certainly not feature Gaussian distributions and the linear approximation model will be suboptimal. On the other hand, PCA is fairly robust to the input data distribution and generally just works.

Heimann T, Meinzer HP.
Statistical shape models for 3D medical image segmentation: A review
In: Medical Image Analysis. Vol. 13(4), pp. 543-563, 2009

Applications

- Model-based segmentation

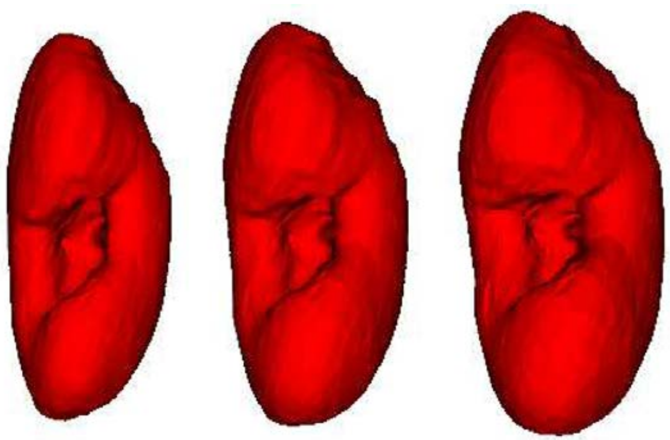


Figure: First mode of variation.

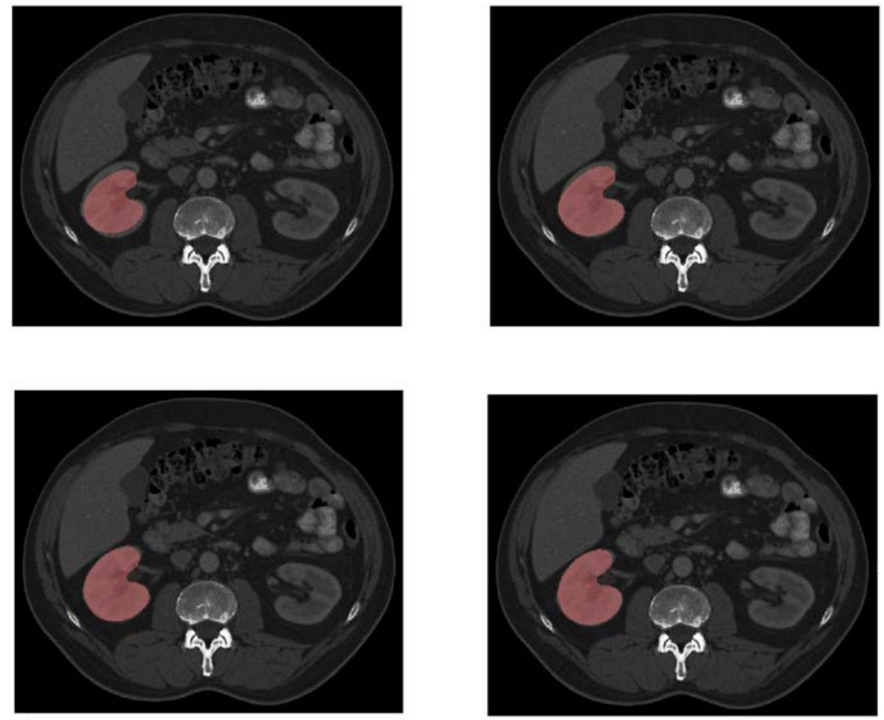


Figure: Kidney segmentation using ASMs.

