Medical Image Analysis 17 (2013) 959–973

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

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

Posterior shape models

Thomas Albrecht *, Marcel Lüthi, Thomas Gerig, Thomas Vetter

Department of Mathematics and Computer Science, Universität Basel, Bernoullistrasse 16, 4056 Basel, Switzerland

ARTICLE INFO

Article history:
Received 31 August 2012
Received in revised form 16 May 2013
Accepted 27 May 2013
Available online 15 June 2013

Keywords:
Statistical shape model
Conditional shape model
Posterior shape model
Image segmentation
Trochlear dysplasia

ABSTRACT

We present a method to compute the conditional distribution of a statistical shape model given partial data. The result is a "posterior shape model", which is again a statistical shape model of the same form as the original model. This allows its direct use in the variety of algorithms that include prior knowledge about the variability of a class of shapes with a statistical shape model. Posterior shape models then provide a statistically sound yet easy method to integrate partial data into these algorithms. Usually, shape models represent a complete organ, for instance in our experiments the femur bone, modeled by a multivariate normal distribution. But because in many application certain parts of the shape are known a priori, it is of great interest to model the posterior distribution of the whole shape given the known parts. These could be isolated landmark points or larger portions of the shape, like the healthy part of a pathological or damaged organ. However, because for most shape models the dimensionality of the data is much higher than the number of examples, the normal distribution is singular, and the conditional distribution not readily available. In this paper, we present two main contributions: First, we show how the posterior model can be efficiently computed as a statistical shape model in standard form and used in any shape model algorithm. We complement this paper with a freely available implementation of our algorithms. Second, we show that most common approaches put forth in the literature to overcome this are equivalent to probabilistic principal component analysis (PPCA), and Gaussian Process regression. To illustrate the use of posterior shape models, we apply them on two problems from medical image analysis: model-based image segmentation incorporating prior knowledge from landmarks, and the prediction of anatomically correct knee shapes for trochlear dysplasia patients, which constitutes a novel medical application. Our experiments confirm that the use of conditional shape models for image segmentation improves the overall segmentation accuracy and robustness.

1. Introduction

Statistical shape models have become an indispensable tool in medical image analysis. In essence, statistical shape models can be seen as a probability distribution (usually a normal distribution), which assigns the anatomically normal shapes of an anatomical structure a high probability, while pathological and other shapes that do not correspond to the modeled anatomical structure are assigned a low probability. Their power and versatility can be explained by the fact that they provide a quantitative answer to two fundamental questions in medicine: *(1) How does a normal instance of a given anatomical structure look like? (2) Is a specific anatomical structure normal or pathological?* Statistical shape models thus allow us to develop algorithms whose solution space is restricted to anatomically normal shapes. Such a strong prior on the solution makes the algorithm more robust, leads to easier optimization problems, and even allows us to infer a solution when only partial data is given. Consequently, applications such as implant design, surgery planning, or even medical image segmentation, for which it is clear that the result has to be a normal shape, have been shown to greatly benefit from the use of shape models (Heimann and Meinzer, 2009). In this paper we show how we can build a statistical shape model that even better restricts the solution space for the case when a part of the solution shape is already known. The new model answers the question: *Given a part of an anatomical structure, how does a normal instance of the full shape look like?* Knowledge of a part of the structure is often immediately available in practice. In surgery planning for example, a part of a shape may be missing due to a trauma or tumor, but the remaining part of the shape is known to be intact. It is thus a priori known that the solution needs to correspond to the shape of the part that is still intact. Another typical scenario is that a number of landmark points are available, which need to be matched by an algorithm.

In the following we sketch the main idea behind our method: A PCA-based statistical shape models is a generative model of the form:

$$\mathbf{s} = \mathbf{s}(\boldsymbol{\alpha}) = \boldsymbol{\mu} + \mathbf{Q}\boldsymbol{\alpha}, \tag{1}$$

* Corresponding author. Tel.: +41 61 267 0442.
E-mail address: thomas.albrecht@unibas.ch (T. Albrecht).

1361-8415/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2013.05.010

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a vector that represents the mean shape and $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_n) \in \mathbb{R}^{p \times n}$ is a matrix of principal components $\mathbf{q}_i$, derived from training examples. By assuming that the coefficients $\boldsymbol{\alpha} \in \mathbb{R}^n$ in (1) follow a standard normal distribution $p(\boldsymbol{\alpha}) \sim \mathcal{N}(0, \mathbf{I}_n)$, a probability distribution $p(\mathbf{s}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}\mathbf{Q}^T)$ is induced on the shape space. For a known, given part of a shape $\mathbf{s}_g \in \mathbb{R}^q$, we wish to compute a new (normal) distribution $p(\boldsymbol{\alpha}|\mathbf{s}_g) \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Lambda})$. Using this distribution as a model for the coefficients $\boldsymbol{\alpha}$ in (1) yields a new shape model, which represents shapes whose fixed part corresponds to $\mathbf{s}_g$. The new model, whose mathematical form is identical to that of the original model, can thus be used to strengthen the prior assumptions for any method that uses shape models.

In most shape models, the number of examples n is less than the dimensionality of the shape space p. This makes the covariance matrix $\mathbf{Q}\mathbf{Q}^T$ and therefore the normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}\mathbf{Q}^T)$ singular and the conditional distribution is more difficult to compute than it seems at first glance. We follow the most common approach from prior work, which is to regularize the covariance matrix, or the part of it corresponding to the given data, by adding a small multiple of the identity matrix $\sigma^2\,\mathbf{I}$. This can be interpreted as modeling the noise or deviation from the model in the partial data. We show the connection of this approach to probabilistic principal component analysis (PPCA) and Gaussian process regression.

We demonstrate two prototypical application of this model in medical applications. The first application targets an atlas-based segmentation of the femur bone from CT images using statistical model fitting. Here, a sparse set of landmark points is used to constrain the shape space, and thus simplify the actual fitting task. The second application targets operation planning for trochlear dysplasia patients. Trochlear dysplasia is a deformity of the knee that is treated surgically by remodeling the joint surface. Our application demonstrates how the shape model can be used to infer the normal shape of the pathological region from the intact part. This constitutes a novel application of (posterior) shape models for surgery planning.

In summary, we have the following main contributions: (1) We show how to efficiently compute the conditional distribution $p(\boldsymbol{\alpha}|\mathbf{s}_g)$ and the resulting posterior shape model, which is again a statistical shape model of the form (1). (2) We show the connection of this method to Probabilistic PCA (Tipping and Bishop, 1999) and Gaussian Process regression Rasmussen and Williams (2006). (3) We provide novel applications of our method to two problems in medical image analysis. (4) We provide a C++ implementation, as an integrated part of the freely available *statismo* library (Lüthi et al., 2012).[1]

1.1. Related work

Since their invention, statistical shape models have been used to infer the full shape from partial or "sparse" data. Often, only the maximum a posterior solution (MAP), i.e. the single most probable shape given the partial data is sought. Of the many papers computing the MAP, we only mention Blanz and Vetter (2002), as it is closest to this work. It uses a regularization term of the form $\sigma^2\mathbf{I}$ to compute a conditional distribution, but only computes the MAP and not the full posterior.

We are interested in computing this posterior model. In previous work (Albrecht et al., 2008), we derived a statistical model matching the given data using a heuristic method. In Lüthi et al. (2009), a similar model was more rigorously derived as the conditional probability of a PPCA formulation (Tipping and Bishop, 1999) of the statistical model given the partial data. The derivation of the conditional models we present here is similar, but it simplifies the formulation by separating the modeling of the partial data and the concept of PPCA models.

Other research groups have also investigated partially determined shape models. In Liu et al. (2004), canonical correlation analysis (CCA) is used to predict an unknown or diseased part of a shape from the healthy part. In Blanc et al. (2009) the given data is not a part of the shape, but given in the form of "surrogate variables" such as weight, sex, or age of a patient. In Blanc et al. (2012), this model is extended to also include partial shape data. In Blanc and Szekely (2012), the confidence of the reconstruction is evaluated, with a focus on including also the uncertainty involved in estimating correspondence between the given data and the model. These last two papers mention conditional shape models in the form we consider here in passing, but do not discuss the technical details or compute the actual shape model of the posterior distribution.

