Patch-Based Low-Rank Matrix Completion for Learning of Shape and Motion Models from Few Training Samples

Jan Ehrhardt(✉), Matthias Wilms, and Heinz Handels

Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
ehrhardt@imi.uni-luebeck.de

Abstract. Statistical models have opened up new possibilities for the automated analysis of images. However, the limited availability of representative training data, e.g. segmented images, leads to a bottleneck for the application of statistical models in practice. In this paper, we propose a novel patch-based technique that enables to learn representative statistical models of shape, appearance, or motion with a high grade of detail from a small number of observed training samples using low-rank matrix completion methods. Our method relies on the assumption that local variations have limited effects in distant areas. We evaluate our approach on three exemplary applications: (1) 2D shape modeling of faces, (2) 3D modeling of human lung shapes, and (3) population-based modeling of respiratory organ deformation. A comparison with the classical PCA-based modeling approach and FEM-PCA shows an improved generalization ability for small training sets indicating the improved flexibility of the model.

Keywords: Statistical modeling · High-dimension-low-sample-size problem · Low-rank matrix completion · Virtual samples

1 Introduction

Statistical models play an important role in several tasks in computer vision and image analysis, such as image segmentation and object classification. These models aim to represent properties like shape or intensity of a class of objects based on a population of observed instances. However, collecting an adequately large and representative training population is often laborious and challenging, particularly if dimensionality and complexity of the observed objects increase. Therefore, many applications suffer from the high-dimension-low-sample-size (HDLSS) problem. In the application of statistical shape models (SSMs) [7] or eigenfaces (eigenimages) [27] for segmentation or recognition tasks, a small sample size results in a limited flexibility of the model and details can not be represented adequately (see Fig. 1).

This paper proposes a method for statistical shape, appearance, and motion modeling with increased ability to adapt to local details, thus, increasing the

B. Leibe et al. (Eds.): ECCV 2016, Part IV, LNCS 9908, pp. 712–727, 2016.
DOI: 10.1007/978-3-319-46493-0_43

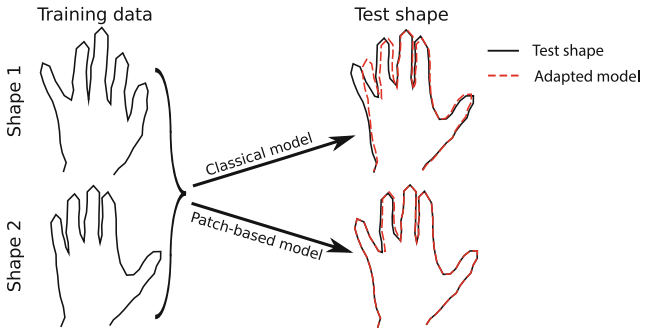


Fig. 1. Example application of the patch-based modeling approach using only two training shapes: classical models only learn the global transition between the two shapes. The patch-based model combines local shape details, and can adapt to test shapes showing local properties of both shapes.

flexibility of models generated from few training samples. The method is based on the assumption of locality, i.e. we assume that local variations in shape, intensity, or motion have limited effects in distant areas. This allows the model to combine local variations observed in different training samples while preserving overall object properties, i.e. generating valid instances. During the learning phase, the objects are partitioned into patches and distant patches of different samples are fused into virtual samples. By doing so, a very large number of virtual samples can be generated from few training instances. To avoid the problem of discontinuities at patch borders, a sparse sampling is performed and the model generation is formulated as a low-rank matrix completion problem. Thus, its left to the model to "fill in" fitting information between patches of different training instances. We validate the proposed approach using three exemplary applications in the fields of computer vision and medical image analysis: (1) 2D shape modeling of faces, (2) 3D modeling of human lung shapes, and (3) population-based modeling of respiratory organ deformation. Evaluations show that our model features an increased generalization ability for problems of different complexity while generated instances are still valid.

1.1 Related Work

Statistical Shape Models. Since their introduction in the early 1990s [7], statistical shape models have proven to be effectively addressing a large number of image segmentation problems. The most generic method to generate a SSM is to build a *point distribution model* (*PDM*) by applying a principal component analysis (PCA) to the sample matrix $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ containing a given set of n training shapes. Each training shape $\boldsymbol{x}_i$ is represented as a m-dimensional vector composed by landmark points or pseudo-landmarks $\boldsymbol{x}_i = (x_1, y_1, z_1, \ldots, x_{\frac{m}{3}}, y_{\frac{m}{3}}, z_{\frac{m}{3}})^T$. There is a variety of other shape representations (see [13] for an overview), but landmark-based systems are the most popular ones.

Eigenimages and Population-Based Deformation Models. Eigenfaces (or eigenimages) [27] is the application of PCA-based methods for intensity modeling, where each training image is represented in a sample vector $\boldsymbol{x}_i$ by concatenating the pixel values. Similar to eigenfaces, PCA-based methods can be applied to a population of deformation fields. This approach has many applications in medical image analysis, e.g. to model respiratory or cardiac organ deformations [10], for morphometric studies in computational anatomy or as priors for atlas-patient matching [24,25]. The typical dimensionality of those models is between several ten thousands and several millions, and the HDLSS problem is intensified by the limited availability of medical image data and the laborious generation of application-specific training sets.

