Multi-resolution multi-object statistical shape models based on the locality assumption

Matthias Wilms*, Heinz Handels, Jan Ehrhardt

Institute of Medical Informatics, University of Lübeck, Germany

Abstract

Statistical shape models learned from a population of previously observed training shapes are nowadays widely used in medical image analysis to aid segmentation or classification. However, providing an appropriate and representative training population of preferably manual segmentations is typically either very labor-intensive or even impossible. Therefore, statistical shape models in practice frequently suffer from the high-dimension-low-sample-size (HDLSS) problem resulting in models with insufficient expressiveness.

In this paper, a novel approach for learning representative multi-resolution multi-object statistical shape models from a small number of training samples that adequately model the variability of each individual object as well as their interrelations is presented. The method is based on the assumption of locality, which means that local shape variations have limited effects in distant areas and, therefore, can be modeled independently. This locality assumption is integrated into the standard statistical shape modeling framework by manipulating the sample covariance matrix (non-zero covariances between distant landmarks are set to zero). To allow for multi-object modeling, a method for computing distances between points located on different object shapes is proposed. Furthermore, different levels of locality are introduced by deriving a multi-resolution scheme, which is equipped with a method to combine variability information modeled at different levels into a single shape model. This combined representation of global and local variability in a single shape model allows the use of the classical active shape model strategy for model-based image segmentation.

An extensive evaluation based on a public data base of 247 chest radiographs is performed to show the modeling and segmentation capabilities of the proposed approach in single- and multi-object HDLSS scenarios. The new approach is not only compared to the classical shape modeling method but also to three state-of-the-art shape modeling approaches specifically designed to cope with the HDLSS problem. The results show that the new approach significantly outperforms all other approaches in terms of generalization ability and model-based segmentation accuracy.

Keywords: statistical shape models, multi-resolution, multi-object segmentation, high-dimension-low-sample-size problem

1. Introduction

When applied to segmentation tasks in medical imaging, typical low-level segmentation algorithms that only explore local image information (e.g. gray values or image gradients) often fail to produce reasonably accurate results due to factors such as image noise, modality-related image artifacts, neighboring structures exhibiting similar gray value characteristics, or the presence of pathologies that all lead to missing or misleading boundary information and, hence, prohibit sufficient success of basic algorithms. To overcome this problem, model-based segmentation approaches are now widely used in medical image analysis, which aim at robustifying the segmentation process by incorporation of a-priori information about the object of interest's shape (and appearance).

Starting with the introduction of simple shape matching approaches in the early 1980s, model-based segmentation has been an active field of research and various approaches utilizing different types of prior information, modeling strategies and ways to use the prior information during segmentation have been developed (Jain et al., 1998). One of the most influential approaches is the active shape model (ASM) method proposed by Cootes et al. (1995). The key component of the ASM method is a statistical shape model (SSM) that compactly represents the space of plausible shapes of an object class (shape space) by a mean shape and a variability model learned from a population of class instances provided as training data. During segmentation, the SSM is employed to constrain the result to plausible shapes. Throughout the years, ASM-like methods and other methods based on SSMs have been successfully used to solve many different segmentation tasks arising in medical imaging (see (Heimann and Meinzer, 2009) for an overview).

When building a statistical shape model based on a given population of training shapes, one first needs to decide on a

*Corresponding author. Universität zu Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

Email address: wilms@imi.uni-luebeck.de (Matthias Wilms)

This is a preprint version of the MedIA article, which (slightly) differs from the final publication available at Elsevier via http://dx.doi.org/10.1016/j.media.2017.02.003

proper representation of the data. A simple but yet generic, flexible, and widely used idea is to represent a shape as a set of points (landmarks) distributed across its surface. Statistical shape models utilizing such a point-based shape representation are usually called point distribution models (PDMs) (Cootes et al., 1992). In this framework, a SSM is then build by (1) representing the points of each training shape in a corresponding shape vector, (2) computing a mean shape, and (3) performing a principal components analysis (PCA) by computing the eigenvectors of the sample covariance matrix of all vectorized training shapes to obtain a compact model that best describes the shape variability observed in the training data.

The basic point-based modeling approach proposed by Cootes et al. is not only applicable to single-object shape modeling (e.g., one organ) but can also be used to build multi-object shape models, which capture the variability of multiple individual structures and their spatial relations. This can be achieved without any modifications to the modeling scheme, shape vectors of different objects only need to be concatenated. Those multi-object models can then be employed to simultaneously segment the individual structures. Compared to the use of several single-object shape models to successively segment different but related structures, these multi-object shape models have the potential to enhance the robustness of the segmentation process as the relations between the structures are encoded in the model and, thus, can provide additional valuable information to guide the segmentation. Multi-object shape models built by employing this technique have been successfully used in several applications such as the segmentation of multiple structures in thoracic radiographs (van Ginneken et al., 2006), neuroanatomic structures in brain MRI (Duta and Sonka, 1998), and heart structures in cardiac MRI (Lötjönen et al., 2004).

To sum up, a SSM provides a compact representation of the shape space of an object/multiple objects of interest derived from a population of observed training instances, which can subsequently be used as prior information to guide segmentation algorithms. The quality of the resulting model, i.e. the correctness of the learned shape space, is related to the quantity and quality of the training data used.

In medical applications, collecting an adequately large and representative training population of (preferably) manual segmentations is often laborious and challenging, particularly if dimensionality (= number of points) and complexity of the observed objects increase – something which is especially true when it comes to multi-object modeling. Therefore, SSMs typically suffer from the so-called high-dimension-low-sample-size (HDLSS) problem: Many landmarks are required to accurately describe the shapes while the number of training samples is limited. For segmentation tasks, this results in a limited flexibility of the model and details can not be represented adequately because the model is over-constrained.

1.1. Related work

Over the years, several approaches to cope with the HDLSS problem in statistical shape modeling have been proposed. In (Cootes and Taylor, 1995), additional variability is introduced by augmenting the sample covariance matrix with information from vibrational modes of simple finite element models (FEMs) of the training shapes that describe physically motivated deformations of the data available. In addition to the question whether these FEM-related deformations are characteristic for the biological variability of a specific object of interest, the choice of the parameter used to control the amount of FEM-related variability added to the model is quite important but non-trivial. Koikkalainen et al. (2008) present several other techniques for an artificial enlargement of the training set such as a non-rigid movement algorithm, which randomly applies smooth local deformations to the training shapes. Their key finding is that both the FEM-based approach of Cootes and Taylor (1995) and their random non-rigid movement technique outperform all other approaches tested and give comparable results in terms of segmentation accuracy on cardiac data.

Zhao et al. (2005) presented a patch-based approach for shape modeling based on a small number of training shapes. In their approach, shapes are subdivided to model different parts of an object independently to allow for a higher flexibility. However, their approach does not generate a consistent model and thus the consistency needs to be enforced during model application.

Davatzikos et al. (2003) presented one of the first hierarchical shape modeling approaches. Their approach builds on the wavelet transform that is employed to decompose 2D shapes into a hierarchy of different levels of details represented by different numbers of bands. Finally, a model is build independently for each band to reduce the effects of the HDLSS problem. In its initial formulation, this approach is only suitable for single-object shape modeling but it was recently generalized to the multi-object case by Cerrolaza et al. (2011). Furthermore, Nain et al. (2007) presented an extension of the single-object approach to the 3D case and proposed a way to optimize the bands by spectral partitioning. As pointed out in (Cerrolaza et al., 2012), the original approach of Davatzikos et al. (2003) has two main drawbacks: the independent modeling of each band might offer too much flexibility and the generic decomposition scheme does not provide a way to explicitly account for inter-object relations. These drawbacks are explicitly addressed by Cerrolaza et al. (2012, 2015) who also use the wavelet transform but (automatically) perform a decomposition of the multi-object data into a hierarchy of objects/sub-objects to explicitly characterize inter- and intra-object relations and build shape models at each scale. From a practical point of view, all of these wavelet-based approaches require an additional (complex) wavelet-based shape decomposition/synthesis step to be integrated into the segmentation algorithm, which can be considered as a general disadvantage of them.

1.2. Contributions

In this paper, we present a novel approach for learning representative multi-object statistical shape models from a small number of training samples that adequately model the variability of each individual object as well as their interrelations. The method is based on the assumption of locality, i.e. we assume that local shape variations have limited effects in distant areas

and, hence, can be modeled independently. This locality assumption is integrated into the standard statistical shape modeling framework by manipulating the sample covariance matrix (covariances between distant landmarks are set to zero). To allow for multi-obbject modeling, a method for computing distances between points located on different object shapes is proposed. Furthermore, different levels of locality are introduced by deriving a multi-resolution scheme, which is equipped with a method to combine variability information modeled at different levels into a single shape model. This combined representation of global and local variability in a single shape model allows the use of the classical ASM strategy for segmentation. To show its modeling and segmentation capabilities, this novel approach is extensively evaluated on a public data base of chest radiographs.

2. Methods

This section starts by formally introducing the classical point-based statistical shape modeling framework conceived by Cootes et al. (1995). Based on this framework, our novel multi-resolution, multi-object shape modeling approach is derived starting with the explanation of the locality assumption in Sec. 2.2, followed by the multi-resolution strategy in Sec. 2.3 and the multi-object extension in Sec. 2.4.

2.1. Statistical shape models

From here on, we assume a set $\{s_i\}_{i=1}^N$ of N d-dimensional ($d = 2$ or 3) training shapes s_i to be given. Each shape $s_i = \{\vec{x}_{1,i}, \ldots, \vec{x}_{j,i}, \ldots, \vec{x}_{M,i}\}$ is represented by a set of M points (landmarks) $\vec{x}_{j,i} = (x_{1,j,i}, \ldots, x_{d,j,i})^T \in \mathbb{R}^d$ distributed across its surface. To allow for a meaningful statistical analysis, these landmarks have to be in correspondence across the training population and all shapes have to be aligned in a common coordinate system, which can be achieved with a multitude of (registration) methods (see (Heimann and Meinzer, 2009) for an overview). Then, the goal is to build a SSM consisting of a mean shape and a compact variability model that adequately approximates the shape space (a subspace of $\mathbb{R}^{dM}$) of the object of interest by performing a PCA of the aligned training data.