De Bruijne et al. (2007) compute a conditional shape model of a human vertebra given its neighboring vertebrae. They compute the conditional distribution with a regularization term of the form $\sigma^2\mathbf{I}$ and use the posterior shape model to classify fractures of the vertebrae. This posterior model seems very similar to our approach, but no details of its computation, especially for datasets larger than 2D vertebra shapes are given. In Baka et al. (2010) and Tomoshige et al. (2012) the simple regularization term $\sigma^2\,\mathbf{I}$ is replaced with a more general matrix reflecting the uncertainty for each given value individually. No explicit form of the posterior shape model is given in these papers. Their idea of replacing the regularization term can be employed in our approach, if individual uncertainty estimates for the given values are available. For our experiments, however, we use the standard regularization term.

Metz et al. (2010) use a combined model of shape and motion to infer cardiac motion from given shapes. They do not use a regularization term but compute the conditional distribution "after applying PCA", which amounts to simply projecting the given data onto the span of the example data and ignoring how far it actually is from this span. No posterior model is computed. Petersen et al. (2011) aim at computing the conditional distribution of a combined model of shape and rigid alignment, given partial data like landmark points. By including the rigid alignment, their conditional model becomes a non-linear manifold. This is then again linearized using a Laplace approximation (see Bishop (2006) for instance), in order to draw samples from the distribution. While this method has the advantage of incorporating the alignment into the model, no analytic expression of the model and no explicit posterior shape models are given.

To sum up, while all of these papers introduce some form "conditional model", the detailed derivation, explicit and efficient computation of the posterior shape model in the form of a standard shape model, are novel.

The viewpoint of interpreting a shape model as a Gaussian processes has been put forward by Joshi et al. (1997). A very comprehensive overview of their group's approach to shape modeling can be found in Grenander and Miller (1998). The use of Gaussian Process Regression for incorporating additional prior information, or the computation of conditional shape modes has to the best of our knowledge not been discussed, neither the connection to PPCA.

Regarding the applications we present in this paper, the surgical treatment of trochlear dysplasia is presented in Verdonk et al. (2005). In Pfirrmann et al. (2000) a statistical study of trochlear dysplasia is performed based on manual measurements of a few selected geometric criteria. The use of statistical shape model in this area is novel.

Statistical shape models have been used in the context of image segmentation since their invention, see Heimann and Meinzer (2009) for a recent and extensive review. In the terminology of

[1] available at: http://www.statismo.org.

the review, we use a volume-to-volume atlas matching scheme with a strict shape constraint to achieve segmentation. In this sense it is similar to the approach introduced by Davatzikos et al. (2002). While Cootes and Taylor (2001) included landmark positions, they simply added a penalty term and did not use a conditional model.

2. Background

In this section we introduce the basic concept underlying PCA-based statistical models. We then discuss posterior models and show the connections to Probabilistic PCA and Gaussian Process regression. While our method can be applied to any PCA-based statistical model, for ease of exposition and visualization, we will focus our discussion on PCA-based shape models, which includes the popular Active Shape Model (Cootes and Taylor, 1992) and Morphable Model (Blanz and Vetter, 1999) as special cases.

2.1. PCA models

Statistical shape models represent a class of shapes based on a set of representative example shapes. Let us assume that $n \in \mathbb{N}$ example shapes are given. In order to treat these shapes as vectors, they need to be brought into correspondence. While this is an important step in the construction of shape models, we assume that the reader is familiar with this concept. A short overview of how we interpret it can be found in Appendix A. Once the shapes are in correspondence, each shape can be represented as a vector $\mathbf{s}_i \in \mathbb{R}^p,\ i = 1, \ldots, n$. It is assumed that the class of shapes is a linear space in $\mathbb{R}^p$ and shapes can be formed by linear combinations of the example shapes. This allows the direct calculation of the sample mean $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{s}_i$ and covariance matrix $\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{s}_i - \boldsymbol{\mu})(\mathbf{s}_i - \boldsymbol{\mu})^T$ of the shape vectors. Statistically, the class of shapes is then typically modeled by the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A principal component analysis (PCA), which amounts to an eigenvalue decomposition of the covariance matrix $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}^2\mathbf{U}^T$ permits a compact representation of all the shapes in the model as:

$$s = s(\boldsymbol{\alpha}) = \boldsymbol{\mu} + \mathbf{U}\mathbf{D}\boldsymbol{\alpha} =: \boldsymbol{\mu} + \mathbf{Q}\boldsymbol{\alpha}. \tag{2}$$

This representation is convenient because the matrix $\mathbf{U}$, which contains the eigenvectors or "principal components" of $\boldsymbol{\Sigma}$ is orthonormal. Each eigenvector represents an independent characteristic shape variation of the shape class and the corresponding eigenvalues in $\mathbf{D}^2$ quantify their variance. Because the eigenvalues and their corresponding eigenvectors are typically ordered from largest to smallest, the first few principal components represent the "main modes of variation" of the shape class.

In principle we have $\mathbf{U}, \mathbf{D}, \mathbf{Q} \in \mathbb{R}^{p \times p}$, where p is the dimensionality of the shape vectors $\mathbf{s}_i$. Typically, p is the number of points used in the discretization times the space dimension. For most shape models used in practice, p is much larger than the number of example shapes n. Therefore, the rank of $\boldsymbol{\Sigma}$ is never larger than n. Thus, its last $p - n$ eigenvalues are zero, and the matrices can, without loss of information, be truncated to $\mathbf{U}, \mathbf{Q} \in \mathbb{R}^{p \times n}$ and $\mathbf{D} \in \mathbb{R}^{n \times n}$.

If $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated as the sample mean and covariance, the rank of the covariance matrix diminishes by one, so we have at most $n - 1$ non-zero eigenvalues, and some authors even choose to omit the smallest ones, which represent the least variance of the model.

The shape models described here are often called "PCA models". However, the use of PCA is not essential. In principle, Eq. (2) simply describes a generative linear model and other combinations of the vector $\boldsymbol{\mu}$ and the matrix $\mathbf{Q}$ that model the variability of the class of shapes are possible.

If the coefficient vectors $\boldsymbol{\alpha}$ are distributed according to $\mathcal{N}(0, \mathbf{I}_n)$, the shapes $s(\boldsymbol{\alpha}) = \boldsymbol{\mu} + \mathbf{Q}\boldsymbol{\alpha}$ are distributed according to

$$\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}\mathbf{Q}^T) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{3}$$

2.2. The singular distribution

At first glance, it seems as if the model (3) defines a valid probability distribution on the space of all possible shape deformations $\mathbb{R}^p$. Upon closer inspection however, it becomes apparent that all the probability mass is concentrated on the n-dimensional space spanned by the columns of $\mathbf{Q}$, which is the span of the example shapes of the model. In other words, because $\boldsymbol{\Sigma}$ does not have full rank if $n < p$, the distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is singular. In many applications of statistical shape models, it suffices to work entirely in the span of the examples. For other applications however, like the conditional distributions computed in this paper, this will not be sufficient.

A small hypothetical experiment can illustrate why the singular distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is not sufficient to model shapes outside of the span of the examples. Suppose we leave out one of the n example shapes and build the shape model from the remaining $n - 1$ examples. Because the left out example is from the same class as the other shapes, it will, as a vector in $\mathbb{R}^p$, be very close to the remaining shapes, but it will in general not lie exactly *in* their span. Therefore, the probability density function of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ assigns a value of 0 to this shape, although we know that it is in fact a member of the object class that the model should represent, and would intuitively expect a high value.

In their paper on covariance estimation, Schäfer and Strimmer (2005) offer a good theoretical explanation for this problem. The sample covariance $\mathbf{Q}\mathbf{Q}^T$ is an estimator for the true covariance of the object class we wish to model. Unfortunately, if the number of examples n is smaller than the dimension of the space p, it is a very poor estimator of the true covariance: It is singular and assigns a probability density of 0 to most samples from the true distribution, like the left out example from above.

Many authors treat this problem by considering not the shape vector but its projection onto the span of the other examples, e.g. Metz et al. (2010). This projection would then have a high value in the probability density of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. However, its distance from the span is ignored. In some scenarios this may be appropriate, in others it can lead to gross misinterpretations.