The HDLSS Problem in Statistical Modeling. Generally, two different ways exist to tackle the HDLSS problem in statistical modeling: (1) Only observed training samples are used but changes are made to the modeling process to allow for higher flexibility. (2) The modeling approach remains (largely) unchanged but additional virtual training samples are generated. Common patch-based and hierarchical techniques applied in shape modeling, object classification, or recognition tasks [5,8,16,19,29–31] belong to the first category. In patch-based approaches for object classification [16,31], the samples are subdivided to model features of small regions independently. These approaches are, however, limited to classification tasks as they do not learn a consistent generative model as needed for, e.g., segmentation tasks. The patch-based shape modeling approach for medical image segmentation proposed in [30] follows the same idea and independently models different parts of an object. This approach also does not generate a consistent model and thus the consistency needs to be enforced during model application. Hierarchical approaches used in shape modeling [5,8] subdivide the (generative) model into several parts for increased flexibility. Common approaches belonging to the second category generate virtual samples by applying (random/heuristic) transformations to individual training instances or simulate different noise levels, location errors or lightning effects [6,15,29].

Conceptually, our approach belongs to the second category. However, our work differs from previous works and contributes to the state-of-the art in two major aspects: (1) We use sparse virtual samples combining information of *different* observed samples and (2) elegantly unify virtual sample completion and learning of a generative model. The second aspect guarantees the reconstruction of valid (but perpetuated) instances, and although our method is motivated by the HDLSS problem occurring in medical applications, it could be of interest for other applications, e.g. for data augmentation in the context of deep learning.

2 Methods

We begin by briefly describing the generation of classical point distribution models and its relation to matrix factorization. As introduced in the last section, our training samples are given by m-dimensional vectors $\boldsymbol{x}_i$ that are assembled in

the data matrix $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. The central step to build a PDM is a principal component analysis of the data matrix $\mathbf{X}$ and a dimensionality reduction by selecting only the principal components corresponding to the k largest eigenvalues [7]. The problem can be formulated as the low-rank approximation

$$\hat{\mathbf{M}} = \arg\min_{\mathbf{M}} \|\tilde{\mathbf{X}} - \mathbf{M}\|_F^2 \quad \text{s.t. } rank(\mathbf{M}) = k, \tag{1}$$

where $\tilde{\mathbf{X}}$ is the centered data matrix. A singular value decomposition (SVD) $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$ can be applied to solve Eq. (1), followed by selecting the right and left singular vectors associated with the k largest singular values $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_k$. $\hat{\mathbf{M}}$ is then given by the truncated matrices $\hat{\mathbf{M}} = \tilde{\mathbf{U}}_k\tilde{\mathbf{\Sigma}}_k\tilde{\mathbf{V}}_k^T$. The number of basis vectors k is usually controlled by setting a threshold as follows

$$\sum_{i=1}^{k} \sigma_i^2 / \sum_{i=1}^{n} \sigma_i^2 \geq \tau, \tag{2}$$

where common values for τ are 0.9 – 0.98. The truncated matrix $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times k}$ defines an orthonormal basis and together with a distribution of the shape parameters (the weights associated with each basis vector) they define the *classical statistical model*.

Now, let us assume that data is missing in the observed samples and let Ω be the subset of $[m] \times [n]$ of the available entries in the data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$. A statistical model can be generated by solving the following low-rank matrix completion problem:

$$\hat{\mathbf{M}} = \arg\min_{\mathbf{M}} \|\mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{M})\|_F^2 \quad \text{s.t. } rank(\mathbf{M}) = k, \tag{3}$$

with $\mathcal{P}_\Omega$ being the projection operator

$$(\mathcal{P}_\Omega(\mathbf{X}))_{ij} = \begin{cases} \mathbf{X}_{ij} & (i,j) \in \Omega \\ 0 & \text{else} \end{cases}. \tag{4}$$

If a solution of Eq. (3) is found, basis and variances of the model are given by the SVD $\hat{\mathbf{M}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

The theoretical properties of low-rank matrix completion (MC) and conditions for a successful matrix recovery are well studied, see e.g. [4]. Available methods to solve Eq. (3) can be roughly divided into methods based on nuclear norm minimization (e.g., singular value thresholding (SVT) [3] and the projected proximal point algorithm (ProPPA) [17]) and algorithms based on minimization on the Grassmann manifold (e.g., OptSPACE, GROUSE and GRASTA [12]). A disadvantage of the Grassmannian-based approaches is that an upper-bound guess of the desired rank is needed, on the other hand, some of these algorithms, like GROUSE and GRASTA, allow online matrix completion and therefore the application in large scale problems.

In our applications, difficulties arise from the ill-conditioned nature of the data matrix $\mathbf{X}$ resulting in large reconstruction errors and slow convergence for many MC algorithms [21]. However, recently, several algorithms were proposed to improve the performance for those matrices [14,21,22]. Among these, polar incremental matrix completion (PIMC) is based on the GROUSE algorithm and can be applied to streaming data [14].

2.1 Low-Rank Matrix Completion of Ill-Conditioned matrices

This section briefly introduces the polar incremental matrix completion (PIMC) algorithm used to solve Eq. (3) in our applications. For a more detailed derivation and description we refer the reader to [14].