First, each training shape s_i is transformed into a vectorial representation

$$\vec{s}_i = (\vec{x}_{1,i}^T, \ldots, \vec{x}_{j,i}^T, \ldots, \vec{x}_{M,i}^T)^T \,, \tag{1}$$

by concatenating the coordinates of each landmark. Then, the statistical analysis starts by computing the mean shape

$$\vec{\mu} = \frac{1}{N} \sum_{i=1}^{N} \vec{s}_i \,, \tag{2}$$

and the $dM \times dM$ sample covariance matrix

$$\mathbf{C} = \frac{1}{N-1} \sum_{i=1}^{N} (\vec{s}_i - \vec{\mu})(\vec{s}_i - \vec{\mu})^T \,, \tag{3}$$

followed by an eigendecomposition $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ where the columns of $\mathbf{U} \in \mathbb{R}^{dM \times dM}$ are eigenvectors of $\mathbf{C}$ and the diagonal elements of diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{dM \times dM}$ denote the corresponding eigenvalues. Here, an eigenvector $\vec{u}_i$ represents a mode of variation of the training shapes and the associated eigenvalue λ_i describes the variance of the training data along its direction.

Aiming at a compact model, the dimensionality of the problem is then reduced by keeping only the eigenvectors associated with the k largest eigenvalues (= most important variation modes) with $\sum_{j=1}^{k} \lambda_j \geq f \sum_{j=1}^{dM} \lambda_i$ (assuming $\lambda_j \geq \lambda_{j+1}, \forall j$), typically with $f \in [0.95, 0.98]$, which corresponds to the portion of the variability in the training set to be retained. Note that for $N < dM$ the rank of the covariance matrix is rank($\mathbf{C}$) $< N$ and that at most $N - 1$ non-zero eigenvalues exist. Finally, the selected eigenvectors are concatenated in a matrix $\mathbf{P} = (\vec{u}_1|\vec{u}_2|\ldots|\vec{u}_k) \in \mathbb{R}^{dM \times k}$ whose columns form an orthonormal basis of the approximated shape space. Instances of this shape space can then be approximated by

$$\hat{\vec{s}} = \vec{\mu} + \mathbf{P}\vec{b} \,, \tag{4}$$

where $\vec{b} \in \mathbb{R}^k$ holds the shape parameters. For a given shape $\vec{s}$, parameters optimally approximating the shape in a least-squares sense can be computed by

$$\vec{b} = \mathbf{P}^T(\vec{s} - \vec{\mu}) \,. \tag{5}$$

To allow only the generation of plausible shapes in Eq. (4), typically the shape parameters are constrained to lie in a hyper-rectangle

$$b_j \leq \alpha \sqrt{\lambda_j} \,, \tag{6}$$

with α between 1 and 3.

An active shape model (Cootes et al., 1995) combines a SSM with a local appearance model that are together used in an iterative algorithm that fits the shape model to the image. During segmentation, the iterative algorithm alternates between the search for new optimal image locations for each landmark (candidate points) of a shape estimated in the last iteration and constraining the shape formed by the new candidate points to the space of plausible shapes by applying Eq. (5) and Eq. (4). Locally optimal image locations for each landmarks are found by matching a landmark-specific appearance model learned from the training samples to the image. Traditionally, this appearance model is based on the normalized image gradient along a 1D profile orthogonal to the shape surface at the landmark location (see (Cootes et al., 1995, 2000) for details).

2.2. Locality assumption

In medical image analysis, SSMs oftentimes suffer from the HDLSS problem because the number of available training shapes is small compared to the shape dimensionality ($N \ll dM$). This hinders the adequate learning of the underlying shape space and leads to a reduced flexibility of the generated model. This problem is illustrated in Fig. 1 based on 2D hand

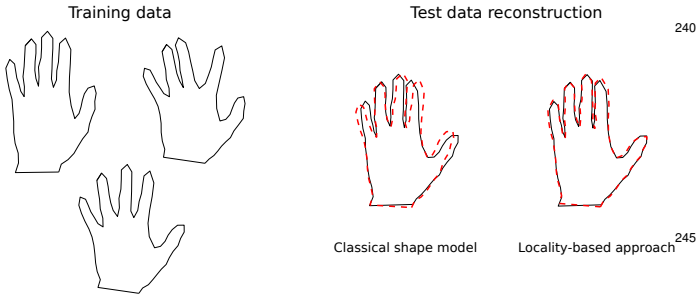


Figure 1: Exemplary illustration of the HDLSS problem when using a classical SSM compared to a model built by using the proposed locality-based approach. Based on a training set of 3 hand contours, a classical SSM and a locality-based model with 4 levels are trained and used to approximate an unseen test contour with a combination of finger orientations not represented in the training set. The approximations generated by the models are shown in red (dashed contour).

contours consisting of 72 landmarks made available by Timothy Cootes[1]. This initial example clearly shows the inability of the classical shape model to adequately represent the test shape. This is mainly a result of the *global* relation between pinky and thumb orientations indicated by the training shapes and learned by the shape model. In contrast, our new approach is guided by the assumption of *locality*, i.e. we assume that local shape variations have only limited influence in distant areas. A model built on the same training shapes using our locality-based approach is able to generate a much better approximation of the test shape in Fig. 1.

In the standard shape modeling framework described in Sec. 2.1, the degree of (linear) relation between two landmarks $\vec{x}_i$ and $\vec{x}_j$ is encoded in the associated part $\mathbf{C}_{i,j} = \mathrm{cov}(\vec{x}_i, \vec{x}_j)$ of the sample covariance matrix $\mathbf{C}$ (cf. Eq. (3)). In the following, we will explain our modeling approach with a slight abuse of notation in terms of correlations $\frac{\mathrm{cov}(\vec{x}_i,\vec{x}_j)}{\sigma_i\sigma_j}$ (with $\sigma_i^2 = \mathbf{C}_{i,i}$) between landmarks instead of covariances. This will allow us to easily preserve the original variances in Eq. (9) and will become clear later.

We assume that the correlation between distant landmarks $\vec{x}_i$ and $\vec{x}_j$ is zero and introduce this assumption into the standard shape modeling framework. The correlation of two landmarks $\vec{x}_i$ and $\vec{x}_j$ is then estimated by:

$$\check{\rho}_{i,j} = \begin{cases} \frac{\mathrm{cov}(\vec{x}_i,\vec{x}_j)}{\sigma_i\sigma_j} & \text{if } d_{geo}(\vec{x}_i,\vec{x}_j) \le \tau, \\ 0 & else. \end{cases} \tag{7}$$

Note that for $d > 1$, $\check{\rho}_{i,j}$ as well as $\mathrm{cov}(\vec{x}_i, \vec{x}_j)$, σ_i and σ_j are $d \times d$ sub-matrices associated with the d-dimensional point coordinates. To compute Eq. (7), we have to select a threshold τ and to define a distance $d_{geo}(\cdot,\cdot)$ between landmarks on the surface. A natural choice is to use the geodesic distance on the mean shape of the training set, which can be easily approximated for piecewise linear shapes, e.g. triangulated surfaces or polylines.

[1] http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/hand_data.html

All estimated matrices $\check{\rho}_{i,j}$ resulting from Eq.(7) are then organized in a matrix $\check{\mathbf{R}} = \begin{pmatrix} \check{\rho}_{1,1} & \cdots & \check{\rho}_{1,M} \\ \vdots & \ddots & \vdots \\ \check{\rho}_{M,1} & \cdots & \check{\rho}_{M,M} \end{pmatrix}$ whose parts related to landmarks close to each other ($d_{geo} \le \tau$) equal the corresponding sub-matrices of the original correlation matrix associated with $\mathbf{C}$ while all other elements are zero. Furthermore, matrix $\check{\mathbf{R}}$ exhibits a band-like structure, and for $N \ll dM$ and a sufficient small threshold τ, $\mathrm{rank}(\check{\mathbf{R}}) > \mathrm{rank}(\mathbf{C})$. Note that $\check{\mathbf{R}}$ is symmetric because $\check{\rho}_{i,j} = \check{\rho}_{j,i}$ but not necessarily positive semi-definite. In this case, we project $\check{\mathbf{R}}$ to the nearest positive semi-definite correlation matrix

$$\begin{aligned} \tilde{\mathbf{R}} \quad = \quad & \min_{\mathbf{A}} \|\mathbf{A} - \check{\mathbf{R}}\|_F \\ \text{s. t.} \quad & \det(\mathbf{A}) \ge 0 \\ \text{and} \quad & \mathrm{diag}(\mathbf{A}) = 1 \end{aligned} \tag{8}$$

using the algorithm proposed by Higham (2002).

The local SSM for a given threshold τ is now computed by solving the eigensystem

$$\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{dM} \end{pmatrix} \tilde{\mathbf{R}} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{dM} \end{pmatrix} = \mathbf{U}_\tau \mathbf{\Lambda}_\tau \mathbf{U}_\tau^T, \tag{9}$$

where σ_i are the original standard deviations of the landmark coordinates. For small sample sizes, more than $N-1$ non-zero eigenvalues exist due to $\mathrm{rank}(\tilde{\mathbf{R}}) > \mathrm{rank}(\mathbf{C})$. Selection of the largest eigenvectors, computation of shape parameters and model fitting is performed equivalently to the classical SSM described in Sec. 2.1. The resulting local model is defined by a mean shape and a set of selected eigenvectors and eigenvalues: $\vec{\mu}, \mathbf{P}_\tau, \vec{\lambda}_\tau$.

2.3. Multi-resolution shape modeling

By selecting different distance thresholds τ in Eq.(7), different levels of locality are defined for the resulting model. For $\tau \ge \max_{i,j} d_{geo}(\vec{x}_i, \vec{x}_j)$, a classical global SSM is computed and for $\tau = 0$ all landmarks will be assumed to be uncorrelated (resulting in a useless model). By defining a sequence of thresholds $\tau_1 > \tau_2 > \cdots > \tau_L$, a multi-resolution scheme is defined resulting in a set of shape models $\{\vec{\mu}, \mathbf{P}_1, \ldots, \mathbf{P}_L, \vec{\lambda}_1, \ldots, \vec{\lambda}_L\}$ (see Fig. 2 for an illustration). However, those models are highly dependent and redundant (cf. Fig. 2), and at first it remains unclear how to combine the shape spaces represented by those models into a consistent shape space approximation defined by one orthonormal basis and containing information from all levels. Therefore, in the remainder of this section, a SSM that combines shape information at different resolutions/levels, defined by a sequence of thresholds $\tau_1 > \cdots > \tau_L$, in a single model $\vec{\mu}, \mathbf{P}_{MR}, \vec{\lambda}_{MR}$ is derived. We start by defining distances between shape spaces/subspaces that allow us to compare the information represented by different local models (cf. Sec. 2.3.1). These distance definitions are then used to estimate a joint shape space (cf. Sec. 2.3.2).