Spiegel M, Hahn D, Daum V, Wasza J, Hornegger J.
Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration
In: Computerized Medical Imaging and Graphics. Vol. 33, pp. 29-39, 2009

Applications

- 4-D shape priors for respiratory motion management

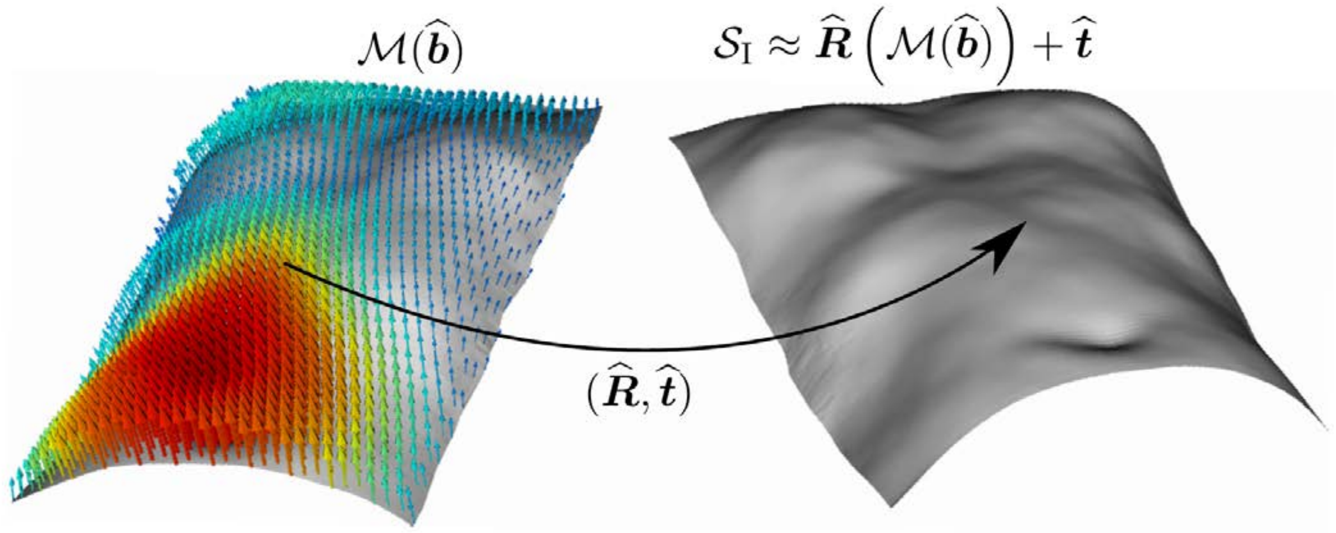


Figure: Motion compensated patient positioning

Wasza J, Bauer S, Hornegger J.
Real-time Motion Compensated Patient Positioning and Non-Rigid Deformation Estimation using 4-D Shape Priors
In: MICCAI 2012, accepted for publication

Applications

- Respiratory model of anatomical motion

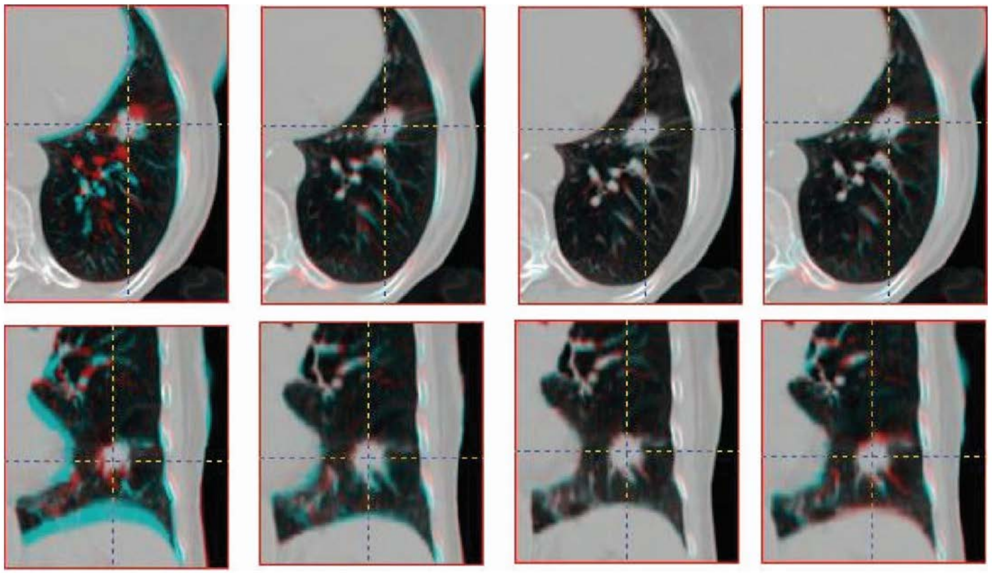


Figure: Actual CT and reconstructed volume.