To enable an online update of the model, the incremental update of basis $\mathbf{U}$ and singular values $\mathbf{\Sigma}$ is needed. Let $\mathbf{M}_t = \mathbf{U}_t\mathbf{R}_t^T$ be the estimated rank k-factorization of the (sparse) data matrix $\mathbf{X}_t \in \mathbb{R}^{m\times t}$ for t observed samples. Given a new sample $\boldsymbol{x}_{\Omega_t}$ with observed entries $\Omega_t \subset \{1,\dots,m\}$, we can compute weights $\boldsymbol{w}_t = \arg\min_{\boldsymbol{w}} \|\mathbf{U}_{\Omega_t}\boldsymbol{w} - \boldsymbol{x}_{\Omega_t}\|_2^2$ to interpolate values at unobserved entries

$$\tilde{\boldsymbol{x}}_t = \begin{cases} \boldsymbol{x}_{\Omega_t} & \text{on } \Omega_t \\ \mathbf{U}_t\boldsymbol{w}_t & \text{otherwise} \end{cases}, \tag{5}$$

where $\mathbf{U}_{\Omega_t}$ contains only the rows Ω_t of $\mathbf{U}$. To update $\mathbf{M}_t$ according to the new sample, we have to solve for

$$\min_{\mathbf{M}} \|[\mathbf{U}_t\mathbf{R}_t^T \ \tilde{\boldsymbol{x}}_t] - \mathbf{M}\|_F^2 \quad \text{s.t. } rank(\mathbf{M}) = k. \tag{6}$$

Given that $\mathbf{R}_t = \mathbf{V}_t\mathbf{\Sigma}_t$ for an orthogonal matrix $\mathbf{V}_t$, iterative SVD [2] can be used to efficiently solve Eq. (6) using

$$[\mathbf{U}_t\mathbf{R}_t^T \ \tilde{\boldsymbol{x}}_t] = \begin{bmatrix} \mathbf{U}_t & \frac{\boldsymbol{r}_t}{\|\boldsymbol{r}_t\|} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_t & \boldsymbol{w}_t \\ 0 & \|\boldsymbol{r}_t\| \end{bmatrix} \begin{bmatrix} \mathbf{V}_t^T & 0 \\ 0 & 1 \end{bmatrix} \tag{7}$$

and performing an SVD on the central $(k+1)\times(k+1)$ matrix $\begin{bmatrix} \mathbf{\Sigma}_t & \boldsymbol{w}_t \\ 0 & \|\boldsymbol{r}_t\| \end{bmatrix} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^T$ followed by the update

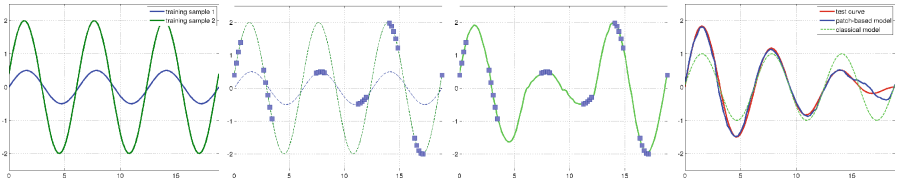
$$\begin{aligned} \mathbf{U}_{t+1} &= \begin{bmatrix} \mathbf{U}_t & \frac{\boldsymbol{r}_t}{\|\boldsymbol{r}_t\|} \end{bmatrix} \hat{\mathbf{U}}, \mathbf{V}_{t+1} = \begin{bmatrix} \mathbf{V}_t^T & 0 \\ 0 & 1 \end{bmatrix} \hat{\mathbf{V}} \\ \mathbf{\Sigma}_{t+1} &= \hat{\mathbf{\Sigma}}, \text{ or } \mathbf{R}_{t+1} = \mathbf{V}_{t+1}\hat{\mathbf{\Sigma}} \end{aligned} \tag{8}$$

and dropping the smallest singular value and corresponding singular vector to obtain a rank k factorization.

As shown in [2,14], this algorithm is equivalent to GROUSE for a specific step size, if setting $\mathbf{\Sigma}_t = \mathbf{I}$ and $\mathbf{V}_t = \mathbf{R}_t$ in Eq. (7) and using the updates in Eq. (8). This reveals the sensitivity of GROUSE to ill-conditioned matrices, because constant singular values are assumed.

To overcome this restriction, the authors of [14] propose to use the following approach to update the model in each step: Let $\mathbf{R}_t = \tilde{\mathbf{V}}_t\tilde{\mathbf{S}}_t$ be the polar decomposition of $\mathbf{R}_t$ into the matrix $\tilde{\mathbf{V}}_t \in \mathbb{R}^{m\times k}$ with orthonormal columns, and a positive semidefinite matrix $\tilde{\mathbf{S}}_t \in \mathbb{R}^{k\times k}$. Although $\tilde{\mathbf{S}}_t$ is not diagonal it presents an estimate of the singular values in the current subspace. Let further $\gamma_t = \frac{\kappa_0 \sum_t \|\boldsymbol{x}_{\Omega_t}\|_2}{\|\tilde{\mathbf{S}}_t\|_F}$ be a scaling value, then Eq. (7) can be rewritten as

$$[\mathbf{U}_t\mathbf{R}_t^T \ \tilde{\boldsymbol{x}}_t] = \begin{bmatrix} \mathbf{U}_t & \frac{\boldsymbol{r}_t}{\|\boldsymbol{r}_t\|} \end{bmatrix} \begin{bmatrix} \gamma_t\tilde{\mathbf{S}}_t & \boldsymbol{w}_t \\ 0 & \|\boldsymbol{r}_t\| \end{bmatrix} \begin{bmatrix} \frac{1}{\gamma_t}\tilde{\mathbf{V}}_t^T & 0 \\ 0 & 1 \end{bmatrix}, \tag{9}$$



(a) training set (b) sparse patch selection (c) completed virtual sample (d) model adaptation

Fig. 2. Demonstrative example describing the patch-based modeling method using two sinusoidal training shapes (see text for details).

and Eq. (8) are used to update the model. In contrast to iterative SVD, the re-computation of $\tilde{\mathbf{S}}_t$ using the polar decomposition is required, because wrong estimates of the singular values may appear due to the missing data. Further, the interpolated data vectors $\tilde{\boldsymbol{x}}_t$ are used in the update, i.e. the singular values will increase according to the interpolated data and not according to the observed data $\boldsymbol{x}_{\Omega_t}$. Therefore, a rescaling to the norm of the actual observed data is performed using the parameter γ_t, with $\kappa_0 \ll 1$ preventing abrupt changes.