Therefore, the aim is clear: We need to employ a statistical model that assigns a high probability to shapes that are in or near the span of the examples and a low, but not zero, probability to shapes further away. We begin by considering the problem for partial data.

3. Posterior PCA models

Our aim is to compute conditional distributions given partial information about the shapes. This partial information can for instance be a (healthy) part of a bone, which shall be used to predict other missing or damaged parts. Or the partial information can consist of a few isolated landmark points that have been supplied by a user to aid in fitting the model to difficult data. For now, we assume that this partial information is given in the form of $q < p$ known entries in the shape vector $\mathbf{s}$, typically we have $q > n$. We will denote the given entries as $\mathbf{s}_g \in \mathbb{R}^q$. Similarly we can define the sub-vector $\boldsymbol{\mu}_g \in \mathbb{R}^q$ and sub-matrix $\mathbf{Q}_g \in \mathbb{R}^{q \times n}$ by selecting those entries and rows from the model's full $\boldsymbol{\mu}$ and $\mathbf{Q}$ that correspond to the given entries.

As outlined above, unless the given data is directly derived from the example data for the model, it will in general not lie exactly in

the span of the model, i.e. there is no $\boldsymbol{\alpha}$ such that $\mathbf{s}_g = \boldsymbol{\mu}_g + \mathbf{Q}_g\boldsymbol{\alpha}$ and the singular distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or rather $\mathcal{N}\left(\boldsymbol{\mu}_g, \mathbf{Q}_g\mathbf{Q}_g^T\right)$ cannot be used to interpret the partial shape directly. Therefore, we propose to explicitly model the distance of the data set $\mathbf{s}_g$ from the model space with a "noise" or "slack" variable $\varepsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I}_q)$ with a small variance of σ^2. I.e. we represent the given data as:

$$\mathbf{s}_g = \boldsymbol{\mu}_g + \mathbf{Q}_g\boldsymbol{\alpha} + \varepsilon. \tag{4}$$

These shapes are therefore distributed according to:

$$\mathcal{N}(\boldsymbol{\mu}_g, \mathbf{Q}_g\mathbf{Q}_g^T + \sigma^2\mathbf{I}_q). \tag{5}$$

This approach of adding ε can be considered standard. In terms of Schäfer and Strimmer (2005) it is an improved covariance estimation by adding the *shrinkage term* $\sigma^2\mathbf{I}_q$. In other areas of mathematics the same approach is known as ridge regression, Tykhonov regularization etc. It has been used in connection with statistical shape models almost since their inception. In connection with conditional distributions it is used by the papers mentioned in the Related Work Section 1.1. A strategy for choosing the variance parameter σ^2 will be given in Section 3.3. The approach can be generalized by allowing other distributions for ε that model the noise or deviation more specifically, e.g. individually for each point (Tomoshige et al., 2012; Baka et al., 2010). For ease of exposition, we use the standard uniform shrinkage term here, but all formulas can be directly generalized by replacing the identity matrix.

Eq. (5) describes a non-singular (!) distribution on the space $\mathbb{R}^q$ of all partial data vectors $\mathbf{s}_g$, which allows us to compute with standard methods the conditional distribution $p(\boldsymbol{\alpha}|\mathbf{s}_g)$ of the model parameters $\boldsymbol{\alpha}$, given the partial information $\mathbf{s}_g$. Because the shape parameters represent full shapes $\mathbf{s}(\boldsymbol{\alpha})$ this also defines the conditional distribution $p(\mathbf{s}|\mathbf{s}_g)$ of the full shapes given the partial data.

Because the distribution of the parameters $\boldsymbol{\alpha}$ and the conditional distribution of the observed partial data given the parameters are both normal,

$$p(\boldsymbol{\alpha}) = \mathcal{N}(0, \mathbf{I}_n), \tag{6}$$

$$p(\mathbf{s}_g|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\mu}_g + \mathbf{Q}_g\boldsymbol{\alpha}, \sigma^2\mathbf{I}_q), \tag{7}$$

the conditional distribution $p(\boldsymbol{\alpha}|\mathbf{s}_g)$ is again a multivariate normal distribution which can be calculated using Bayes' rule. The calculations leading to the explicit form of the mean and covariance matrix of this distribution are rather lengthy (For a full derivation see e.g. Bishop (2006), Chapter 2.3) Here, we present the final expression:

$$p(\boldsymbol{\alpha}|\mathbf{s}_g) = \mathcal{N}\left(\mathbf{M}^{-1}\mathbf{Q}_g^T(\mathbf{s}_g - \boldsymbol{\mu}_g), \sigma^2\mathbf{M}^{-1}\right) \tag{8}$$

$$=: \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Lambda}) \tag{9}$$

$$\text{where } \mathbf{M} = \mathbf{Q}_g^T\mathbf{Q}_g + \sigma^2\mathbf{I}_n. \tag{10}$$

For the full shapes $\mathbf{s}(\boldsymbol{\alpha})$ defined in Eq. (2), we therefore have the posterior distribution:

$$p(\mathbf{s}|\mathbf{s}_g) = \mathcal{N}(\boldsymbol{\mu} + \mathbf{Q}\ \boldsymbol{\eta}, \mathbf{Q}\ \boldsymbol{\Lambda}\ \mathbf{Q}^T) =: \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_C). \tag{11}$$

Written out, we have the posterior mean:

$$\boldsymbol{\mu}_c = \boldsymbol{\mu} + \mathbf{Q}\ (\mathbf{Q}_g^T\mathbf{Q}_g + \sigma^2\mathbf{I}_n)^{-1}\mathbf{Q}_g^T\ (\mathbf{s}_g - \boldsymbol{\mu}_g) \tag{12}$$

and covariance:

$$\boldsymbol{\Sigma}_c = \sigma^2\mathbf{Q}(\mathbf{Q}_g^T\mathbf{Q}_g + \sigma^2\mathbf{I}_n)^{-1}\mathbf{Q}^T. \tag{13}$$

3.1. Correspondence and alignment

Mathematically, this seems very straight-forward but a few clarifications are in order. In order to select the appropriate sub-vector and matrix $\boldsymbol{\mu}_g$, $\mathbf{Q}_g$, the given points need to be *in correspondence* with the shapes. This means that we need to know which entries in $\mathbf{s}_g$ belong to which entries in $\mathbf{s}$ in order to select the correct entries and rows from $\boldsymbol{\mu}$, $\mathbf{Q}$. For more details on correspondence, see Appendix A.

Furthermore, we assume that the given points are correctly *aligned* to the shape model, because in Eq. (8) we model only shape variations and not variations in rigid alignment. These prerequisites are not always given in general, correspondence and alignment of the given points to the shape model may need to be computed beforehand. The most obvious approach is to do this, is to *fit* the model to the given data, i.e. to find the optimal rigid alignment and model parameters so that the fitted model shape (almost) coincides with the given data. This raises another problem: As Baka et al. (2010) point out, the alignment of the given data should be consistent with that of the original model. Therefore, as they propose, we realign the original model based only on the parts of the shape that correspond to the given data $\mathbf{s}_g$. This alignment may be sub-optimal with respect to the full shapes, but is consistent with the given data. A more mathematical discussion of this problem is given in Section 4.3.