Zhang Q, Pevsner A, Hertanto A, Hu Y, Rosenzweig K, Ling C, Mageras G.
A patient-specific respiratory model of anatomical motion for radiation treatment planning
In: Medical Physics. Vol. 34(12), pp. 4772-4781, 2007

Non-linear Dimensionality Reduction

- Kernel Principal Component Analysis (KPCA)
- Manifold Learning and Graph-Based Methods
- Applications

Kernel Principal Component Analysis (KPCA)

- Basic idea [4]: compute PCA in a **feature space** $\phi(\boldsymbol{x}_i) \in \mathcal{H}$ instead of the original input space $\boldsymbol{x}_i \in \mathbb{R}^D$
  - In general: $\mathcal{H} = \mathbb{R}^\infty$
  - Kernel trick: $\phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$
- Covariance matrix in feature space $\mathcal{H}$

$$\boldsymbol{C} = \frac{1}{n} \sum_i \phi(\boldsymbol{x}_i) \phi(\boldsymbol{x}_i)^\top$$

- This is equivalent to non-linear MDS
  - Kernel matrix contains generalized inner products in feature space
  - Spectral decomposition of the Gram matrix in feature space

Kernel Principal Component Analysis (KPCA)

- Eigendecomposition of the kernel matrix $\boldsymbol{K}_{i,j} = k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$

$$\boldsymbol{K} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^\top = \boldsymbol{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top$$

$$\boldsymbol{E} = \left[\boldsymbol{e}_i, \ldots, \boldsymbol{e}_n\right] \ , \ \boldsymbol{\Lambda} = diag\left(\lambda_1, \ldots, \lambda_n\right)$$

- Note that the number of modes equals the number of samples! This is in contrast to conventional PCA.

- Projection on the p-th Eigenvector $\boldsymbol{V}^p$

$$\boldsymbol{y}_i^p = \boldsymbol{V}^p \cdot \phi\left(\boldsymbol{x}_i\right) = \sqrt{\lambda_p}\boldsymbol{e}_i^p$$

- Note that we assume the data to be mean-centered in feature space. This in general not valid, for details see [4].

Kernel Principal Component Analysis (KPCA)

- *Out-of-sample* problem
  - Not as easy as with conventional PCA
  - Related to the Nyström extension [5]

$$\widehat{\boldsymbol{y}}^p = \boldsymbol{V}^p \cdot \phi\left(\widehat{\boldsymbol{x}}\right) = \frac{1}{\sqrt{\lambda_p}} \sum_{i=1}^{n} \boldsymbol{e}_i^p \left(\phi\left(\boldsymbol{x}_i\right) \cdot \phi\left(\widehat{\boldsymbol{x}}\right)\right) = \frac{1}{\sqrt{\lambda_p}} \sum_{i=1}^{n} \boldsymbol{e}_i^p k\left(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}\right)$$

- *Pre-image* problem
  - In feature space like regular PCA
  - From feature space to input space an ill-posed problem
  - Usually approximated by minimizing an error function [5]

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\|\boldsymbol{V}\phi\left(\boldsymbol{x}\right) - \widehat{\boldsymbol{y}}\right\|^2$$

Kernel Principal Component Analysis (KPCA)

- Choosing the kernel function
  - Radial basis functions (RBF)
  - Polynomial kernels
  - Linear functions
  - See also Kernel-SVM and Kernel-SVR

- Parameters
  - Highly depend on the problem
  - Slightly different parameters usually produce complete different results

There is no rule which kernel to choose and how to select its parameters!