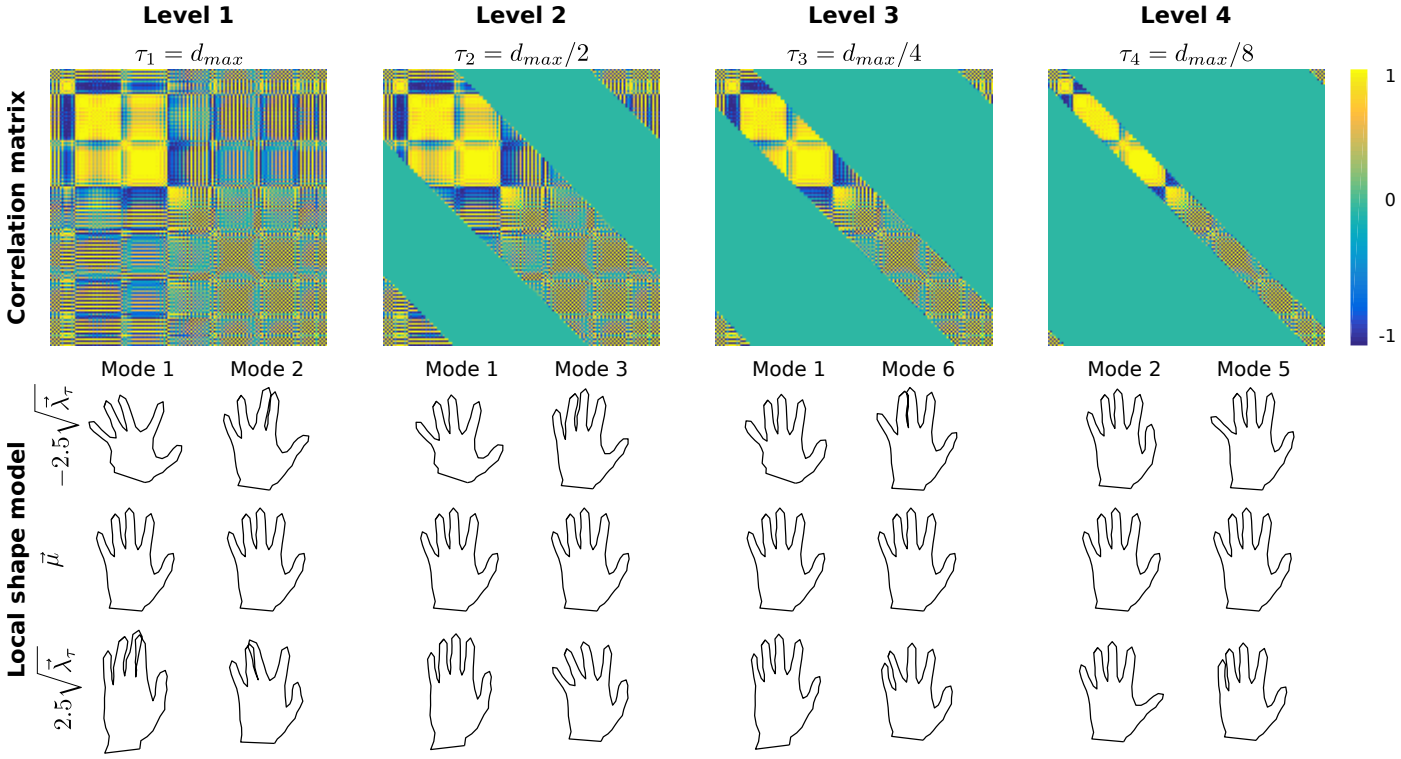


Figure 2: Illustration of thresholded correlation matrices and corresponding local shape models obtained for different distance thresholds τ (levels) when using the 3 training shapes shown in Fig. 1 with $d_{max} = \max_{i,j} d_{geo}(\vec{x}_i, \vec{x}_j)$. For the local shape models two exemplary eigenvectors (modes) of the resulting models are shown, which illustrate the details modelled (global and/or local) at each level and their redundancies. Please note that the model for level 1 (τ_1) is the classical shape model and only consists of 2 eigenvectors. For the first 3 levels, the first modes shown are very similar/redundant and model rather global effects. With decreasing τ, the locality of the effects modeled by the second mode depicted for each level increases. At level 4, (nearly) complete independence of all fingers is achieved.

2.3.1. *Distances between shape spaces*

The shape space spanned by the eigenvectors $\mathbf{P}_{\tau_l}$ of a local model is a subspace of $\mathbb{R}^{dM}$. Let $\mathcal{G}(k,m)$ be the Grassmann manifold (Grassmannian) of linear subspaces of dimension k in $\mathbb{R}^m$. Each point $\mathcal{A} \in \mathcal{G}(k,m)$ can be described by an orthonormal matrix $\mathbf{A} \in \mathbb{R}^{m\times k}$ whose columns form an orthonormal basis of $\mathcal{A} = \text{span}(\mathbf{A})$.

Distance measures on the Grassmannian can be conveniently defined by using the concept of principal angles (Absil et al., 2004). Given two points $\mathcal{A} \in \mathcal{G}(k,m)$ and $\mathcal{B} \in \mathcal{G}(k,m)$, a sequence of k principal angles $0 \le \theta_1 \le \ldots \le \theta_k \le \pi/2$ between these subspaces can be recursively defined as

$$\cos\theta_i = \max_{\vec{q}_i \in \mathcal{A}} \max_{\vec{p}_i \in \mathcal{B}} \vec{q}_i^T \vec{p}_i \tag{10}$$

subject to $\|\vec{q}_i\|_2 = \|\vec{p}_i\|_2 = 1$ and $\|\vec{q}_i^T \vec{q}_j\|_2 = \|\vec{p}_i^T \vec{p}_j\|_2 = 0,\ \forall j \in \{1,\ldots,i-1\}$ (Absil et al., 2004). Based on the associated orthonormal bases $\mathbf{A}$ and $\mathbf{B}$ the principal angles can be efficiently computed using the singular value decomposition

$$\mathbf{A}^T\mathbf{B} = \mathbf{U}(\cos\boldsymbol{\Theta})\mathbf{V}^T\ , \tag{11}$$

with $\boldsymbol{\Theta} = diag(\theta_1,\ldots,\theta_k)$. Given the principal angles, a number of different distances between subspaces can be defined (see (Ye and Lim, 2014) for an overview). For example, the shortest distance between $\mathcal{A}$ and $\mathcal{B}$, measured along the geodesic in $\mathcal{G}(k,m)$ is given by $d_{\mathcal{G}(k,m)} = \|\boldsymbol{\Theta}\|_F$.

Let us now consider two orthonormal matrices $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{m\times l}$ defining subspaces of dimension k and l ($k \le l$), both embedded in $\mathbb{R}^m$. Distances between $\mathcal{A} \in \mathcal{G}(k,m)$ and $\mathcal{B} \in \mathcal{G}(l,m)$ can be defined by finding the k-dimensional subspace contained in $\mathcal{B}$ that is nearest to $\mathcal{A}$, *or* by finding the l-dimensional subspace that contains $\mathcal{A}$ and is nearest to $\mathcal{B}$. It was shown (see e.g. (Ye and Lim, 2014)), that the distances defined by both approaches are equal and define a distance between $\mathcal{A}$ and $\mathcal{B}$. Moreover, an efficient method is given in Alg. 1 to compute the nearest $k-$ and $l-$dimensional subspaces (Ye and Lim, 2014).

Algorithm 1 Computing nearest subspaces

Require: $\mathbf{A} \in \mathbb{R}^{m\times k}$, $\mathbf{B} \in \mathbb{R}^{m\times l}$ with $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{k\times k}$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{l\times l}$, and $k < l$

Perform singular value decomposition:
$[\mathbf{U}(\cos\boldsymbol{\Theta})\mathbf{V}^T] \leftarrow \text{svd}(\mathbf{A}^T\mathbf{B})$
Compute principal vectors between $\mathcal{A}$ and $\mathcal{B}$:
$(\vec{p}_1|\vec{p}_2|\ldots|\vec{p}_k) \leftarrow \mathbf{AU}$
$(\vec{q}_1|\vec{q}_2|\ldots|\vec{q}_l) \leftarrow \mathbf{BV}$
Nearest subspace to $\mathcal{A}$ contained in $\mathcal{B}$:
$\hat{\mathbf{A}} \leftarrow (\vec{q}_1|\vec{q}_2|\ldots|\vec{q}_k)$
Nearest subspace to $\mathcal{B}$ containing $\mathcal{A}$:
$\hat{\mathbf{B}} \leftarrow (\vec{p}_1|\ldots|\vec{p}_k|\vec{q}_{k+1}\ldots|\vec{q}_l)$

Output: orthogonal bases $\hat{\mathbf{A}} \in \mathbb{R}^{m\times k}$, $\hat{\mathbf{B}} \in \mathbb{R}^{m\times l}$

In Alg. 1, $\mathbf{AU}$ and $\mathbf{BV}$ represent rotations of the bases $\mathbf{A}$ and $\mathbf{B}$ to minimize the angles between the first k corresponding vectors.[2] In the next section, the idea of Alg. 1 is used to compute SSMs that combine shape information from different resolutions.

2.3.2. Building multi-resolution shape models

In order to derive the next steps, we recall the following dependency: Given a data matrix $\mathbf{X} \in \mathbb{R}^{m\times k}$ with zero mean and associated covariances $\mathbf{\Sigma_X} = E(\mathbf{X}^T\mathbf{X}) \in \mathbb{R}^{k\times k}$, the covariance of the transformed data matrix $\mathbf{Y} = \mathbf{XM}$ is

$$\mathbf{\Sigma_Y} = \mathbf{M}^T\mathbf{\Sigma_X}\mathbf{M}\,. \tag{12}$$

Furthermore, note that from Eq. (11) follows $\mathbf{AU} = \mathbf{BV}(\cos\mathbf{\Theta})^T$ and the nearest basis $\hat{\mathbf{B}}$ in Alg. 1 is given by

$$\hat{\mathbf{B}} = \mathbf{BVS}^T\,, \tag{13}$$

where $\mathbf{S} = \begin{pmatrix} \cos\mathbf{\Theta}_{k\times k} & 0 \\ 0 & \mathbf{I}_{(l-k)\times(l-k)} \end{pmatrix} \in \mathbb{R}^{l\times l}$ is a diagonal matrix with k elements containing the cosine of the principal angles while the remaining $k-l$ diagonal elements are filled with ones.