2.2 Patch-Based Model Generation

We now come back to the HDLSS problem in statistical modeling (cf. Sect. 1). To overcome this problem, several approaches for SSM generation propose to learn local models by dividing the shape into parts. A difficulty arises in recombining these local models to a global shape.

The main idea of the presented method is to combine local information of different training shapes in artificially generated virtual samples. Instead of learning individual local models or recombining these local variations into complete new training samples, we provide only partial information to learn the model. Thus, we exploit the low-rank structure of the subspace to find a model that fits the generated samples. In previous work probabilistic PCA was used to learn shape models from partial information [20]. Our approach for model generation is illustrated using the simple one-dimensional example in Fig. 2. The training samples are given by two scaled and shifted sine functions. A large number of sparse virtual samples is generated by selecting small patches randomly from both training samples (Fig. 2(b)). The virtual samples are agglomerated in the sparse data matrix $\mathbf{X}$ and the low-rank matrix completion problem given in Eq. (3) is solved to compute the completed matrix $\mathbf{M} = \mathbf{U}\mathbf{R}^T$ of given rank k. Figure 2(c) shows the reconstructed virtual samples. The computed basis $\mathbf{U}$ can now be applied to approximate a new sample $\boldsymbol{y}$ by

$$\hat{\boldsymbol{y}} = \mathbf{U}\hat{\boldsymbol{w}} \quad \text{with} \quad \hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\mathbf{U}\boldsymbol{w} - \boldsymbol{y}\|_2^2, \tag{10}$$

as illustrated in Fig. 2(d). Here, a damped sine function is used to show the approximation quality. Although the function resembles the second training

sample near zero and approaches the first sample near 5π, classical global models can not achieve a good fitting result. In contrast, the patch-based method can combine *local* properties of both training samples and allows for a good approximation of the damped sine function. Interestingly, the approximation quality decreases in the interval $[5\pi, 6\pi]$ and the approximated curve is forced to be periodic. This behavior results from the fact that we learned the periodicity of the training functions by applying periodic boundary condition during the patch-selection (see Fig. 2(c)).

We can summarize the patch-based model generation algorithm as follows:

Algorithm 1. Patch-based model generation (batch)

Require: N training samples $\boldsymbol{x}_i \in \mathbb{R}^m, i = 1, \ldots, N$, estimated rank k, sparsity $p \in (0, 1)$, number of virtual samples $n \gg N$

Generate sparse virtual samples:
for each $\boldsymbol{x}_{\Omega_j}, j = 1, \ldots, n$ **do**
Apply a *patch selection strategy* to select random patches from different training samples until at least $p \cdot m$ entries of $\boldsymbol{x}_{\Omega_j}$ are filled
Agglomerate $\boldsymbol{x}_{\Omega_j}$ in sparse matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$
end for
Solve low-rank matrix completion:
Compute $\mathbf{M} = \mathbf{U}\mathbf{R}^T$ of rank k by solving Eq. (3)
using the algorithm described in Sect. 2.1
Estimate a distribution of the shape parameters from matrix $\mathbf{R}$ or from estimated singular values $\hat{\boldsymbol{\Sigma}}$ (Eq. (8))

Output: Model defined by orthonormal basis $\mathbf{U}$ and associated distribution of the shape parameters, e.g. $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}, diag(\sigma_1, \ldots, \sigma_k))$

To select suitable parameters for Algorithm 1, the dependency $p \cdot mn \geq \zeta(m + n - k)k$ can be used, where ζ is the *oversampling ratio* [21]. Most algorithms for matrix completion yield robust results with $\zeta \approx 6$. In our applications we selected a sparsity of $p = 0.3$, lower values increase the training size and higher values complicate the selection of suitable patches. The estimated rank k is application dependent and will be discussed in Sect. 3. To estimate a distribution of the model parameters, one possibility is to assume a normal distribution $\mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\sigma_1}, \ldots, \hat{\sigma_k})$ and compute standard deviations and mean from the matrix $\mathbf{R}$.

2.3 Patch Selection and Domain Partitioning

The remaining component of the algorithm is the generation of virtual samples, i.e. the selection of patches to fill the entries of $\boldsymbol{x}_{\Omega_j}$. This step affects the properties of the generated model and at the same time interacts with the matrix factorization algorithm. In general, the applied patch selection strategy depends on the regarded application. Here, we purposely apply simple sampling strategies to show the strength of the presented approach in different applications without elaborate fine tuning.

The guiding assumption is that local variations have limited influence in distant areas. Consequently, global variations in pose and orientation have to be removed from the training set before model generation. Furthermore, prior knowledge about the minimum distance between independent areas is needed to partition the domain of interest (surface mesh or image space), and each partition should be able to reflect local domain properties, e.g. orientation or curvature. For each virtual sample to generate, partitions are randomly assigned to different training samples. The number of training samples used to generate each virtual sample influences the globality of the model and the needed rank – a small number (two or three) was sufficient in our tests.