Kernel Principal Component Analysis (KPCA)

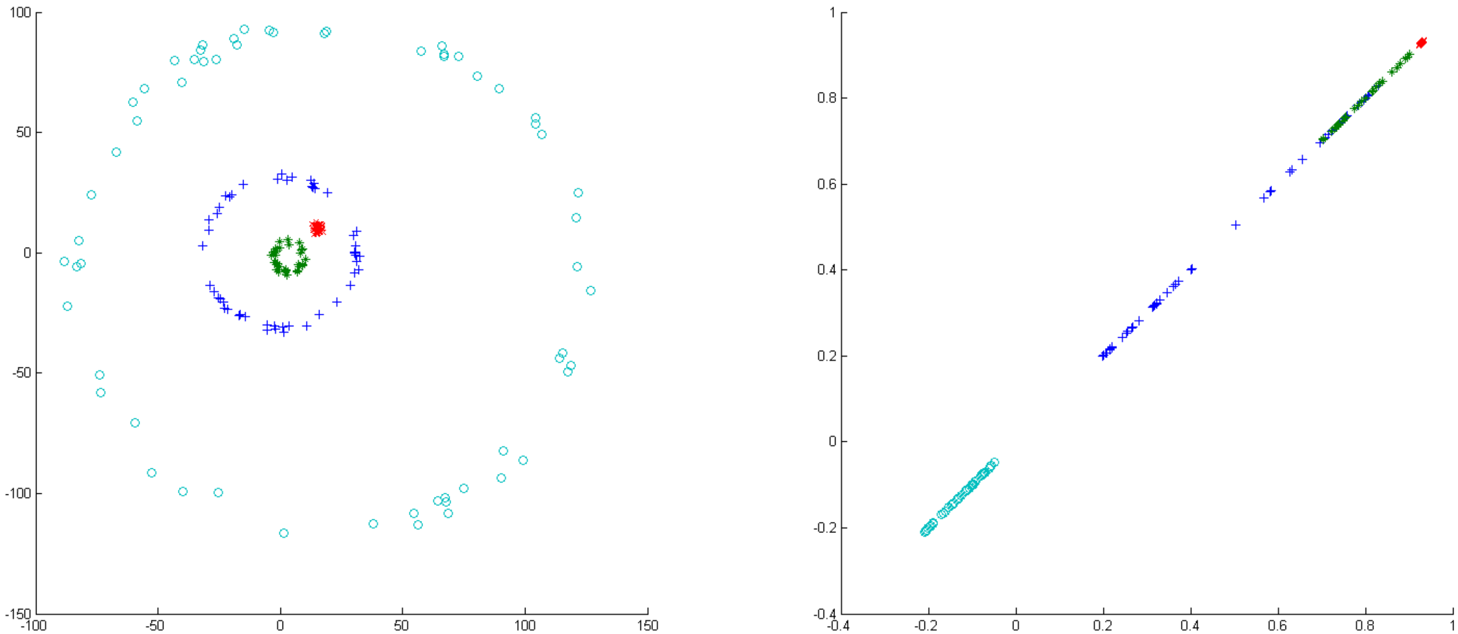


Figure: Dimensionality reduction using KPCA.

Manifold Learning and Graph-Based Methods

- Basic idea: duplicate the behavior of PCA on **manifolds** instead of linear subspaces
- Manifolds:
  - Low-dimensional structure embedded in a high-dimensional space
  - Geodesic distances vs. euclidean distances