We now use the derived methodology to combine two local shape models $\mathbf{P}_{\tau_1}$ and $\mathbf{P}_{\tau_2}$ with eigenvalues $\vec{\lambda}_{\tau_1}$ and $\vec{\lambda}_{\tau_2}$ and $\text{rank}(\mathbf{P}_{\tau_1}) = k < l = \text{rank}(\mathbf{P}_{\tau_2})$ into a single model in way that their redundancy is removed while preserving the characteristic unique information contained in each model as much as possible. The general idea is to use Alg.1 to compute a basis $\hat{\mathbf{P}}_{\tau_1,\tau_2}$ of rank l that fully contains the shape space spanned by the model $\mathbf{P}_{\tau_1}$ from level 1 and is nearest to the shape space described by $\mathbf{P}_{\tau_2}$, the model estimated at level 2.

Following Alg.1, we first perform a singular value decomposition of $\mathbf{P}_{\tau_1}^T\mathbf{P}_{\tau_2}$ to obtain left- and right-singular vectors $\mathbf{U} \in \mathbb{R}^{k\times k}$ and $\mathbf{V} \in \mathbb{R}^{l\times l}$, and the cosines of the principal angles contained in $\mathbf{S}$ (cf. (Eq. 13)). The combined basis $\hat{\mathbf{P}}_{\tau_1,\tau_2}$ is then computed using Eq. (13).

Based on the transformed covariance matrices of both local shape models, $\hat{\mathbf{\Sigma}}_{\tau_1} = \left(\sigma_{i,j}^{\tau_1}\right) = \mathbf{U}^T\begin{pmatrix} \lambda_{\tau_1,1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{\tau_1,k} \end{pmatrix}\mathbf{U}$ and $\hat{\mathbf{\Sigma}}_{\tau_2} = \left(\sigma_{i,j}^{\tau_2}\right) = \mathbf{V}^T\begin{pmatrix} \lambda_{\tau_2,1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{\tau_2,l} \end{pmatrix}\mathbf{V}$, and Eq.(12), the covariances of the basis vectors in $\hat{\mathbf{P}}_{\tau_1,\tau_2}$ are obtained as $\hat{\mathbf{\Sigma}}_{\tau_1,\tau_2} = \left(\sigma_{i,j}^{\tau_1,\tau_2}\right) \in \mathbb{R}^{l\times l}$ with

$$\sigma_{i,j}^{\tau_1,\tau_2} = \begin{cases} \sigma_{i,j}^{\tau_1} & \text{if } i,j \in [1,k]\,, \\ \sigma_{i,j}^{\tau_2} & \text{if } i,j \in [k+1,l]\,, \\ 0 & \text{else}\,. \end{cases}$$

Here, $\sigma_{i,j}^{\tau_1}$, $\sigma_{i,j}^{\tau_2}$, and $\sigma_{i,j}^{\tau_1,\tau_2}$ denote the scalar elements of the covariance matrices. Note, that due to the transformations applied, $\hat{\mathbf{\Sigma}}_{\tau_1,\tau_2}$ is not diagonal in general, thus the basis vectors are correlated. To find uncorrelated basis vectors the eigensystem of $\hat{\mathbf{P}}_{\tau_1,\tau_2}\hat{\mathbf{\Sigma}}_{\tau_1,\tau_2}\hat{\mathbf{P}}_{\tau_1,\tau_2}^T$ is computed. By iteratively repeating this process a shape model combining several multi-resolution levels can be computed. The complete model generation process is outlined in Alg.2 and an illustrative example can be found in Fig. 3.

[2]Note, that the principal angles θ_i are zero for $i > k$.

Algorithm 2 Multi-resolution Shape Models

Require: Data matrix $\mathbf{X} = (\vec{s}_1|\vec{s}_2|\ldots|\vec{s}_N) \in \mathbb{R}^{m\times N}$ and distance thresholds $\tau_1 > \tau_2 > \ldots > \tau_L$.

Compute mean shape: $\vec{\mu} = \frac{1}{N}\sum_{i=1}^{N}\vec{s}_i$

Define distance d_{geo} *on mean shape*

for $r = 1$ to L **do**

Compute local shape model $\mathbf{P}_{\tau_r}$, $\vec{\lambda}_{\tau_r}$ *according to Sec. 2.2.*

if $r > 1$ **then**

Combine shape model:

$\mathbf{U}(\cos\mathbf{\Theta})\mathbf{V}^T \leftarrow \text{svd}(\mathbf{P}_{\text{MR}}^T\mathbf{P}_{\tau_r})$

$\mathbf{S} = \begin{pmatrix} \cos\mathbf{\Theta}_{k\times k} & 0 \\ 0 & \mathbf{I}_{(l-k)\times(l-k)} \end{pmatrix}$

$\hat{\mathbf{B}} = \mathbf{P}_{\tau_r}\mathbf{VS}^T$ *(compute transformed basis)*

Compute transformed covariances:

$\hat{\mathbf{\Sigma}}_{\tau_{\text{MR}}} = \left(\sigma_{i,j}^{\tau_{\text{MR}}}\right) = \mathbf{U}^T\begin{pmatrix} \lambda_{\tau_{\text{MR}},1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{\tau_{\text{MR}},k} \end{pmatrix}\mathbf{U}$

$\hat{\mathbf{\Sigma}}_{\tau_r} = \left(\sigma_{i,j}^{\tau_r}\right) = \mathbf{V}^T\begin{pmatrix} \lambda_{\tau_r,1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{\tau_r,l} \end{pmatrix}\mathbf{V}$

$\hat{\mathbf{\Sigma}}_{\text{MR},r} = \left(\sigma_{i,j}^{\tau_{\text{MR}},\tau_r}\right)$ with

$$\sigma_{i,j}^{\tau_{\text{MR}},\tau_r} = \begin{cases} \sigma_{i,j}^{\tau_{\text{MR}}} & \text{if } i,j \in [1,k]\,, \\ \sigma_{i,j}^{\tau_r} & \text{if } i,j \in [k+1,l]\,, \\ 0 & \text{else}\,. \end{cases}$$

Compute uncorrelated basis vectors and eigenvalues:

$[\mathbf{P}_{\text{MR}}, \vec{\lambda}_{\text{MR}}] \leftarrow \text{eig}(\hat{\mathbf{B}}\hat{\mathbf{\Sigma}}_{\text{MR},r}\hat{\mathbf{B}}^T)$

else

$\mathbf{P}_{\text{MR}} \leftarrow \mathbf{P}_{\tau_r}$, $\vec{\lambda}_{\text{MR}} \leftarrow \vec{\lambda}_{\tau_r}$

end if

end for

Output: Multi-resolution shape model $\vec{\mu}, \mathbf{P}_{\text{MR}}, \vec{\lambda}_{\text{MR}}$

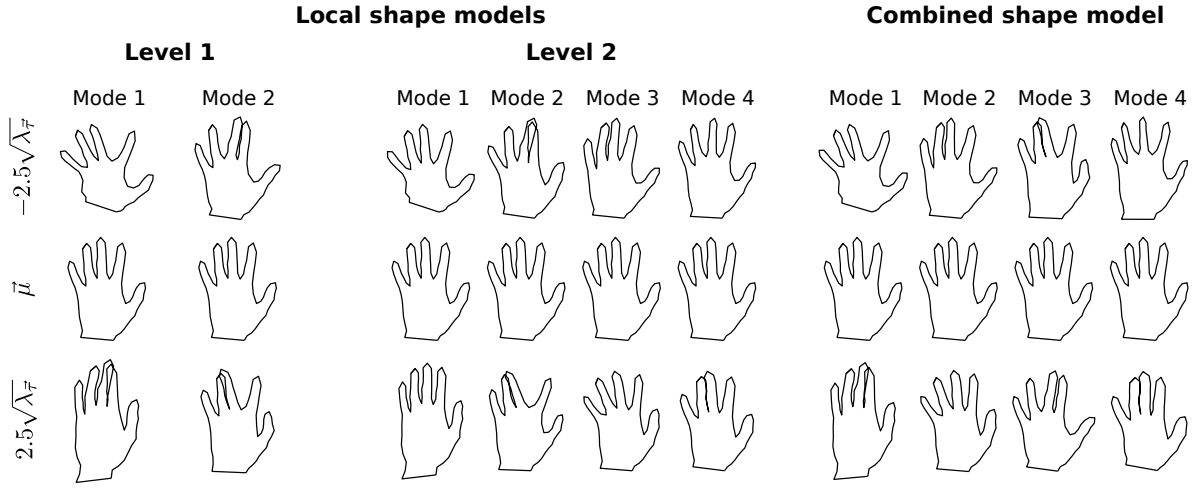


Figure 3: Illustration of the resulting combined multi-resolution shape model resulting from the combination of the local models obtained for distance thresholds τ_1 and τ_2 (cf. Fig. 2) for the training shapes in Fig. 1. Please note, that the model for level 1 (global model) only consists of 2 eigenvectors (modes). For level 2, the eigenvectors corresponding to 95% of the represented variability are shown.

2.4. Multi-object shape modeling

The approach for single-object local shape modeling presented in Sec. 2.2 can be extended to a multi-object modeling approach by computing distances between points located on different object shapes. Therefore, we propose to combine the geodesic distance d_{geo} used in Eq. (7) to compute distances between landmarks on a surface with a modified Euclidean distance. The geodesic distance is infinity for points located on different shapes, whereas the Euclidean distance can be computed between arbitrary points.