3.2. Principal components of the posterior model

The distribution of the conditional shape model Eq. (11) defines a multivariate normal distribution of shapes, represented by vectors in $\mathbb{R}^p$. In this sense, it is simply a statistical shape model. We have explicit representations of the mean and covariance of this distribution, $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$. In many applications, the mean is all we are interested in, because it represents the most likely shape given the partial data. Often though, we would like to visualize and represent the shape model in terms of its most prominent modes of variation, i.e. its principal components. For this we can perform an additional eigenvalue decomposition of the covariance matrix

$$\boldsymbol{\Sigma}_c = \sigma^2\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q}^T. \tag{14}$$

In principal, the eigenvalue decomposition could be computed directly, but this is a $(p \times p)$ matrix, which may be *very* large (for instance 10^{12} entries for a typical shape model), and the decomposition may not be tractable. With the specific expression for $\mathbf{Q} = \mathbf{UD}$ from Eq. (2) however, we can reformulate and perform a PCA of the much smaller inner matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{n\times n}$:

$$\boldsymbol{\Sigma}_c = \mathbf{U}\underbrace{\sigma^2\mathbf{D}\ \mathbf{M}^{-1}\ \mathbf{D}}_{=:\boldsymbol{\Sigma}_i = \mathbf{U}_i\mathbf{D}_i^2\mathbf{U}_i^T}\ \mathbf{U}^T \tag{15}$$

to arrive at the final PCA:

$$\boldsymbol{\Sigma}_c = (\mathbf{U}\mathbf{U}_i)\ \mathbf{D}_i^2\ (\mathbf{U}_i\mathbf{U})^T \tag{16}$$

more efficiently. The new matrix of principal components are then given as $\mathbf{Q}_c = (\mathbf{U}\mathbf{U}_i)\mathbf{D}_i$ and the shapes of the conditional model can be represented as $\mathbf{s}_c(\boldsymbol{\alpha}) = \boldsymbol{\mu}_c + \mathbf{Q}_c\boldsymbol{\alpha}$.

3.3. The choice of the parameter σ^2

As outlined in the previous sections, the parameters σ^2 models the deviation of the given data $\mathbf{s}_g$ from the model. This deviation may be due to noise in the data, but even for noise-free data it is necessary to avoid a singular distribution. In some cases, we can have a clear notion of what this parameter should be, e.g. the expected precision of user-placed landmarks in millimeters. In other cases it is hard to derive exact values for the parameters. For these, we derive a maximum likelihood estimation from the data.

Suppose we have an example data set $\mathbf{s}_g$ that we know to be part of a normal member of the object class modeled by the original statistical model. Then we can fit the model to this data to find

the best approximation of this data set within the model span $\hat{\mathbf{s}}_g$. Then, according to Eq. (7) the likelihood of σ^2 is:

$$L(\sigma) = \frac{1}{z}\exp\left(\mathbf{s}_g - \hat{\mathbf{s}}_g\right)^T \sigma^{-2}\mathbf{I}_p\left(\mathbf{s}_g - \hat{\mathbf{s}}_g\right). \tag{17}$$

Maximizing with respect to σ^2 gives the maximum likelihood solution:

$$\sigma^2 = \frac{1}{q}\|\mathbf{s}_g - \hat{\mathbf{s}}_g\|_2^2, \tag{18}$$

the mean squared residual between $\mathbf{s}_g$ and its best approximation in the model. This makes sense because, knowing that $\mathbf{s}_g$ is a member of the modeled object class, the residual represents the deviation we will most likely have to expect for other members. Obviously, using more examples gives a better estimation. For m examples we get, assuming independence:

$$\sigma^2 = \frac{1}{mq}\sum_{i=1}^{m}\|\mathbf{s}_{gi} - \hat{\mathbf{s}}_{gi}\|_2^2. \tag{19}$$

4. Alternative views: PPCA and Gaussian process regression

In this section, we give two alternate views of the same problem. First we show the connection of our approach to probabilistic PCA (PPCA), which essentially extends the idea of modeling the deviation from the model span to the complete data sets **s**. Then, we show the connection to Gaussian process regression, which allows us to treat the problem of inferring the missing parts of a shape as a standard regression problem.

4.1. PPCA

In Eq. (4) the given information $\mathbf{s}_g$ was modeled by a linear shape model plus a noise or slack variable $\varepsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I}_q)$. It is possible to model not only the partial shape but all complete shapes s in a similar manner, using a noise variable $\omega \sim N(0, \rho^2\mathbf{I}_p)$:

$$\mathbf{s} = \boldsymbol{\mu} + \mathbf{Q}\,\boldsymbol{\alpha} + \omega. \tag{20}$$

This leads to a generalization of PCA introduced as Probabilistic PCA (PPCA) by Tipping and Bishop (1999) or Sensible PCA by Roweis (1998). This is a direct generalization of Eq. (2), as by choosing a variance of $\rho^2 = 0$, we arrive at the standard PCA formulation.

PPCA can be motivated by assuming that every data set **s** may be afflicted by noise modeled by $\omega \sim N(0, \rho^2\mathbf{I}_p)$, which seems to be a reasonable assumption in medical imaging. The distribution of the model parameters is still assumed to be $\boldsymbol{\alpha} \sim N(0, \rho^2\mathbf{I}_n)$, and therefore the shape conditioned on the parameters is distributed as follows:

$$p(\mathbf{s}|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\mu} + \mathbf{Q}\boldsymbol{\alpha}, \rho^2\mathbf{I}_p). \tag{21}$$

For $\rho^2 \neq 0$ PPCA therefore defines a non-singular distribution

$$p(\mathbf{s}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}\mathbf{Q}^T + \rho^2\mathbf{I}_p). \tag{22}$$

on the space of all shapes $\mathbb{R}^p$, which can be advantageous in many situations. On the other hand, this forces us to choose the variance parameter ρ^2. This parameter may or may not be the same as the parameter σ^2 in Eq. (4). Both parameters model the amount of noise, or deviation from the model space. ρ^2 for the training data **s**, and σ^2 for given partial data $\mathbf{s}_g$. A maximum likelihood estimation for both parameters can be computed as described in Section 3.3, where, for the parameter ρ^2 a leave-one-out experiment with the model's input data can be performed, yielding the maximum likelihood estimator:

$$\rho^2 = \frac{1}{np}\sum_{i=1}^{n}\|\mathbf{s}_i - \hat{\mathbf{s}}_i\|_2^2. \tag{23}$$

In particular, for noise-free example data we can choose $\rho^2 = 0$ and $\sigma^2 > 0$, which brings us back to the formulation introduced in the previous section.

In the PPCA model, as before before, **Q** can be any $(p \times n)$-matrix describing the variability of the shapes. But typically it is the matrix of principal components computed by PCA. As somewhat of a technicality however, by computing a maximum likelihood estimator for **Q** from Eq. (22), we arrive not at the standard form $\mathbf{Q} = \mathbf{UD}$ from Eq. (2) but at the slightly modified version

$$\mathbf{Q}_{\mathrm{ML}} = \mathbf{U}\left(\mathbf{D}^2 - \rho^2\mathbf{I}_p\right)^{\frac{1}{2}}, \tag{24}$$

see Tipping and Bishop (1999); Roweis (1998) for details. While the PPCA model defines a non-singular distribution on the whole space $\mathbb{R}^p$, it is just as compact as a standard PCA model, because we still only need to store the (small) matrices **U** and **D** and the additional value ρ^2. Regarding the computation of the principal components of the posterior model described in the previous section, the inner matrix from Eq. (15) takes the slightly more complicated form:

$$\boldsymbol{\Sigma}_{\mathrm{i}} = \sigma^2\left(\mathbf{D}^2 - \rho^2\mathbf{I}_n\right)^{\frac{1}{2}}\mathbf{M}^{-1}\left(\mathbf{D}^2 - \rho^2\mathbf{I}_n\right)^{\frac{1}{2}}. \tag{25}$$

4.2. The Gaussian process view

For many application, it is convenient to view statistical shape models as a Gaussian Process that defines probabilistic model over the displacements from a mean shape. This interpretation of statistical shape models goes back at least to Grenander and Miller (1998). Unfortunately it has not been widely adopted in the community, despite the fact that it provides a conceptually clean view on shape models and allows for the application of the many powerful techniques and results from statistics and machine learning (see e.g. Rasmussen and Williams, 2006).