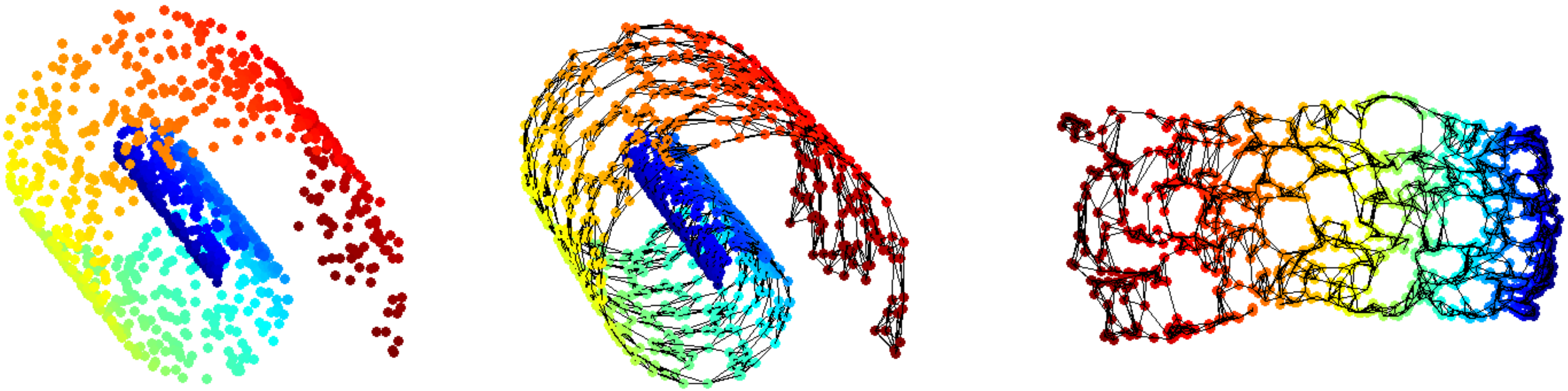


Figure: The swissroll manifold. Left: original data. Middle: connected data. Right: unfolded.

Manifold Learning and Graph-Based Methods

- Multitude of approaches (see [6,7] for overviews)
  - Isomap
  - Locally Linear Embedding
  - Laplacian Eigenmaps
  - Maximum Variance Unfolding
  - ...
- Key issues
  - Discrete manifold approximations (graphs)
  - Geodesic distances (shortest paths)

Manifold Learning: Isomap

- Idea: Preserve distances between input patterns **as measured along the manifold from which they were sampled** → Geodesic distances

- Estimation of geodesic distances
  - *k*-nearest neighbor graph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$

    $$\mathcal{V} = \{\boldsymbol{x}_i\}_{i=1}^{n} \ , \ \mathcal{E}_k = \left\{ \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \mid \boldsymbol{x}_j \in \mathcal{N}^k\left(\boldsymbol{x}_i\right) \right\}$$

  - Geodesic distance matrix

    $$D_{i,j} = \mathrm{ShortestPath}\left(\mathcal{G}^k, \boldsymbol{x}_i, \boldsymbol{x}_j\right)$$

- Classical MDS on the geodesic distances

  $$\mathcal{J}_{\mathrm{IsoMDS}} = \sum_{i,j} \left(D_{i,j} - \boldsymbol{y}_i \cdot \boldsymbol{y}_j\right)^2$$

Manifold Learning and Graph-Based Methods

- *Out-of-sample* problem
  - Depends on the method
  - Often related to the Nyström extension
- *Pre-image* problem
  - Depends on the method
  - Closely related to the *out-of-sample* problem
- Common problems
  - All algorithms require a neighborhood size
  - Theoretical performance results often not available
  - Target dimensionality must be specified in general
  - Do real world data exhibit manifold structures?