Given a shape vector $\vec{s}$ containing landmarks of O objects

$$\vec{s} = (\vec{x}_{1,1}^T, \ldots, \vec{x}_{1,M_1}^T, \vec{x}_{2,1}^T, \ldots, \vec{x}_{2,M_2}^T, \ldots, \vec{x}_{O,1}^T, \ldots, \vec{x}_{O,M_O}^T)^T$$

with M_i the number of landmarks of shape $i \in \{1, \ldots, O\}$. We start by defining an undirected graph $\mathbb{G}_g(\mathbb{V}, \mathbb{E}_g)$ with nodes $\mathbb{V} = \{\vec{x}_{i,j} | i \in \{1, \ldots, O\}, j \in \{1, \ldots, M_i\}\}$ and edges $\mathbb{E}_g = \{(\vec{x}_{i,j}, \vec{x}_{i,k}) | i \in \{1, \ldots, O\}, j \in \{1, \ldots, M_i\}, k \in \mathcal{N}(\vec{x}_{i,j})\}$, where $\mathcal{N}(\vec{x}_{i,j})$ is the direct neighborhood of point $\vec{x}_{i,j}$ on object i, e.g. given by a triangulation or polylines. The weight $w^g_{i,j,k}$ of edge $(\vec{x}_{i,j}, \vec{x}_{i,k})$ is given by the Euclidean distance between the points:

$$w^g_{i,j,k} = \|\vec{x}_{i,j} - \vec{x}_{i,k}\|.$$

The geodesic distance between two points $d_{geo}(\vec{x}, \vec{y})$ on the object surface can then be approximated by the shortest path in $\mathbb{G}_g$. Note, that no edges between different objects exist in $\mathbb{G}_g$ and, therefore, the geodesic distance between points on different objects is infinity.

We define a second undirected graph $\mathbb{G}_e(\mathbb{V}, \mathbb{E}_e)$, which is fully connected with edge weights

$$w^e_{i,j,l,k} = \eta \|\vec{x}_{i,j} - \vec{x}_{l,k}\| + \kappa, \quad \eta, \kappa \in \mathbb{R} \tag{14}$$

representing a scaled and shifted Euclidean distance. The shortest path in $\mathbb{G}_e$ defines a distance $d_e(\vec{x}, \vec{y})$, with $\vec{x}, \vec{y} \in \mathbb{V}$. Clearly, for $\eta \leq 1, \kappa = 0$ holds $d_e(\vec{x}, \vec{y}) \leq d_{geo}(\vec{x}, \vec{y})$, and for $\eta \geq 1, \kappa \geq \max_{i,j,k} d_{geo}(\vec{x}_{i,j}, \vec{x}_{i,k})$ holds $d_e(\vec{x}_{i,j}, \vec{x}_{i,k}) > d_{geo}(\vec{x}_{i,j}, \vec{x}_{i,k})$ for points on the same object.

To further understand the distance defined in Eq. (14), we can use a physical interpretation: Lets assume the energy needed to travel from $\vec{x}$ to $\vec{y}$ onto the object surface is equal to the distance $d_{geo}(\vec{x}, \vec{y})$. The factor η represents the proportion of energy needed to travel the same distance in the embedding space $\mathbb{R}^d$, thus describing the relative viscosity of the embedding space, and κ represents the energy needed to overcome adhesion forces to leave the object surface.

The combined distance $d(\vec{x}, \vec{y})$ between two points $\vec{x}, \vec{y} \in \mathbb{V}$ is now the path with minimum energy, when traveling either on object surfaces or in the embedding space. We can compute $d(\vec{x}, \vec{y})$ as shortest path in the combined fully connected graph $\mathbb{G}(\mathbb{V}, \mathbb{E})$ with edge weights

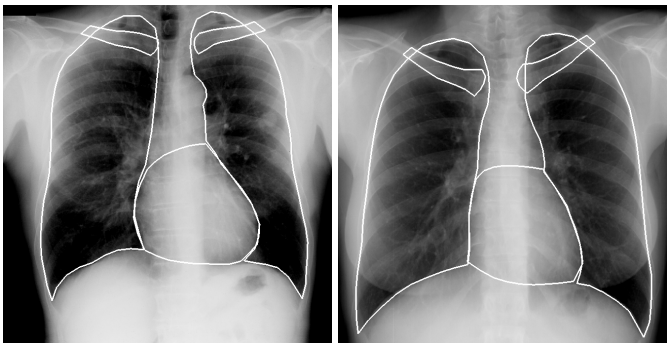
$$w_{i,j,l,k} = \begin{cases} \min(w^g_{i,j,k}, w^e_{i,j,l,k}) & \text{if } (\vec{x}_{i,j}, \vec{x}_{l,k}) \in \mathbb{E}_g \\ w^e_{i,j,l,k} & \text{else.} \end{cases}$$

Note, that the shortest path might combine paths on surfaces and in the embedding space and depends on η and κ. Furthermore, these parameters need to be chosen in accordance with the distance threshold(s) τ because objects are only fully seperated for $\tau < \kappa$ with $\eta \geq 0$.

3. Experiments

In the evaluation part of this paper, the novel shape modeling approach presented in Sec. 2 is extensively evaluated to analyze its capabilities for shape modelling and model-based segmentation. The approach is primarily designed to overcome shortcomings of the classical shape modelling approach in HDLSS scenarios and the goals of our evaluation are therefore:

1. Analysis of the effect of different numbers of training data sets on the *shape model properties* of models generated by the new approach in single- and multi-object scenarios to show its general suitability to tackle the HDLSS problem.



(a) JPCLN001 (b) JPCNN001

Figure 4: Two sample cases from the JSRT data base with corresponding manual segmentations for 5 different anatomical structures (white contours) provided by the SCR data base.

2. Analysis of the effect of different numbers of training data sets on the *segmentation performance* of models generated by the new approach in challenging single- and multi-object scenarios.
3. Systematic comparison of the new approach with the classical SSM method and other state-of-the-art shape modelling approaches specifically designed to cope with the HDLSS problem.

In the following, the data used in the experiments, the other approaches employed for the systematic comparison, the general experimental design, and the parameter selection are detailed.

Data. To allow for reproducibility of the results, all experiments are based on the publicly available JSRT data base (Shiraishi et al., 2000) that contains 247 chest radiographs (2047 $\times$ 2047 pixels; 0.175 mm pixel spacing) with and without lung nodules and the also publicly available SCR data base (van Ginneken et al., 2006) that provides manual segmentations of 5 anatomical structures (right lung, left lung, heart, right clavicle, left clavicle; cf. Fig. 4) for all 247 JSRT cases. The segmentation of these structures is a non-trivial and challenging problem due to the projective nature of the data, fuzzy organ boundaries, and large anatomical variability. These data bases have been widely used to evaluate the performance of different segmentation algorithms including ASM-like methods (e.g. (van Ginneken et al., 2006; Cerrolaza et al., 2011)).

Each manual segmentation of the SCR data base consists of a set of 2D points in correspondence across the whole population. The authors of the data base divided the data (the images and related contours) into two distinct sets of 124 (*fold1*) and 123 cases (*fold2*) with the same proportion of pathological and healthy cases. For this evaluation, *fold2* is used as training data while *fold1* serves as test data. In a preprocessing step, all 247 contours are aligned in a common coordinate system by using similarity transforms to account for differences in pose and size.

Approaches. In addition to the classical SSM (cf. Sec. 2.1), the performance of the new locality-based SSM (Locality-SSM) presented in this work is compared to three state-of-the-art approaches for shape modelling in HLDSS scenarios. The FEM-SSM approach of Cootes and Taylor (1995) is chosen as a representative of approaches that manipulate the training shapes or augment the resulting sample covariance matrix. From the hierarchical wavelet-based approaches, the original method of Davatzikos et al. (2003) (Wavelet-SSM) and the recent method of Cerrolaza et al. (2012) (MOWavelet-SSM) are selected. Both approaches address different aspects: While the Wavelet-SSM method was developed for single-object scenarios and is not suitable for multi-object modeling, the MOWavelet-SSM method explicitly accounts for the inter-object relations in multi-object scenarios.

Due to the lack of publicly available implementations of these approaches, all approaches employed in this study were implemented in MATLAB by the authors of this paper based on the information provided in the respective original papers. This source code (including code for Locality-SSM), is freely available online[3].

Experimental design. The experiments are designed with respect to the previously defined three goals. In general, the evaluation consists of two different parts: In the first part, the properties (generalization ability, specificity and compactness) of the shape models generated by the different approaches with varying numbers of training data are evaluated by using the contours of the JSRT/SCR data bases. The second part deals with the evaluation of the segmentation performance of the shape models in a real segmentation application. This part is, therefore, based on the contours and the associated images of the JSRT/SCR data bases.

Furthermore, each evaluation part is subdivided into two subparts: a single-object and a multi-object scenario. In the single-object scenario, only the right lung contours of the JSRT/SCR data bases consisting of 44 corresponding landmarks are employed. Conversely, in the multi-object scenario, all 5 anatomical structures with a total of 166 corresponding landmarks (right lung: 44, left lung: 50, heart: 26, right clavicle: 23, left clavicle: 26) are used.

The general design of the experiments for both evaluation parts (model properties and segmentation) is the same: For the different approaches, models are generated for varying numbers $N \in \{5, 10, 15, 20, 30, 40, 70, 120\}$ of available training samples. The N training samples are randomly chosen subsets of the 123 cases in the training set. Identical subsets are used for the different approaches for a training set size N.

In the first part, for each model built, its generalization ability, specificity and compactness (Davies et al., 2002) are computed. The generalization ability describes the ability to model formerly unseen shapes, the specificity indicates the validity of the contours produced by the model, and the compactness is the number of parameters of the model. To compute the generalization ability, for each contour in the test set the mean landmark distance to its closest model instance is measured by projecting the test contour to the subspace spanned by the eigenvectors of the SSM. For estimating the specificity, 1000 random model

[3] https://imi.uni-luebeck.de/multi-resolution-multi-object-statistical-shape-models

instances are generated and the minimum mean landmark distance to one of the available samples in the whole data base is computed. The compactness corresponds to the number of eigenvectors used to represent the variability. For each training set size N, the experiments are repeated 50 times and the resulting measures are averaged to reduce the bias introduced by the random selection process. The Wavelet-SSM is not used for the multi-object scenario as it is only applicable to single-object data.