Sampling large partitions would lead to block-like structures in the data matrix, which impedes the convergence of the matrix completion algorithm [14]. Therefore, many smaller patches drawn from each partition are used to fill the (incomplete) data matrix. Patches sampled from different training shapes should be *detached* to avoid the learning of discontinuities at patch stitches.

A partitioning of the domain is obvious if multiple objects are modeled together, e.g. the facial structures in Fig. 3. For applications like eigenfaces or deformation models the partitioning is simple because the rectangular image space can easily be divided. For surface models existing mesh partitioning methods [26] can be used. In our experiments, we apply a generic mesh partitioning approach, which randomly partitions a triangulated genus-0 surface into equally sized parts. First, the triangulated surface is mapped onto a unit sphere using an unconstrained energy-based method [11] and subsequently the unit sphere is partitioned into regions of equal area [18].

3 Experiments and Applications

To demonstrate the practicability of our approach, we evaluate the patch-based model for three different types of models: 2D contour data, 3D surface meshes, and 2D deformation fields. The proposed algorithm is compared with the classical modeling approach (see Sect. 2) and systematically evaluated using different training sizes. For 2D contour data, our approach is further compared with the FEM-PCA model of Cootes and Taylor [6] combining the standard PCA and finite element method (FEM). This approach also addresses the HDLSS problem and has shown to perform among the best in [15]. FEM-PCA manipulates the data covariance matrix and can therefore only be applied for data of moderate size.

Generalization error and *specificity error* introduced by Davies et al. [9] are used as quantitative performance measures for the statistical models. The generalization error describes the ability to model unseen shapes, and is measured by the distance of the closest model instance to the samples in a test set. The specificity error indicates the validity of the shapes produced by the model. For specificity estimation, a high number of random model instances are generated and the minimal distance to one of the samples in the database is computed. Note that small values indicate better models for both measures.

Fig. 3. 50 example contours of the IMM face database [23] (left), and randomly generated contours from $N = 4$ training samples using the classical model (middle) and the patch-based model (right). The patch-based model shows a higher variability while the overall shape is preserved.

3.1 2D Contour Data of the IMM Face Database

In our first experiment, we apply our algorithm to facial annotations contained in the IMM face database [23]. This database provides 58 facial landmarks of 40 subjects with 6 different expressions (240 in total). Three of the expressions in the database contain rotations of the head, which contradicts our assumption of locality. Therefore, these expressions are excluded and 120 samples (40 subjects, 3 expressions) were used in total. Figure 3 shows 50 example contours of the provided faces after alignment with similarity transformations.

Experiment Design: To evaluate the performance of the patch-based approach, models are generated for varying numbers N of available training samples. For each model generalization ability and specificity are computed and compared to the classical model using identical training and test sets. The N training samples are chosen randomly from the 120 available samples together with a disjunct test set of 30 samples. For each size N the experiments are repeated 60 times and the resulting measures are averaged. After model computation each sample in the test set is approximated using Eq. (10) and average landmark distances are computed to determine the generalization ability. To measure the specificity, 1000 random samples are generated using the computed model basis $\mathbf{U}$ and normally distributed weights $\boldsymbol{w} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} = \operatorname{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2))$, with standard deviations and mean computed from the estimated matrix $\mathbf{R}$.

Virtual Sample Generation: Each facial landmark in the IMM database is labeled with the associated facial structure (eyebrows, eyes, nose, mouth and jaw) so that a partitioning is already given. The different parts are not connected whereby the patch selection strategy is further simplified. The following strategy is applied to generate one virtual sample $\boldsymbol{x}_{\Omega_j}$: Two training samples $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ are selected randomly (possibly $A = B$) and each partition is randomly assigned to one of the drawn samples $\boldsymbol{x}_A$ or $\boldsymbol{x}_B$. Then, patches of size 1 (landmarks) are drawn randomly and depending on the associated facial structure the coordinates (values) are taken from $\boldsymbol{x}_A$ or $\boldsymbol{x}_B$. The last step is repeated until $p \cdot m$ of the entries of $\boldsymbol{x}_{\Omega_j}$ are known.

Determining the Model Parameters: We computed a rank of 13 for the complete set of 120 samples using Eq. 2 with a threshold of $\tau = 0.95$ for the ratio of the total variance. Therefore, for the classical PDM the rank $k = 13$ if $N > 13$ and $k = N$ otherwise is used. The proposed patch-based modeling method is designed to learn additional variations beside the inter-sample variance, this is taken into account by using a higher rank of $k = 13 + \delta$, where δ is set arbitrary to 10 in this experiment. To show the ability to generate a reasonable and feasible model independent of the number of training samples the rank is left constant for all training sizes. The other parameters are chosen as $p = 0.3$ and $n = 400$. The FEM-PCA model uses the same rank as the patch-based approach and the control parameter is set to $\alpha = \alpha_1/N$ as suggested in [6] with $\alpha_1 = 20$.

3.2 3D Lung Surfaces of the LIDC Database

In the second experiment, 3D shape models of the right lung are generated. This experiment is based on image data from the publicly available LIDC-IDRI database [1] that provides > 1000 thoracic 3D CT images of patients with lung nodules. Here, we use a subset of $N = 160$ randomly selected images and extract the lungs via thresholding. Based on these segmentations, an average lung shape is computed as described in [10], which is subsequently triangulated and decimated to obtain 2000 pseudo-landmarks. After registering the atlas to all images by applying an open-source registration method [28], the resulting deformation fields are used to propagate the landmarks to all 160 lungs to define correspondences for the shape modeling process.