In order to formulate the Gaussian Process view, we need to change the notation slightly. So far we have assumed that a shape is represented as a vector $\mathbf{s} \in \mathbb{R}^p$. For the Gaussian Process view, we interpret these vectors as a representation of discrete functions, defined on a finite domain $\Omega = (x_1, \ldots, x_N)$. More specifically, as we have assumed correspondence between all the example vectors, we can define the function

$$s_i : \Omega \to \mathbb{R}^d$$

such that $s_i(x_j) \in \mathbb{R}^d$ refers to the jth point of the ith example shape. The mean vector $\boldsymbol{\mu}$ can be interpreted as a the mean function

$$\begin{aligned} &\mu : \Omega \to \mathbb{R}^d \\ &\mu(x) = \frac{1}{N}\sum_{i=1}^{N} s_i(x) \end{aligned} \tag{26}$$

and the covariance matrix becomes the covariance function

$$\begin{aligned} &\Sigma : \Omega \times \Omega \to \mathbb{R}^{d\times d} \\ &\Sigma(x, x') = \sum_{i=1}^{N}\left(s_i(x) - \mu(x)\right)\left(s_i(x') - \mu(x)\right)^T. \end{aligned} \tag{27}$$

As for the normal distribution, the mean and covariance functions uniquely define the Gaussian Process $\mathcal{GP}(\mu, \Sigma)$ (Rasmussen and Williams, 2006). From the covariance function (27) we see that a shape model actually defines a model over displacements $u_i(x) := s_i(x) - \mu(x)$. Identifying the domain Ω with the model mean μ, this yields the interpretation of shape models as a zero-mean Gaussian Process model, $GP(0, \Sigma)$ over displacement fields, defined on the mean shape μ.

4.2.1. Gaussian process regression

Using the interpretation of statistical models as a Gaussian Process, the prediction of a missing part of the shape becomes a standard regression problem. Let $s : \Omega \to \mathbb{R}^d$ be a shape from our model and assume that the shape is given at a subset of the points $\Omega_g \subset \Omega$ only. Additionally, we assume that we do not observe the shape directly but only a noisy versions, $t(x_i) = s(x_i) + \varepsilon$, with $\varepsilon = \mathcal{N}(0, \sigma^2)$. Our given data is then $\mathbf{t}_g = \{(x_i, t(x_i)) | x_i \in \Omega_g\}$, and our goal is to infer the complete function s from $\mathbf{t}_g$. Under the assumption that s is a Gaussian process and that the noise ε is uncorrelated Gaussian noise, this problem is known as Gaussian Process regression. Gaussian process regression is a standard inference technique that has become largely popular in machine learning (Rasmussen and Williams, 2006). It does not only allow us to compute the best prediction of s, but we obtain a full posterior distribution $(s | \mathbf{t}_g)$. This distribution is again a Gaussian Process $\mathcal{GP}(\mu_c, \Sigma_c)$ with its mean and covariance function given by:

$$\begin{aligned} \mu_c(x_i) &= \mu(x_i) + \boldsymbol{\Sigma}_g(x_i)^T (\boldsymbol{\Sigma}_{gg} + \sigma^2 \mathcal{I})^{-1} (\mathbf{t}_g - \boldsymbol{\mu}_g), \\ \Sigma_c(x_i, x_j) &= \Sigma(x_i, x_j) - \boldsymbol{\Sigma}_g(x_i)^T (\boldsymbol{\Sigma}_{gg} + \sigma^2 \mathcal{I})^{-1} \boldsymbol{\Sigma}_g(x_j). \end{aligned} \tag{28}$$

Here we define the following block matrices:

$$\boldsymbol{\Sigma}_g(x_i) = (\Sigma(x_i, x_j))_{x_j \in \Omega_g} \in \mathbb{R}^{q \times d}, \tag{29}$$

$$\boldsymbol{\Sigma}_{gg} = (\Sigma(x_i, x_j))_{x_i, x_j \in \Omega_g} \in \mathbb{R}^{q \times q}, \tag{30}$$

$$\boldsymbol{\mu}_g = (\mu(x_j))_{x_j \in \Omega_g} \in \mathbb{R}^{q \times d}. \tag{31}$$

While the formulas look rather different to the one obtained for the PCA solution (11), the solution is equivalent. The (somewhat lengthy) proof of this equivalence is given in Appendix B. We have thus shown that all three views on the problem, PCA with regularization, PPCA and Gaussian process regression are equivalent. In particular, this means that all similar problems in the literature regarding partial shapes can be interpreted as regression problems.

4.3. Alignment of the data

In Section 4.2 (Eq. (27)) we noted a statistical model can be interpreted as a zero mean Gaussian process $GP(0, \Sigma)$ that models the displacements $u(x) := s(x) - \mu(x)$ from the mean shape μ to a shape s. The distribution of these displacements is estimated from training data. Obviously this model will only yield good predictions for new shapes if the displacements that are observed for new shapes follow the same distribution as observed in the training data. This can only be the case if the alignment of the data is consistent between training and test data. Thus, given a shape s_g that is defined at only a subset Ω_g, our method will only yield meaningful results if all training shapes have been aligned with respect to the same points. To ensure this, we follow the strategy proposed by Baka et al. (2010), and simplly realign all the training shapes and compute a new model for each set of observed points Ω_g.

5. Model-based segmentation of femur bones from CT images

In this and the next section, we show the application of the conditional models to two typical problems that arise in surgery planning: The segmentation of a bone from a CT image, and the prediction of anatomically correct reconstructions of damaged or malformed parts of a bone.

5.1. Atlas based segmentation using a statistical deformation model

Although bone segmentation from CT images seems like a straight-forward problem and has been addressed by many researchers over the years, it remains difficult to this day. Fig. 1 shows two typical issues that arise in most bone segmentation problems, illustrated on the femur bone. At the top of the bone, it is difficult to separate the femur from the hip bone and at the bottom, the bone is so thin that the image intensity in this area is indistinguishable from that of soft tissue. While the image comes from a simple threshold segmentation, the same problems plague even the most sophisticated segmentation approaches. In low resolution images, these issues are aggravated by the partial volume effect.

A popular approach to achieve a segmentation even in these difficult regions is to include prior knowledge about the bones in form of a statistical model (Heimann and Meinzer, 2009). By constraining the segmentation result to the span of the model, it is ensured that the segmented object is a valid femur bone, thus preventing the inclusion of adjacent bones or the omission of thin parts. The segmentation approach we use is based on this idea. Instead of a statistical shape model however, we use a statistical deformation model (Rueckert et al., 2003). This model is built from deformation fields obtained by registration of manually segmented images of femur bones. We refer to A.1 for details of this procedure. Using a deformation model instead of a shape model allows us to formulate the model fitting in terms of an image registration problem. This approach is commonly referred to as "atlas matching". The output is a deformation field, which can then be used to transfer a high quality manual segmentation of the reference to the input image. Again, we refer the reader to A.2 for mathematical details.

5.2. Experimental setup

The experimental setup for this section is as follows. For building the statistical model we use the Statismo framework (Lüthi et al., 2012). The model fitting is performed using the Elastix registration framework (Klein et al., 2010), which was extended to allow for transformations defined by a statistical model. In all the experiments, we use the mutual information metric to quantify image similarity. The registration is performed using a 3 level multi-resolution scheme. As an optimizer we use stochastic gradient descent. The landmark uncertainty (i.e. σ in Eq. (3)) is set to 2 mm. This reflects our perceived accuracy in placing the landmarks.

5.3. Femur segmentation with a posterior deformation model

The performance of atlas based segmentation crucially depends on the accuracy of the registration between the atlas and the image to be segmented. Using a statistical deformation model that is built for a specific anatomical structure helps to restrict the registration results to deformations that are anatomically meaningful. However, in practice there are often cases where even this prior does not provide enough information to accurately identify the structure in the image. Such an example is shown in Fig. 3. The red contour shows a result of fitting a deformation model for the segmentation of a femur from a CT image. Overall, the shape is well

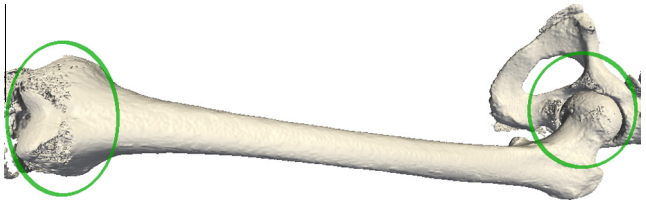


Fig. 1. A surface reconstruction of a segmentation obtained by thresholding a CT image. The pelvis cannot be separated from the femur (right) and the surface shows holes at the condyles (left).

Fig. 2. The 6 landmarks defined on the femur.

Fig. 4. Example of a CT image of an isolated femur bone that was used for the quantitative tests.

matched. However, we see that the femoral head is not accurately segmented. The proposed red outline is smaller than the bone's actual outline. This happens because the intensity differences in this area are very small. To overcome this problem we can further strengthen our prior assumptions and incorporate known correspondences into the deformation model, by means of a posterior model.