In the second part, the segmentation performance of models built for varying numbers of available training numbers is assessed. For this purpose, the standard multi-resolution active shape model fitting algorithm of Cootes et al. (1995, 2000) was implemented and all approaches were integrated into this framework as described by their authors. As before, models are generated for varying numbers $N \in \{5, 10, 15, 20, 30, 40, 70, 120\}$ of available training cases and used to segment the right lung/all 5 structures in the chest radiographs belonging to the test set by applying the ASM algorithm. For each model the segmentation performance is quantitatively measured by calculating mean symmetric contour distances between the contours resulting form the ASM algorithm and the ground truth contours provided by SCR data base. For the multi-object scenario, mean symmetric contour distances are individually calculated for each anatomical structure and finally weighted and summed according to their proportion of the full number of landmarks to get an overall mean contour distance. For each training set size N, the experiments are repeated $\{40, 40, 40, 40, 35, 35, 35, 30, 20\}$ times with less repetitions for larger training sets to reduce the computational effort.

Parameter selection. The parameters were chosen based on systematic experiments in which only the training data was utilized and according to the parameters used by the respective authors in their publications .

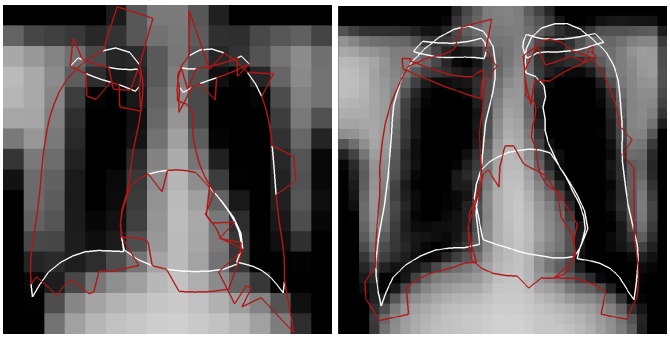
In the following, we list the parameter configurations for all approaches in our experiments to allow for reproducibility of our results. Generally, for all tests and models (and levels therein), the shapes generated by the models where restricted to a hyperrectangle with $\alpha = 2.5$ (cf. Eq. (6)). The remaining parameters for each approach were chosen as follows:

- Locality-SSM: Number of levels/resolutions $L = 5$; variability to be retained at each level $f_L = (0.95, 0.95, 0.95, 0.95, 0.95)$; single-object distance thresholds (in mm, starting with the max. distance in the data that is decreased by a factor of 0.5 at each level) $\tau_L = (957.22, 475.61, 239.30, 119.65, 59.83)$; multi-object distance weights $\eta = 10$, $\kappa = 150$; multi-object distance thresholds (in mm, starting with the max. distance in the data that is decreased by a factor of 0.5 at each level) $\tau_L = (2278.44, 1139.22, 569.61, 284.81, 142.40)$. For the multi-object data, this leads to a separation of all structures at the last level ($\tau_L < \kappa$).
- Classical SSM: Variability to be retained $f = 0.95$
- FEM-SSM: Variability to be retained $f = 0.95$; proportion of FEM variation to include $a = \frac{2500}{N}$ (single-object), $a = \frac{500}{N}$ (single-object segmentation), $a = \frac{35000}{N}$ (multi-object), $a = \frac{10000}{N}$ (multi-object segmentation). See (Cootes and Taylor, 1995) for details on this parameter.
- Wavelet-SSM: Number of levels $L = 4$; Debauchies-7 wavelets are used as proposed in (Davatzikos et al., 2003); variability to be retained at each band $f_B = 0.95$. See (Davatzikos et al., 2003) for details on these parameters.
- MOWavelet-SSM: 5 levels of resolution ($R = 4$); Linear B-Spline wavelets are used as proposed in (Cerrolaza et al., 2012); variability to be retained at each resolution $f_r = (0.95, 0.99, 0.99, 0.99, 0.995)$. For the multi-object scenario, a grouping of the 5 anatomical structures needs to be defined. Here, all structures are grouped individually on the two highest levels, while all structures are modeled together on the two lowest levels. On the remaining level, same side lungs and clavicles are combined and the heart is modeled individually. See (Cerrolaza et al., 2012) for details on these parameters.

With the exception of the FEM-SSM approach, the parameters listed above are used for all experiments (modeling and segmentation performance evaluation). For the FEM-SSM, the weighting parameter a was chosen individually for the modeling and segmentation experiments and each scenario (single-object/multi-object) in a way that the integral of the difference between mean landmarks distances over all training subsets for the classical SSM and the FEM-SSM in the generalization test is maximized.

In addition to the modeling approach-specific parameters, the parameters of the ASM segmentation algorithm employed for the segmentation task need to be chosen. In (van Ginneken et al., 2006), a comparison study of different supervised segmention methods on the JSRT/SCR data base is performed. A key contribution of their work is the investigation of the effects of the different parameters of the ASM segmentation algorithm on the segmentation performance. Therefore, the parameters giving the best results in their work are employed here (ASM tuned; e.g., appearance model: $k = 5$, $L_{Max} = 5$; search algorithm: $n_S = 2$, $N_{max} = 20$, $p_{close} = 1.1$; see (van Ginneken et al., 2006) for details and explanations). However, in contrast to van Ginneken et al. (2006), we do not include pose and size differences into our models and, hence, need to estimate a similarity transform in each iteration of the ASM algorithm. The algorithm is initialized with the mean shape.

Due to the low image resolution at the highest levels (lowest image resultion) of the multi-resolution image pyramid utilized in the ASM algorithm and its rather simplistic gray value profile-based candidate point search, the resulting candidate shapes are noisy (cf. Fig. 5). It turned out that this noise negatively impacts the segmentation performance of the approaches. Therefore, the ASM algorithm was slightly modified in an approach-specific way for the Locality-SSM, FEM-SSM and Wavelet-SSM approaches to improve the segmentation robustness. For the Locality-SSM approach a hierachical fitting scheme was implemented by restricting the number of eigenvectors used for the model fitting step at each of the 5 image



(a) Level 1 (16×16 pixels) (b) Level 2 (32×32 pixels)

Figure 5: The two highest highest levels of the multi-resolution image pyramid of case JPCNN001 depicted in Fig. 4 with exemplary noisy candidate contours obtained by the ASM algorithm when starting from the white contours.

resolution levels to the number of eigenvectors resulting from the combined locality model at this model level (also 5). For the FEM-SSM approach, no explicit global-to-local hierarchy of eigenvectors exist. Such a hierarchy is, therefore, simulated by using a limited number of eigenvectors corresponding to the number of eigenvectors of the classical SSM learned on the same training data for the first 4 levels while the full model is only employed at the lowest pyramid level. For the Wavelet-SSM approach, a Gaussian smoothing of the contour deformation induced by the new candidate points as proposed in (Davatzikos et al., 2003) is performed. Furthermore, a hierarchical scheme was implemented by using the natural hierarchy given by the model by restricting the bands used for the model fitting step at each of the first 3 image resolution levels to the bands at this model level (including bands belonging to lower levels of the model). For the MOWavelet-SSM, no modifications took place as it has shown to be relatively invariant to noise by Cerrolaza et al. (2015) and the complicated wavelet-based shape decomposition does not allow a simple hierarchical adaptation of the fitting algorithm.

4. Results

In this section, the results of both parts of the experiments described in Sec. 3 are summarized.

Quantitative results for the first part of the evaluation (generalization ability, specificity, and compactness) obtained for all approaches in both scenarios are shown in Tab. 1, Tab. 2, and Fig. 6. Regarding the generalization ability, the novel Locality-SSM approach significantly outperforms the classical SSM, the Wavelet-SSM, and the MOWavelet-SSM approaches for all training set sizes in the single-object as well as the multi-object experiments and the differences are especially high for small number of training samples. Here, the statistical significance between the mean results reported in Tab. 1 & 2 for different modeling approaches obtained for the same training set size over all repetitions are assessed by pairing the mean values for each run in paired t-tests with $p < 0.05$. Only the FEM-SSM approach is able to generate better results than the Locality-SSM approach for some training set sizes. However, this behavior is only seen for larger training set sizes (single-object: $N \geq 40$; multi-object: $N \geq 70$).

The superiority of the Locality-SSM approach with respect to the generalization ability comes along with a relatively high specificity compared to most of the competing approaches. In terms of compactness, the number of modes needed to obtain these results can be characterized as average for the single-object scenario and is lower than that of all other approaches (except for the classical SSM) for the multi-object scenario, showing the efficiency of the proposed approach.

Regarding the individual performances of the other approaches with respect to generalization ability, specificity, and compactness, it can be seen that for the single-object experiments the classical SSM and the MOWavelet-SSM approaches generate comparable results in terms of the generalization ability. This can be explained by the fact that the hierarchy built by the MOWavelet-SSM approach is responsible for modeling inter-object relations and, therefore, despite the larger number of modes does not have any accuracy advantages over the classical SSM in single-object scenarios. On the contrary, for the multi-object scenario, the advantages of this approach over the classical SSM are clearly visible. However, the number of modes needed to achieve these results is relatively high (≈ 150) but can be explained by the hierarchy built. The Wavelet-SSM approach is able to significantly outperform the classical SSM approach in terms of generalization ability in the single-object experiments. As for the MOWavelet-SSM approach in the multi-object setup, the number of modes needed to obtain these results is relatively high and can also be explained by the hierarchy built and the resulting large number of bands that are modeled individually. The FEM-SSM approach also significantly outperforms the classical SSM with respect to the generalization ability and is the only approach to also outperform the Locality-SSM approach for some training set sizes. For this approach, the dependence of the specificity on the number of training samples N is striking. This is a results of the assumed (inverse) linear relationship between N and the proportion of FEM variation to be included in the model (cf. parameter a in Sec. 3). With increasing N, the FEM-SSM model (theoretically) approaches the classical SSM.

The results of the model-based segmentation experiments (second part of the evaluation) are also reported in Tab. 1 & 2 and exemplary illustrations of segmentation results can be found in Fig. 7. The results reported for the classical SSM for $N = 120$ training samples are in the range of the results obtained in the comparison study of van Ginneken et al. (2006) on the same data with the same parameters for the ASM algorithm.

Interestingly, the MOWavelet-SSM approach is not able to outperform the classical SSM in the multi-object setting of the the segmentation experiments. The reasons for this behavior, however, remain unclear but may be related to the noisiness of the candidate contours during the model fitting. For the FEM-SSM approach, Fig. 7 shows that the large number of modes of the multi-object models (cf. Fig. 6) leads to highly noisy contours. This behavior seems to neither affect the specificity results nor the segmentation performance, but can be clearly considered a disadvantage as the contours have an highly un-

Table 1: Generalization ability and segmentation performance for the *single*-object scenarios of both evaluation parts obtained by the different shape modeling approaches tested given varying numbers of training samples. Results are given as mean±standard deviation in mm over all test data sets and repetitions.