For virtual sample generation, the lungs are partitioned into 10 randomly placed areas of equal size generated by the approach described in Sect. 2.3. Using these partitions, the same strategy as used for the face data in the first experiment is applied. The rank estimation leads to a rank of 21 ($\tau = 0.95$) for the complete set of 160 shapes. The experiments are repeated 25 times, and the parameters of our approach are set to $\delta = 20, p = 0.3$, and $n = 1000$.

3.3 Respiratory Lung Motion

At last, we use the proposed method to generate population-based models for respiratory lung motion. We use 2D sagittal slices of lung CT images of $N = 38$ patients and a size of 160×200 pixels acquired at two breathing phases: end inspiration and end expiration. An open-source image registration toolbox [28] is applied to estimate a dense deformation field to describe the respiration-related organ deformations. Following the approach in [10], all images and the associated deformation fields are transformed into a common atlas space to establish anatomical correspondence between the patients. Figure 4 shows example images and computed motion fields for two patients.

Let $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_N$ be the (aligned) deformation fields with each pixel $\mathbf{\Phi}(x, y) = (u, v)$ describing the displacement from inspiration to expiration. By concatenating the u and v components of all pixels, each image can be represented by a sample vector of dimension $m = 64000$. Virtual sample generation starts by

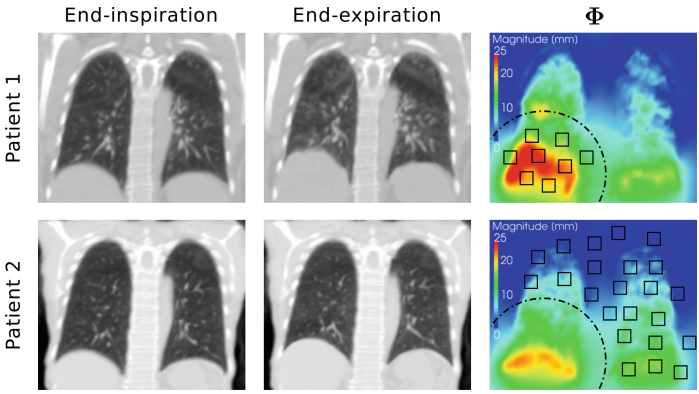


Fig. 4. CT images of two patients for different breathing phases and magnitudes of associated motion fields describing the respiratory lung deformation. The dotted line indicates the partitioning of the image domain, and the squares show the patch selection for one sparse virtual sample.

random selection of two training samples $\mathbf{\Phi}_A$ and $\mathbf{\Phi}_B$ and a seed pixel (s_x, s_y). Then, random patch centers (p_x, p_y) are determined and all displacements inside a patch of size 9×9 are used for the virtual sample. If the distance between (s_x, s_y) and (p_x, p_y) is below ϵ the values are sampled from $\mathbf{\Phi}_A$ and from $\mathbf{\Phi}_B$ otherwise. The distances between patch centers are required to be larger than $\kappa > 9\sqrt{2}$. Patches are selected until $p \cdot m$ entries are filled. The generation of one virtual samples is sketched in Fig. 4(right).

The estimated rank for all available training samples is $k = 27$ ($\tau = 0.95$) and the parameters are set to $\delta = 10, p = 0.3, n = 800, \epsilon = \frac{\sqrt{(160^2+200^2)}}{3}$ and $\kappa = 15$. The experiments are repeated 25 times for each size of the training set.

4 Results

The condition numbers of the complete data matrices were computed for all experiments using the largest and lowest singular values by $\frac{\sigma_1}{\sigma_k}$ where k is the rank at $\tau = 0.95$. The computed condition numbers are 423 for the IMM faces, 783 for the 3D lung surfaces and 95 for the deformation fields, showing that all problems are highly ill-conditioned, particularly the surface modeling. To validate the suitability of PIMC for our applications, random matrices of size 6000×1000 with condition number 1000 and rank 20 were generated as described in [14]. Figure 5 shows the average residual error and the required computation time for GROUSE, PIMC and ProPPA [17] and illustrates the suitability of the PIMC algorithm for that type of problems.

ROC-Like Analysis: $N = 10$ training samples of the IMM face database are used to compute a classical PCA model, a FEM-PCA model, and a patch-based model. Generalization and specificity of these shape models can be controlled by

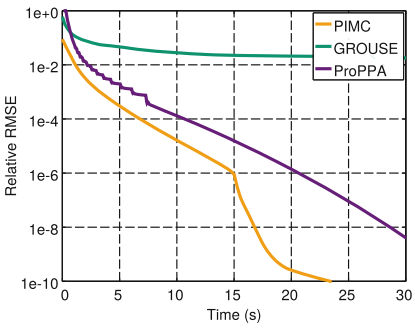


Fig. 5. Performance of three different matrix–completion algorithms for an ill–conditioned matrix (see text for details).

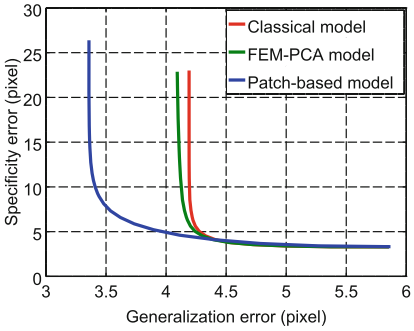


Fig. 6. ROC-like analysis: generalization and specificity for the IMM faces ($N = 10$) and varying variances of shape parameters (see text for details).

restricting the variances of the shape parameters. Figure 6 shows generalization and specificity for varying variances characterizing the dependencies between the two measures for the three different shape modeling approaches (average of 30 repeated experiments). The graph shows that for each given level of specificity our model clearly outperforms PCA and FEM-PCA models in terms of generalization, except for very low specificity errors where our model performs equal to the PCA model. For all remaining experiments we use the same model parameter restrictions for all shape models.