In order to build the posterior model, we define 6 landmark points $L_R = \{p_1^R, \ldots, p_n^R\}$ on the reference image that was used to build the deformation model (Fig. 2). The corresponding landmark points $L_T = \{p_1^T, \ldots, p_n^T\}$ are then marked on the target image. Using these landmarks, we can build a posterior deformation model $p(u|L_T, L_R) \sim \mathcal{N}(\mu_c, \Sigma_c)$, employing the method described in Section 3. As the posterior model is again a deformation model, it can be used for model fitting instead of the unconstrained model, without requiring changes in the procedure. Besides providing a stronger prior, using the posterior model has two further advantages: (1) The mean μ_c provides a superior initialization to the problem, which already matches the landmark points. (2) As the model effectively excludes solutions that do not match the landmarks, the search space is greatly reduced, leading to a simpler optimization problem.

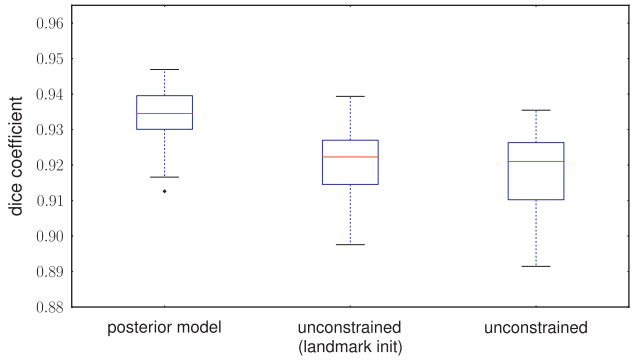
Fig. 3 shows the fitting result when the posterior model is used (green contour). We observe that the fit is greatly improved at the femoral head, as the segmentation outline is forced to adhere to the bone's true outline.

5.4. Quantitative results

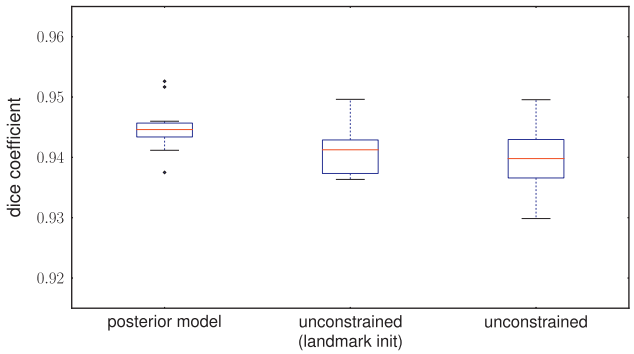
To compare the performance of the posterior model with normal deformation models in this segmentation task, we performed a quantitative evaluation on CT images of two sets of femur bones. (1) A set of 27 surgically extracted femur bones (see Fig. 4 for a typical example). (2) A set of 13 femurs with surrounding tissue, as in Fig. 3. For both sets, manual ground truth segmentations are available for comparison, and as partial information for the posterior model, we manually labeled the 6 landmark points seen in Fig. 2.

The original deformation model is built from 114 manually segmented femur bones. To ensure that this model is representative for the deformations observed for the test images, we rigidly aligned all the femurs using the same 6 landmarks that were defined on the test images (cf. Section 4.3). We applied the deformation model fitting with the original model, the posterior model, and the original model initialized with the posterior mean, which amounts to a landmark-initialization of the original model. For each method we computed the dice coefficient between the result and the ground-truth segmentation. Fig. 5 shows the fitting result for both test sets. It can be seen that fitting with a posterior model (left plot) not only yields a better average accuracy, but also to less variability in the results, i.e. the method becomes more robust. We observe that only initializing the fitting with the landmark, but not constraining the model (middle plot) does not give a clear advantage compared to a normal unconstrained fitting (right plot). It appears as if the effect of the landmark-initialization is undone by the fitting procedure. To further investigate this behavior, we evaluated the initialization and landmark error for our larger test set (Fig. 6). Fig. 6a shows that the landmark-initialization does indeed give a much better starting solution, and Fig. 6b shows that the landmark error after fitting of the landmark-initialized method is actually almost as good as that of the posterior model. So why is the overall performance measured in Fig. 5 hardly better than the unconstrained and uninitialized model, while the posterior model performs much better? This can be explained by the fact that the posterior model greatly constrains the model space, which positively influences the optimization procedure. Indeed, this conclusion is supported by Fig. 7, which shows the convergence of the

Fig. 3. Fitting a deformation model to a CT image of the femur. The red contour shows the fitting result using an unconstrained model, which fails to label the contour of the femoral head correctly. The green contour shows the result for a posterior model, where a landmark was used to enforce the position of the femoral head. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

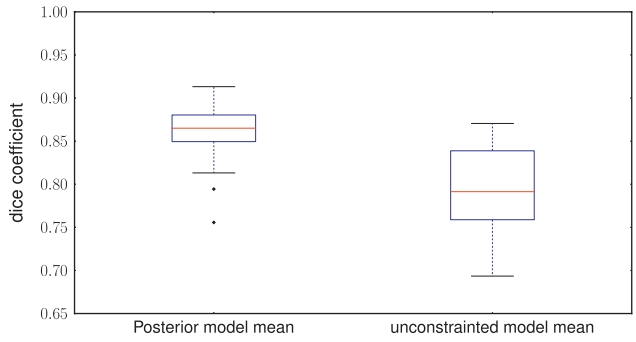


(a) Test set with images of isolated bones

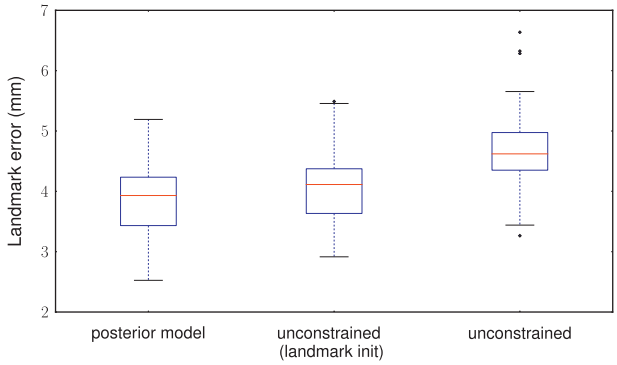


(b) Test set with clinical CT images

Fig. 5. Comparison of the segmentation performance (dice-coefficient) achieved by model fitting. (a) Shows the result for a set of 27 CT images of isolated femur bones and (b) shows the results for 13 clinical CT images.



(a) Initialization



(b) Landmark error

Fig. 6. Comparison of the initialization and landmark error of the different methods. (a) Shows the dice coefficient that is achieved by using the mean deformation of the respective models, before any optimization is performed. (b) Shows the error evaluated on the landmark points only. (Note that in (a) higher values are better, while in (b) lower values indicate better results.)

optimization algorithm averaged over all test cases, out of which the posterior model converges the most quickly.

5.5. Discussion

The results we presented in this section clearly show that posterior models can improve the results for model based segmentation tasks. The obvious application is to constrain the model with landmarks in cases where a fully automatic method fails to find the correct segmentation. Specifying a landmark point and constraining the model to remain fixed on this point provides an easy to use and effective means to incorporate user interaction into an algorithm. However, we also showed that using the posterior model improves the results even for relatively simple segmentation tasks, such as the fitting of dry femur bones. As the posterior model is a normal PCA model, it can readily be used to improve any algorithm that uses a statistical model.

6. Predicting anatomically correct shapes for trochlear dysplasia patients

In this experiment, we apply the conditional shape models to predict anatomically correct bone shapes for trochlear dysplasia patients. The term trochlear dysplasia describes a flattening of the trochlear groove. The trochlear groove, which is marked in Fig. 8, guides the patella during motion of the knee joint. If the groove is too shallow, a correct guiding of the patella is not guaranteed, leading to pain or even a dislocation of the patella. Trochlear dysplasia can be treated surgically by deepening the trochlea groove or augmenting its edges (Verdonk et al., 2005). Knowing the ideal shape of the trochlea would be of great advantage for the planning of such an intervention.