Applications

- Image-based breathing gating

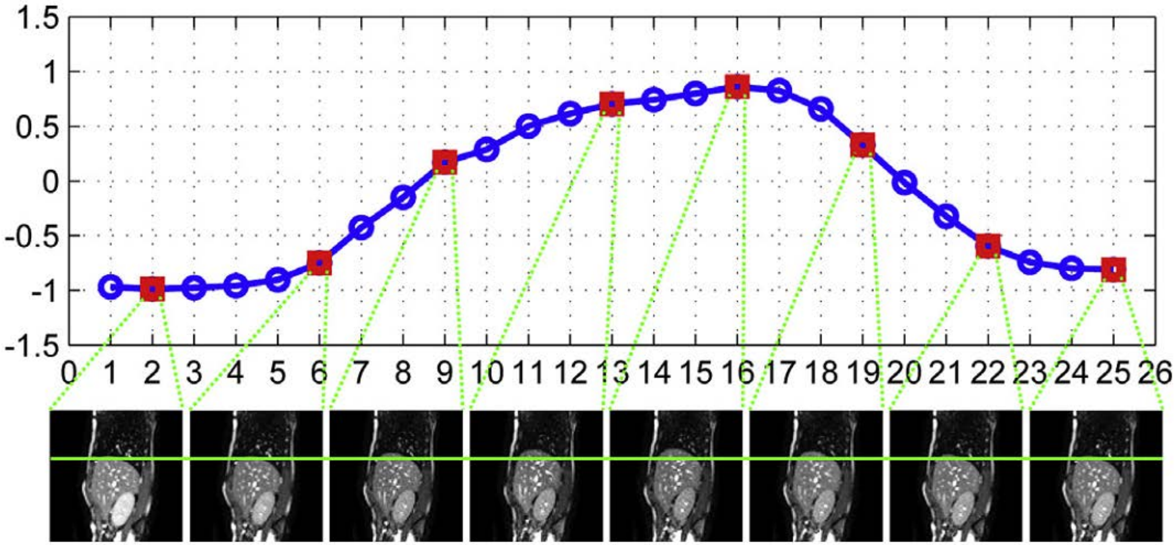


Figure: Image-based breathing signal obtained by manifold learning.

Wachinger C, Yigitsoy M, Rijkhorst E, Navab N.
Manifold learning for image-based breathing gating in ultrasound and MRI
In: Medical Image Analysis, Vol. 16(4), pp.806-818, 2012

Applications

- Registration, Segmentation and Classification

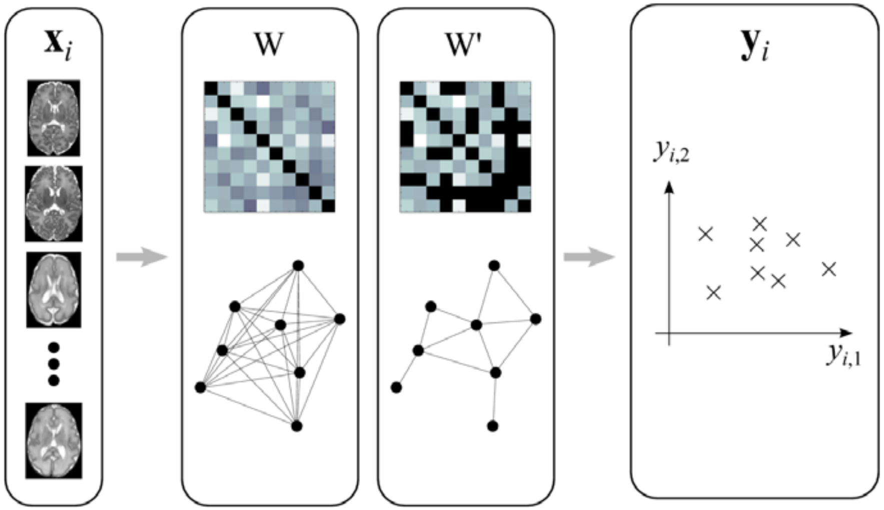


Figure: Schematic overview of manifold learning in medical imaging.

Aljabar, P, Wolz R, Rueckert D.
Manifold Learning for Medical Image Registration, Segmentation and Classification
In: Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis, 2012

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Bibliography

[1] *Stegmann MB, Sjöstrand K, Larsen R*.
Sparse Modeling of Landmark and Texture Variability using the Orthomax Criterion
In: Proc SPIE. Vol. 6144, 2006

[2] *Zou H, Hastie T, Tibshirani R*.
Sparse Principal Component Analysis
In: Journal of Computational and Graphical Statistics. Vol. 15, 2006

[3] *Hein M, Bühler T*.
An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA
In: Advances in Neural Information Processing Systems, 2010

[4] *Schölkopf B, Smola A, Müller K.*
Nonlinear Component Analysis as a Kernel Eigenvalue Problem
In: Neural Computation. Vol 10(5), pp. 1299-1319, 1998

[5] *Arias P, Randall G, Sapiro, G.*
Connecting the Out-of-Sample and Pre-Image Problems in Kernel Methods
In: Proc CVPR, pp. 1-8, 2007

[6] *Cayton L.*
Algorithms for manifold learning
Technical Report CS2008-0923, Univ. of California, 2005

[6] *Saul L, Weinberger K, Sha F, Ham J. Lee, D.*
Spectral Methods for Dimensionality Reduction
In: Semi-Supervised Learning, pp. 293-308, 2006