2D Contour Data of the IMM Face Database: Figure 7 (left) compares the generalization ability and specificity of the patch-based, the classical model, and FEM-PCA for varying numbers of available training samples. For small training sets ($N \leq 13$) the proposed method improves the generalization error by $\approx 22\,\%$, for larger training sets the improvement is less prominent ($\approx 13\,\%$). FEM-PCA shows no substantial improvement compared to the classical model for training sizes $N \leq 20$. As expected, the improvements in terms of generalization ability come along with higher landmark errors in the specificity tests. Both results reveal the desired increased variability of the model. However, model generated instances still represent valid face contours as indicated by an average specificity value of ≈ 10 pixels and as shown in Fig. 3.

3D Lung Surfaces: Figure 8 shows the mean surface and associated deformation modes generated from the patch-based model ($N = 10$ training shapes). Although only local information is provided to the model, it learns global deformations in the most important modes. Figure 7 (middle row) presents generalization ability and specificity depending on the number of training samples. The generalization ability is improved for small training sizes but approaches the classical model for $N \geq 20$. For this data, the localization error of the pseudo-landmarks is ≥ 2 mm due to a voxel spacing of 1.5 mm^3 of the underlying image data and an average registration error of ≈ 1.2 mm (see [28]). Therefore, by

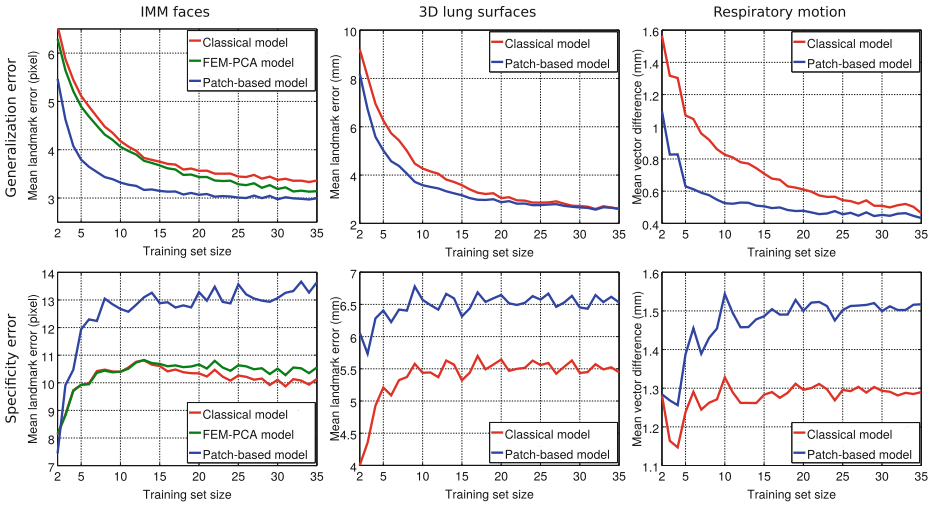


Fig. 7. Generalization and specificity errors for three experiments comparing the patch-based model (blue), the classical model (red), and the FEM-PCA model [6] (green, IMM faces only) given a varying numbers of training samples. Smaller values indicate better models. (Color figure online)

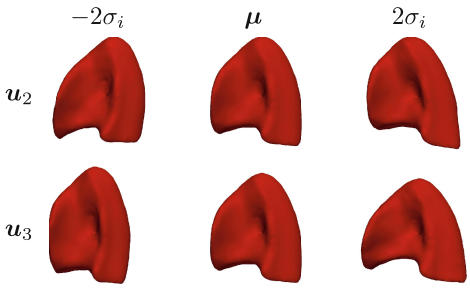


Fig. 8. Mean surface and two associated deformation modes of the 3D lung model generated by the proposed patch-based approach.

using the estimated rank of $k = 21$ a separation of noise and content in the data is obtained, and hence no further improvements in terms of adaption accuracy can be achieved without directly learning the noise.

Respiratory Lung Motion: In this application, the model obtains the most evident improvements compared with the classical model, particularly for small training sizes; e.g. an improvement of 41 % was achieved for $N = 5$, and 36 % for $N = 10$. The L-shape of the curve indicates that even small training sizes produce models with low approximation errors. The specificity measures for lung surfaces and lung motion in Fig. 7 reflects that the dependency of the generated model on the randomly selected training samples is more prominent for small training sizes.

5 Discussion and Conclusion

The contribution of this paper is a patch-based approach for learning of representative statistical shape, appearance, and motion models from few training samples. Our approach is based on the assumption that local variations have limited effects in distant areas and the model generation is formulated as a low-rank matrix completion problem that can be efficiently solved using recent algorithms capable of handling ill-conditioned matrices. In contrast to other patch-based algorithms used to tackle the HDLSS problem, our approach learns a consistent generative model and contrary to hierarchical techniques, our approach does not rely on the explicit and non-trivial definition a hierarchy.