Specifically, we would like to use our model to help us answering the following questions:

1. Does the patient really exhibit an abnormal trochlea shape?
2. How should the trochlea be remodeled? How much should the groove be deepened and/or the edges augmented?
3. How reliable is the prediction?

For this, we use a statistical shape model of the distal femur and construct, individually for each patient, a conditional shape model

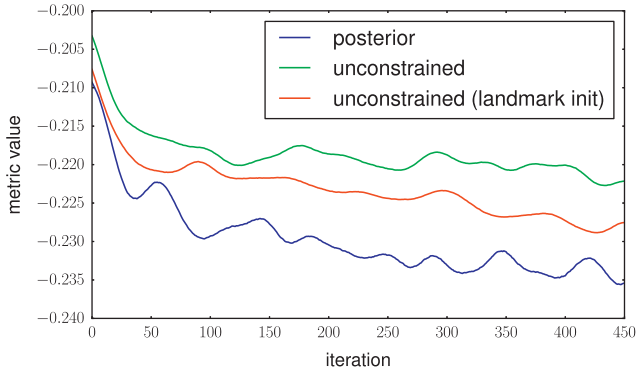


Fig. 7. Convergence of the fitting algorithm on the first resolution level. The curves show the average metric value for the mutual information metric, averaged over the test cases. The plots have been smoothed for clarity of visualization.

Fig. 8. Patella and femur. The trochlea is marked in red. If the trochlear groove, which guides the movement of the patella, is not deep enough, a dislocation of the patella is possible. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

given the healthy part, i.e. excluding the trochlea. The posterior shape model represents all anatomically correct trochlea shapes for the patient.

The children's hospital Basel provided us with a set of 29 MRI scans of trochlear dysplasia patients. All the datasets are clinical scans recorded during the daily routines of the hospital.

6.1. Experimental setup

The posterior shape models introduced in Section 2 rely on a previous shape model of the full shape. For this experiment, we use a statistical shape model of the distal part of the human femur built from 145 healthy example data sets. The model is built as a PPCA model with the parameter ρ estimated by a leave-out-experiment resulting in a value of $\rho = 0.26$. As explained in Section 3.1, the given data for the conditional shape model needs to be in correspondence with the original model. Therefore, we first produce a surface representation of the patient's femur by segmenting it from the MRI data set. This is brought into correspondence with the statistical shape model using an Iterative Closest Point (ICP) fitting (Feldmar and Ayache, 1996), which estimates the position and shape parameters $\boldsymbol{\alpha}$ of the model that best fit the segmented surface of the patient's femur. Due to large variations in the quality and image modality of the MRI scans, the segmentation is performed manually using the 3D Slicer software (Pieper et al., 2004) and not with the segmentation presented in the previous section.

As a next step, we wish to separate the afflicted trochlea region from the healthy part of the patient's bone, which will take the role of the given data $\mathbf{s}_g$. Thanks to the correspondence between the model and the patient's data set, the marking of the trochlea can be transferred automatically from a mask defined on the model. In Fig. 9 the two regions are visualized for an example case. The red trochlea region is then discarded and shall be predicted using the conditional shape model.

The ICP fitting used to establish correspondence with the model was performed using the complete data set including the malformed trochlea. In order to rule out any influence this may have on our prediction, we perform a second ICP fitting using only the healthy part $\mathbf{s}_g$. At this point we can finally apply the methods outlined in Section 2. The deviation parameter σ was estimated

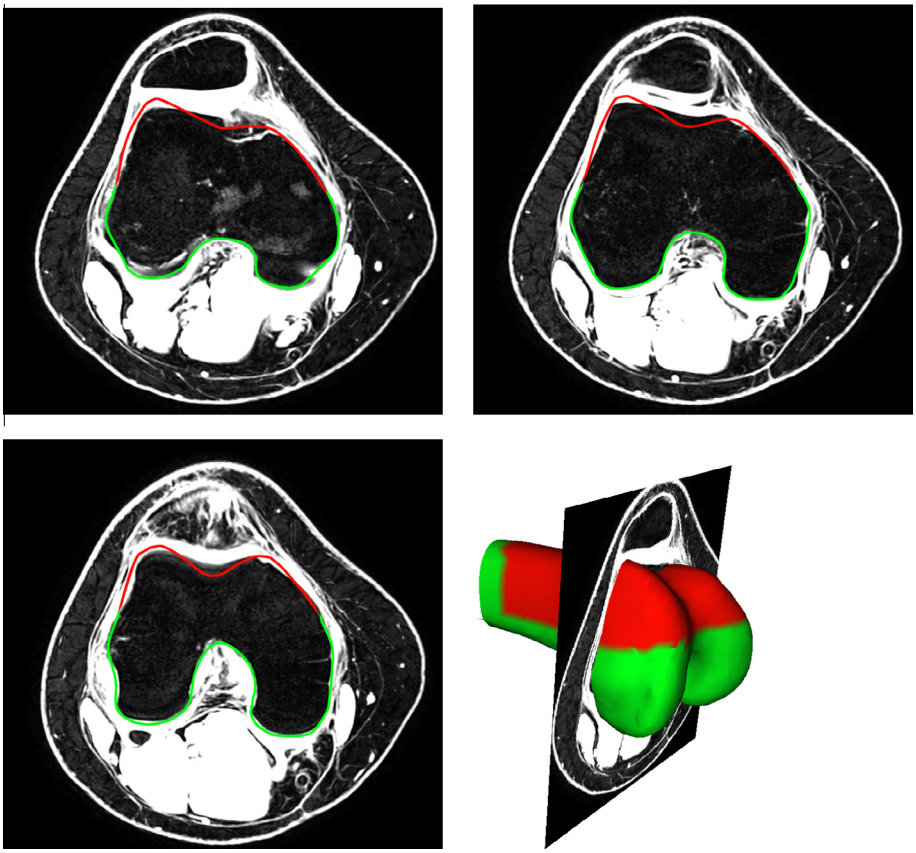


Fig. 9. Estimation of an anatomically correct femur shape. The green contour shows the given points $\mathbf{s}_g$, the red contour the most probable reconstruction $\boldsymbol{\mu}_c$ of the trochlea. It is clearly visible that the patient's original trochlea is too flat to hold the patella in place, whereas the proposed reconstruction has a pronounced groove to guide the patella. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

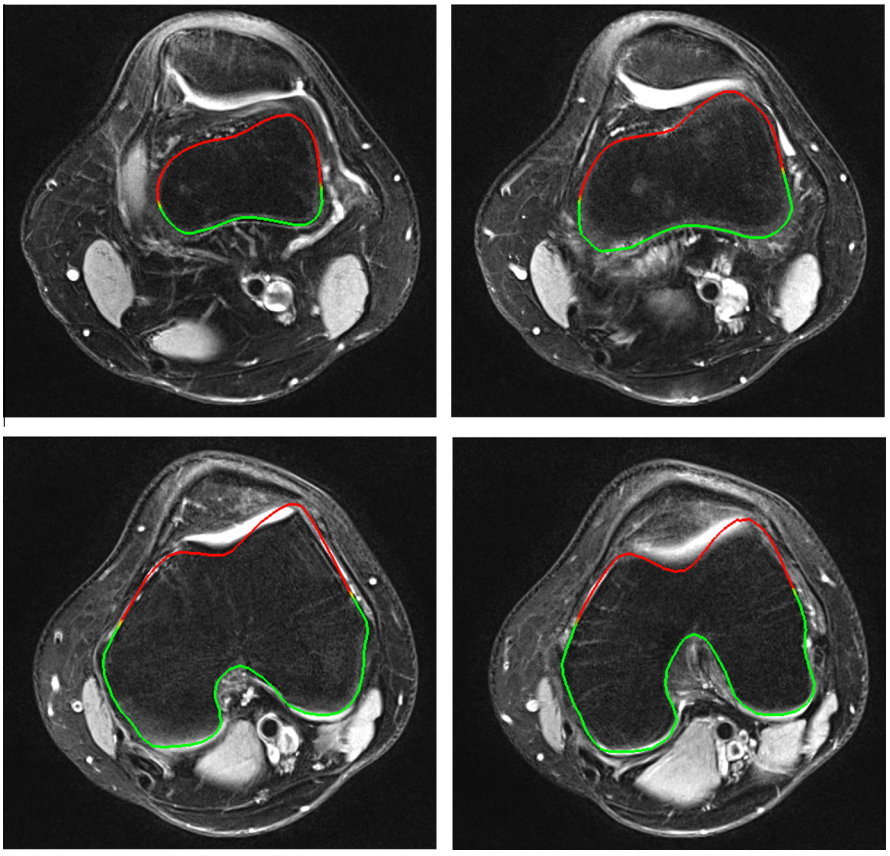


Fig. 10. A second example with given data $\mathbf{s}_g$ in green probable reconstruction μ_c in red. The green part of the contour is the given part $\mathbf{s}_g$. Again, the difference between the patient's flat trochlea and the proposed anatomically correct reconstruction is clearly visible. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