# Training samples	Locality-SSM	Classical SSM	FEM-SSM	Wavelet-SSM	MOWavelet SSM
Generalization ability (mean landmark distance [mm])					
5	2.10±1.27	3.78±1.43	2.99±1.28	2.72±1.38	3.82±1.48
10	1.34±1.05	2.71±0.96	1.90±0.80	1.78±0.93	2.75±1.01
15	1.14±0.91	2.36±0.82	1.56±0.67	1.46±0.75	2.39±0.86
20	1.07±0.92	2.03±0.71	1.31±0.58	1.39±0.73	2.07±0.76
30	0.99±0.87	1.80±0.61	1.06±0.50	1.26±0.66	1.83±0.65
40	0.96±0.86	1.62±0.54	0.91±0.47	1.17±0.63	1.65±0.59
70	0.88±0.82	1.44±0.45	0.70±0.40	1.05±0.56	1.47±0.49
120	0.86±0.81	1.29±0.40	0.72±0.36	0.93±0.49	1.31±0.44
Segmentation performance (mean point-to-contour distance [mm])					
5	2.42±1.27	3.23±1.61	2.90±1.56	2.74±1.38	3.18±1.53
10	2.29±1.28	3.06±1.69	2.77±1.68	2.66±1.42	3.01±1.62
15	2.24±1.19	2.91±1.67	2.62±1.59	2.59±1.36	2.86±1.60
20	2.24±1.21	2.86±1.60	2.61±1.59	2.58±1.38	2.83±1.59
30	2.17±1.14	2.85±1.68	2.62±1.61	2.61±1.38	2.79±1.66
40	2.19±1.18	2.77±1.61	2.60±1.56	2.59±1.37	2.76±1.65
70	2.15±1.13	2.72±1.62	2.63±1.57	2.62±1.35	2.70±1.63
120	2.11±1.09	2.69±1.51	2.67±1.53	2.74±1.41	2.70±1.66

Table 2: Generalization ability and segmentation performance for the *multi*-object scenarios of both evaluation parts obtained by the different shape modelling approaches tested given varying numbers of training samples. Results are given as mean±standard deviation in mm over all test data sets and repetitions.

# Training samples	Locality-SSM	Classical SSM	FEM-SSM	MOWavelet SSM
Generalization ability (mean landmark distance [mm])				
5	4.23±2.13	7.90±2.74	7.08±2.67	7.67±2.67
10	2.94±1.85	6.33±2.31	5.18±2.11	5.83±2.19
15	2.63±1.82	5.48±2.21	4.21±1.95	4.90±2.08
20	2.50±1.79	4.99±2.07	3.63±1.78	4.35±1.96
30	2.28±1.78	4.27±1.97	2.91±1.68	3.64±1.90
40	2.17±1.81	3.86±1.91	2.57±1.67	3.31±1.91
70	2.06±1.82	3.33±1.84	2.01±1.62	2.83±1.84
120	1.95±1.93	2.98±1.85	1.65±1.63	2.53±1.86
Segmentation performance (mean point-to-contour distance [mm])				
5	3.62±1.52	5.01±1.71	4.53±1.66	5.19±1.82
10	3.08±1.33	4.42±1.64	3.95±1.57	4.57±1.73
15	2.95±1.25	4.07±1.59	3.62±1.49	4.23±1.74
20	2.88±1.21	3.87±1.58	3.41±1.42	3.97±1.67
30	2.84±1.17	3.62±1.55	3.22±1.37	3.64±1.58
40	2.82±1.21	3.54±1.58	3.16±1.40	3.46±1.54
70	2.84±1.21	3.27±1.41	2.97±1.26	3.22±1.37
120	2.84±1.17	3.10±1.24	2.84±1.19	3.09±1.24

natural appearance. Of course, this noisiness can be reduced by lowering the proportion of FEM variations included in the model. However, initial tests showed that this would lead to a significantly decreased segmentation accuracy.

The most important result of the segmentation experiments is that the Locality-SSM approach significantly outperforms all other approaches in all segmentation experiments (except for the FEM-SSM at $N = 120$ in the multi-object scenario) despite its relatively high specificity. Its negative effects are effectively avoided by the small but natural modifications of the standard ASM algorithm (integration of a hierarchical fitting scheme). Again, the largest improvements of the Locality-SSM approach over the competing methods are seen for small numbers of training samples. It is also interesting to note that the the Locality-SSM approach is able to achieve a segmentation accuracy with very small numbers of training samples (e.g. $N = 15$) that is better than that obtained by the competing approaches with $N = 120$ training samples. This, again, demonstrates the effectiveness of the proposed locality-based modeling approach and is visible in Fig. 7. These exemplary illustrations of individual segmentation results show that the Locality-SSM approach is much better able to adapt to the individual details of the contours than the classical SSM (cf. first and second row in Fig. 7) for small training populations and also shows a better adaptation to the inter-object relations in multi-object scenarios (cf. last two rows in Fig. 7).

5. Conclusion

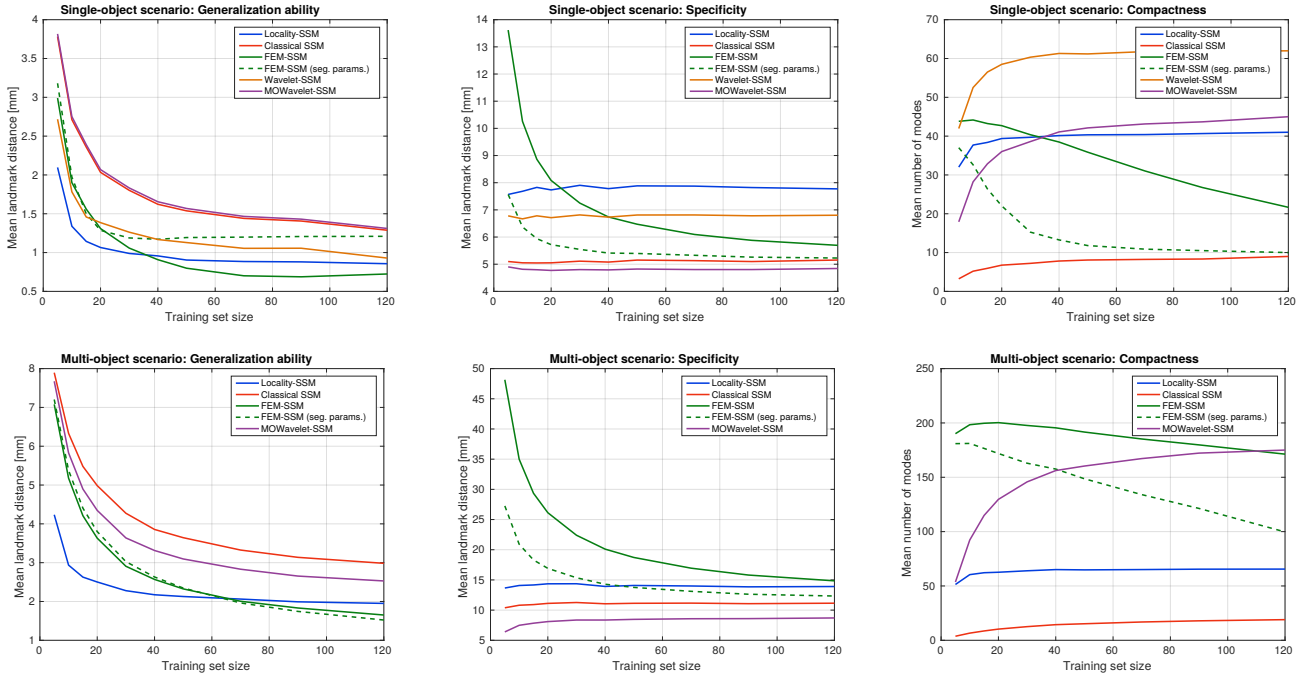
Statistical shape models are widely used in medical image analysis to aid segmentation or classification. These models are typically learnt from a population of previously observed training shapes. However, in the medical context, providing an appropriate and representative training population is cumbersome or even impossible. Therefore, statistical shape models frequently suffer from the high-dimension-low-sample-size (HDLSS) problem resulting in models with insufficient expressiveness.

In this work, a new approach for learning representative multi-object multi-resolution statistical shape models from a small training population was presented to address the HDLSS problem. The method is based on the assumption that local shape variations have limited effects in distant areas and, therefore, can be modeled independently. To integrate this assumption into the classical statistical shape modeling framework, the sample covariance matrix was manipulated by setting the covariances between distant landmarks to zero. A multi-resolution scheme allowing for different levels of locality was derived and equipped with a method to combine variability information modeled at different levels into a combined single shape model. This combined and consistent representation of global and local variability in a single shape model is a major contribution of this work and allows the use of the standard active shape model algorithm for model-based segmentation.

Compared to other approaches that manipulate/augment the sample covariance matrix or artificially enlarge the training population to cope with the HDLSS problem such as the FEM-inspired method of Cootes and Taylor (1995) or the work of Koikkalainen et al. (2008) based on random shape deformations, a methodological advantage of the approach presented in this work is that the resulting model solely relies on shape information already present in the (small) training population. Another methodological advantage is related to the hierarchy defined in the proposed approach by the different levels of locality as it allows to easily derive a hierarchical fitting scheme.

Although the new approach is guided by the assumption that

Figure 6: Generalization ability, specificity, and compactness obtained during the single-object and multi-object experiments for the different shape modelling approaches given varying numbers of training samples. Smaller values indicate better models. For the FEM-SSM approach, the results obtained when using the optimized parameters for the segmentation experiments are given here for comparison as dashed lines.



local shape variations have only limited influence in distant areas, the introduced covariance matrix manipulations do not lead to complete independence of different local areas such as in the patch-based approach of Zhao et al. (2005). This is because even if the covariance of two points of an object is set to zero an indirect dependency still exists due to the neighborhood relations of all points. Hence, generated shapes are smooth and no post-processing (e.g. stitching of piecewise smooth patches, ...) is required.