Our experiments show that the proposed method can be applied for a variety of problems and leads to an increased flexibility and generalization ability while the validity of generated model instances is preserved. We have furthermore shown that the chosen PIMC algorithm is well suited for our intended applications. Its ability to solve the MC problem online is a key advantage for large scale problems as arising in deformation modeling where the data needed oftentimes exceeds the available memory. A disadvantage of this approach is the need to provide an estimated rank, however, in our experience an accurate choice is not crucial. It has to be noted that the presented patch-based modeling approach is not restricted to a specific MC algorithm, and any method that can handle ill-conditioned matrices can be used, e.g. ProPPA [17] (see Fig. 5). An important part of our approach is the patch selection strategy because it affects the generated model directly. But, even simple strategies lead to improved performance compared to other modeling approaches as shown in our experiments. Furthermore, application-specific prior-knowledge such as a left-right symmetry or other non-local relations could easily be incorporated.

An investigation of the computed variation modes revealed that although only local information was provided to our algorithm, the proposed method was able to learn the global shape variability as well (see Fig. 8). In this way, our approach combines local flexibility with well-known properties of classical models. In future work, we will further analyze the effects of the patch selection on the generated model. By using varying partitions (e.g., from large to small) during the virtual sample generation it should be possible to directly enforce the learning of a consistent hierarchical model with different levels of locality.

Acknowledgement. This work was supported by the German Research Foundation (DFG EH 224/6-1).

References

1. Armato, S.G., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. **38**(2), 915–931 (2011)
2. Balzano, L., Wright, S.J.: On GROUSE and incremental SVD. In: Computational Advances in Multi-sensor Adaptive Processing (CAMSAP), pp. 1–4 (2013)

3. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
4. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Found. Comput. Math. **9**(6), 717–772 (2009)
5. Cerrolaza, J.J., Reyes, M., Summers, R.M., Ballester, M.A.G., Linguraru, M.G.: Automatic multi-resolution shape modeling of multi-organ structures. Med. Image Anal. **25**(1), 11–21 (2015)
6. Cootes, T.F., Taylor, C.J.: Combining point distribution models with shape models based on finite element analysis. Image Vis. Comput. **13**(5), 403–409 (1995)
7. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
8. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. IEEE Trans. Med. Imaging **22**(3), 414–423 (2003)
9. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imaging **21**(5), 525–537 (2002)
10. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Handels, H.: Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. IEEE Trans. Med. Imaging **30**(2), 251–265 (2011)
11. Friedel, I., Schröder, P., Desbrun, M.: Unconstrained spherical parameterization. J. Graph. GPU Game Tools **12**(1), 17–26 (2007)
12. He, J., Balzano, L., Szlam, A.: Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. CVPR **2012**, 1568–1575 (2012)
13. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
14. Kennedy, R., Taylor, C.J., Balzano, L.: Online completion of ill-conditioned low-rank matrices. In: IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 507–511 (2014)
15. Koikkalainen, J., Tolli, T., Lauerma, K., Antila, K., Mattila, E., Lilja, M., Lotjonen, J.: Methods of artificial enlargement of the training set for statistical shape models. IEEE Trans. Med. Imaging **27**(11), 1643–1654 (2008)
16. Kumar, R., Banerjee, A., Vemuri, B.C., Pfister, H.: Maximizing all margins: pushing face recognition with kernel plurality. ICCV **2011**, 2375–2382 (2011)
17. Lai, R.Y., Yuen, P.C.: ProPPA: a fast algorithm for ℓ_1 minimization and low-rank matrix completion, May 2012. arXiv preprint arXiv:1205.0088
18. Leopardi, P.: A partition of the unit sphere into regions of equal area and small diameter. Electr. Trans. Numer. Anal. **25**(12), 309–327 (2006)
19. Lu, J., Tan, Y.P., Wang, G.: Discriminative multi-manifold analysis for face recognition from a single training sample per person. ICCV **2011**, 1943–1950 (2011)
20. Lüthi, M., Albrecht, T., Vetter, T.: Building shape models from lousy data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 1–8. Springer, Heidelberg (2009)
21. Mishra, B.: A Riemannian approach to large-scale constrained least-squares with symmetries. Ph.D. thesis, Université de Namur (2014)
22. Ngo, T., Saad, Y.: Scaled gradients on grassmann manifolds for matrix completion. NIPS **2012**, 1412–1420 (2012)
23. Nordstrøm, M.M., Larsen, M., Sierakowski, J., Stegmann, M.B.: The IMM face database-an annotated dataset of 240 face images. Technical report, Technical University of Denmark, DTU Informatics (2004)

24. Onofrey, J., Papademetris, X., Staib, L.: Low-dimensional non-rigid image registration using statistical deformation models from semi-supervised training data. IEEE Trans. Med. Imaging **34**(7), 1522–1532 (2015)
25. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models using non-rigid registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 77–84. Springer, Heidelberg (2001)
26. Shamir, A.: A survey on mesh segmentation techniques. Comput. Graph. Forum **27**(6), 1539–1556 (2008)
27. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: CVPR 1991, pp. 586–591. IEEE (1991)
28. Werner, R., Schmidt-Richberg, A., Handels, H., Ehrhardt, J.: Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: a comparison and evaluation study. Phys. Med. Biol. **59**, 4247–4260 (2014)
29. Zhang, D., Chen, S., Zhou, Z.H.: A new face recognition method based on SVD perturbation for single example image per person. Appl. Math. Comput. **163**(2), 895–907 (2005)
30. Zhao, Z., Aylward, S.R., Teoh, E.-K.: A novel 3D partitioned active shape model for segmentation of brain MR images. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 221–228. Springer, Heidelberg (2005)
31. Zhu, P., Zhang, L., Hu, Q., Shiu, S.C.K.: Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part I. LNCS, vol. 7572, pp. 822–835. Springer, Heidelberg (2012)