In comparison to the diverse hierarchical wavelet-based modeling approaches originating from the work of Davatzikos et al. (2003), a general advantage of the locality-based approach is that no costly wavelet decomposition/synthesis needs to be performed during model application. Compared to the wavelet-based approach of Cerrolaza et al. (2012), which was designed to model inter-object relations by manually defining an object hierarchy, the approach presented here is also able to model intra-object relations and provides a unified and automatic way to build the hierarchy based on the distance thresholds chosen. In (Cerrolaza et al., 2015), (Cerrolaza et al., 2012) was extended by introducing an automatic hierarchy generation method and the possibility to model intra-object relations. However, in contrast to the locality-based approach, there approach performs a post-processing of the generated shape during model application to handle shape inconsistencies resulting from the hierarchical object splitting.

An extensive evaluation based on a public data base of 247 chest radiographs was performed to show the modeling and segmentation capabilities of the novel approach in single- and multi-object HDLSS scenarios. Within the evaluation the new approach was not only compared to the classical shape modeling method but also to three state-of-the-art shape modeling approaches specifically designed for HDLSS scenarios. The new approach outperformed all other approaches in (nearly) all situations in terms of generalization ability and model-based segmentation accuracy and, therefore, illustrated its effectiveness. Most prominent improvements over the competing approaches were observed for small training populations (e.g. 15 samples). It was also observed that the new approach is able to achieve segmentation results with very small numbers of training samples (e.g. 15) that are better than the results obtained by the other approaches with 120 training shapes. Both of these results prove the practical impact of this work in HDLSS scenarios. However, it has to be noted that the new approach suffers from a relatively high specificity but its (possible) negative impact in the segmentation experiments was successfully avoided by employing a hierarchical multi-resolution fitting scheme.

For the initial evaluation performed for this work, the experiments were restricted to the 2D case mainly because of the availability of public data. The methodology developed in this article is not limited to a certain dimensionality of the input data and only depends on the definition of a reasonable distance measure to assess locality. However, in future work, the approach should be evaluated on additional data (2D and 3D) to verify the results presented here. Future work should also investigate ways on how to apply the locality assumption underlying the approach presented here without having to directly modify the sample covariance matrix as this requires an explicit calculation (and storage) of this possibly large matrix. Such a modification would also enable the use of this methodology in

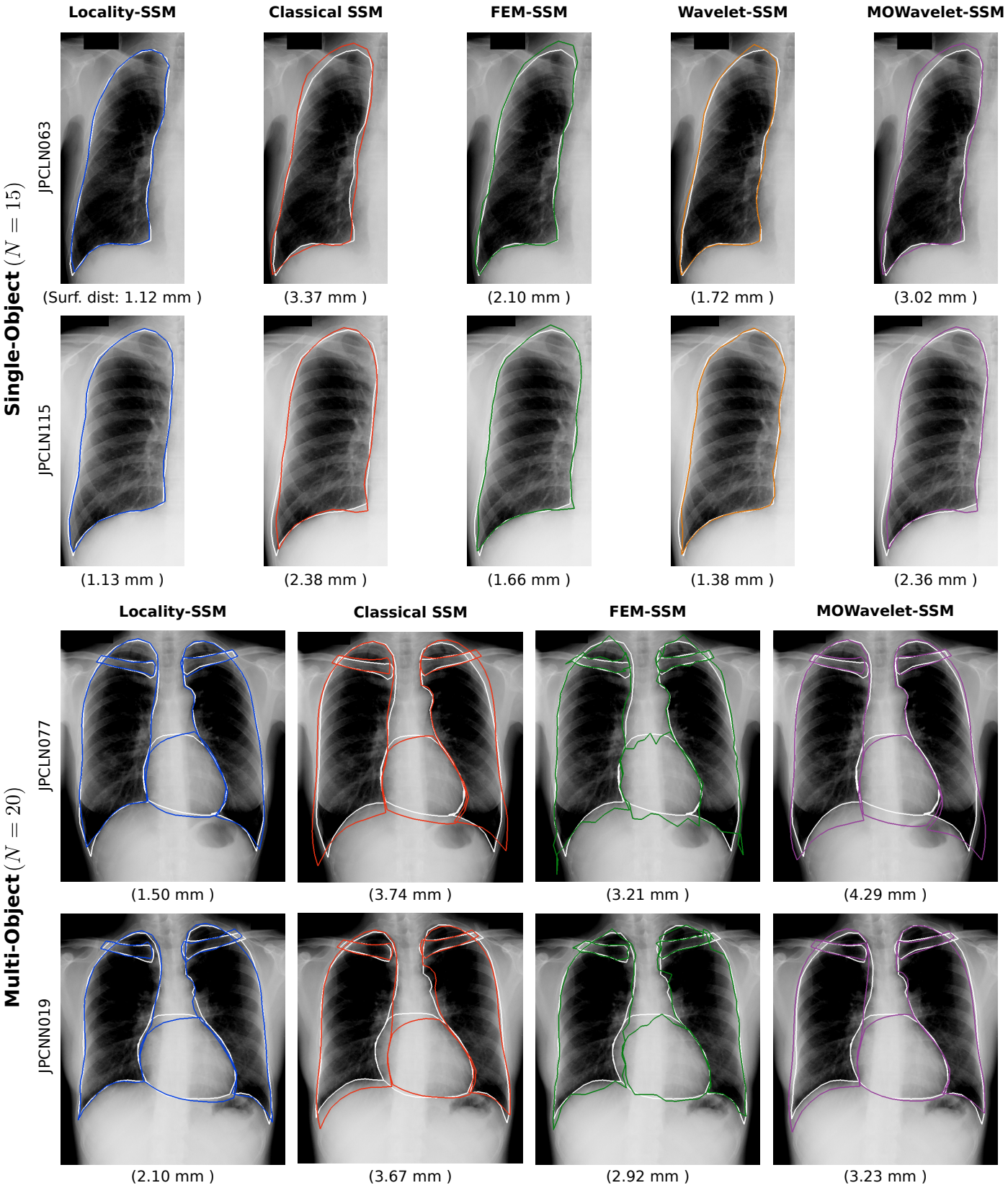


Figure 7: Exemplary results obtained for all modelling approaches for the single- and multi-object scenarios given small numbers of training shapes (15 and 20, respectively). The mean surface distance between the segmentation resulting from the ASM algorithm (colored contours) and the manual ground-truth segmentation (white contour) is given under each image.

scenarios such as statistical deformation modelling, which also suffer from the HDLSS problem and where the dimensionality of the input data is usually magnitudes higher compared to shape modeling (Xue et al., 2006; Onofrey et al., 2015).

Acknowledgment

This work was supported by the German Research Foundation DFG under grant EH 224/6-1.

References

Absil, P.-A., Mahony, R., Sepulchre, R., 2004. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. Acta Applicandae Mathematica 80 (2), 199–220.

Cerrolaza, J. J., Reyes, M., Summers, R. M., Gonzalez-Ballester, M. A., Linguraru, M. G., 2015. Automatic multi-resolution shape modeling of multi-organ structures. Medical Image Analysis 25 (1), 11 – 21.

Cerrolaza, J. J., Villanueva, A., Cabeza, R., 2011. Multi-shape-hierarchical active shape models. In: Proceedings of International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV11). Vol. 1. pp. 137–143.

Cerrolaza, J. J., Villanueva, A., Cabeza, R., 2012. Hierarchical statistical shape models of multiobject anatomical structures: application to brain mri. Medical Imaging, IEEE Transactions on 31 (3), 713–724.

Cootes, T., Baldock, E., Graham, J., 2000. An introduction to active shape models. Image processing and analysis, 223–248.

Cootes, T., Taylor, C., Cooper, D., Graham, J., 1992. Training models of shape from sets of examples. In: British Machine Vision Conference 1992. Springer, pp. 9–18.

Cootes, T. F., Taylor, C. J., 1995. Combining point distribution models with shape models based on finite element analysis. Image and Vision Computing 13 (5), 403–409.

Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J., 1995. Active shape models-their training and application. Computer vision and image understanding 61 (1), 38–59.

Davatzikos, C., Tao, X., Shen, D., 2003. Hierarchical active shape models, using the wavelet transform. Medical Imaging, IEEE Transactions on 22 (3), 414–423.

Davies, R., Twining, C., Cootes, T., Waterton, J., Taylor, C., May 2002. A minimum description length approach to statistical shape modeling. Medical Imaging, IEEE Transactions on 21 (5), 525–537.

Duta, N., Sonka, M., 1998. Segmentation and interpretation of mr brain images. an improved active shape model. Medical Imaging, IEEE Transactions on 17 (6), 1049–1062.

Heimann, T., Meinzer, H.-P., 2009. Statistical shape models for 3d medical image segmentation: A review. Medical Image Analysis 13 (4), 543 – 563.

Higham, N. J., 2002. Computing the nearest correlation matrix – a problem from finance. IMA journal of Numerical Analysis 22 (3), 329–343.

Jain, A. K., Zhong, Y., Dubuisson-Jolly, M.-P., 1998. Deformable template models: A review. Signal Processing 71 (2), 109 – 129.

Koikkalainen, J., Tolli, T., Lauerma, K., Antila, K., Mattila, E., Lilja, M., Lotjonen, J., 2008. Methods of artificial enlargement of the training set for statistical shape models. Medical Imaging, IEEE Transactions on 27 (11), 1643–1654.

Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K., 2004. Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images . Medical Image Analysis 8 (3), 371 – 386.

Nain, D., Haker, S., Bobick, A., Tannenbaum, A., 2007. Multiscale 3-d shape representation and segmentation using spherical wavelets. IEEE Transactions on Medical Imaging 26 (4), 598–618.

Onofrey, J., Papademetris, X., Staib, L., 2015. Low-dimensional non-rigid image registration using statistical deformation models from semi-supervised training data. Medical Imaging, IEEE Transactions on 34 (7), 1522 – 1532.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.-i., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American Journal of Roentgenology 174 (1), 71–74.

van Ginneken, B., Stegmann, M. B., Loog, M., 2006. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Medical Image Analysis 10 (1), 19 – 40.

Xue, Z., Shen, D., Davatzikos, C., 2006. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3d warping. Medical Image Analysis 10 (5), 740 – 751.

Ye, K., Lim, L.-H., 2014. Distance between subspaces of different dimensions. arXiv preprint arXiv:1407.0900v1.

Zhao, Z., Aylward, S. R., Teoh, E. K., 2005. A novel 3D partitioned active shape model for segmentation of brain MR images. In: MICCAI 2005. Springer, pp. 221–228.