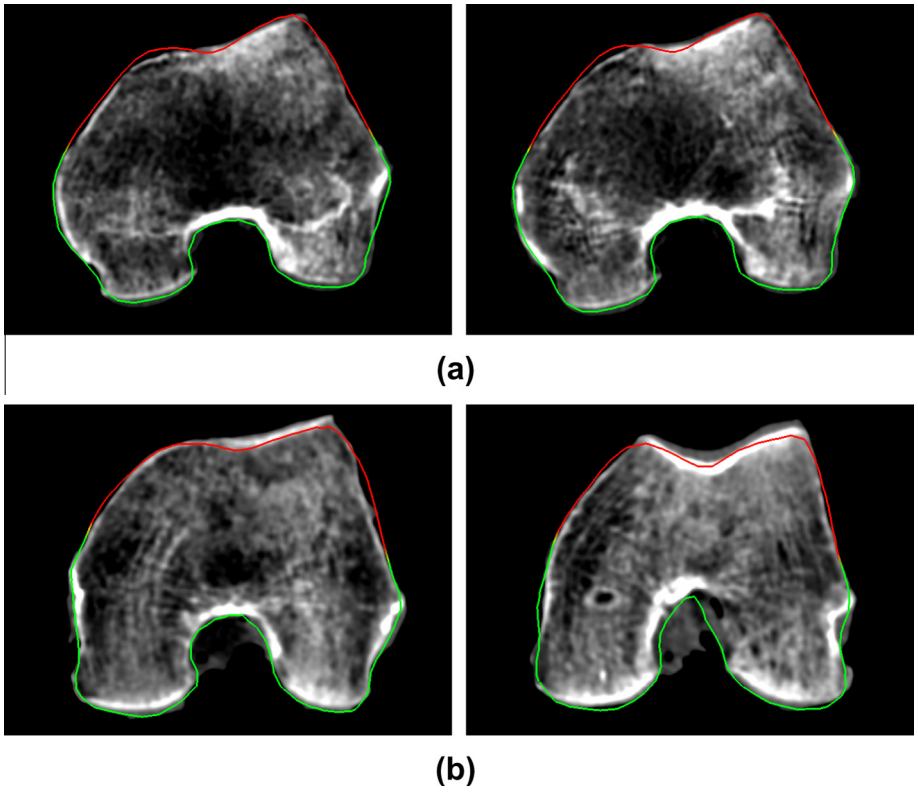
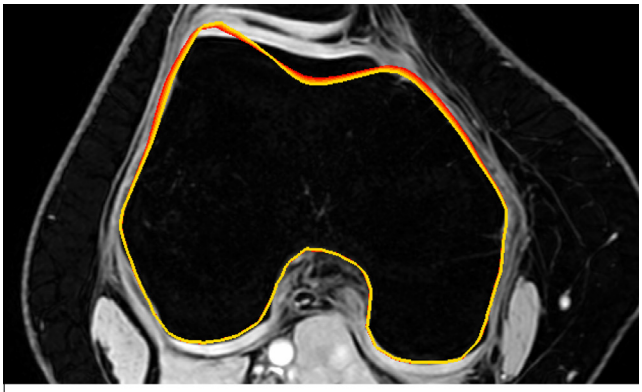
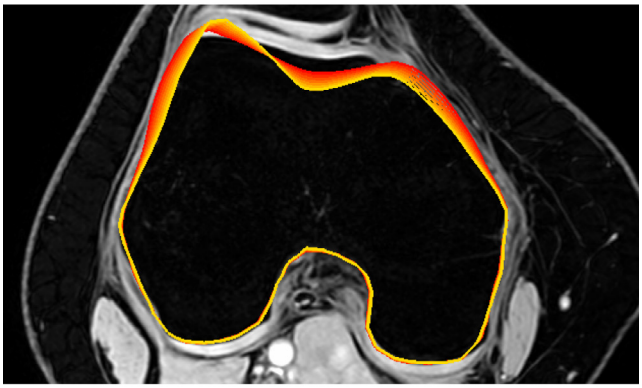


Fig. 11. When the method is applied to healthy bones, the proposed reconstructed shapes are very close to the patient's true trochlea, indicating that neither patient (a) nor (b) suffers from trochlea dysplasia.



(a) ±3 standard deviations



(b) ±9 standard deviations

Fig. 12. The remaining flexibility of the conditional shape model $p(\mathbf{s}|\mathbf{s}_g)$ displayed by adding the first principal component to the mean shape with coefficients between (a) ±9 and (b) ±3 standard deviations.

using the distance from the fitted models to the segmented surfaces, yielding a parameter of $\sigma = 1$. The result is a conditional shape model $p(\mathbf{s}|\mathbf{s}_g)$ as defined in Eq. (11). It models full anatomically correct femur shapes $\mathbf{s}$ that fit the given healthy part $\mathbf{s}_g$. The mean shape of this model represents the most probable shape of an anatomically correct trochlea for the patient given $\mathbf{s}_g$. Additionally, the conditional shape model models the space of all other shapes which fit the given healthy points. Although none of the shapes is more probable than the mean, they can deliver an insight into the remaining flexibility of the model and show the planning physician what other trochlea shapes could be possible.

6.2. Results

Here, we show visualizations of the proposed anatomically correct femur shapes for trochlea dysplasia patients. First, we show the most probable reconstruction. This is the mean $\boldsymbol{\mu}_c$ of the posterior shape model. In Fig. 9d the proposed shape is illustrated as a 3D shape placed into the patient's MRI scan. The remaining images in Fig. 9 show the intersection of the proposed shape with MRI slices. The green portion of these outlines represents the given healthy points $\mathbf{s}_g$. We observe that these fit the individual anatomy of the patient's femur well. The red part represents the trochlea region that is proposed by our posterior model. It exhibits a much deeper trochlear groove than that of the patient, confirming that this patient suffers from trochlea dysplasia. The proposed shape could be used as an operation plan for a surgeon. It is visible to what extent the trochlea groove needs to be deepened and that an augmentation of its edges should also be considered. Most likely, this remodeled shape would greatly improve the guiding of patella movement and prevent future dislocations. Fig. 10 shows the result of another case with similar results.

In both cases, the proposed anatomically correct trochlea differs from the patient's trochlea, confirming the already known diagnosis that he or she suffers from trochlea dysplasia. In contrast, we performed the same experiments on CT scans of two healthy bones in Fig. 11. Here, the proposed reconstruction agrees well with the patient's trochlea, confirming that he or she is not suffering from trochlea dysplasia.

In these images, we have only displayed the most probable reconstruction, the mean of the posterior shape model. However, in order to answer some of the questions posed earlier, we need to consider the full posterior shape model. First, it could be possible that the most probable reconstruction differs from the patient's trochlea, but that his trochlea shape is still in the range of normal shapes, modeled by the posterior model. In Fig. 12, we display the first principal component of the posterior shape model for the patient from Fig. 9. First, we visualize the anatomically normal range of ±3 standard deviations associated with this principal component in Fig. 12a. It shows that while there is still some flexibility regarding the depth of the trochlea groove. The patient's trochlea is far from the normal range. In order to give a better visual indication of the remaining flexibility, we exaggerated the same visualization to ±9 standard deviations in Fig. 12b. We see that the first principal component of the posterior model represents the depth of the trochlea groove. It should be noted however, that ±9 includes highly improbable and/or pathological shapes.

Obviously it is not enough to consider only the first principal component. In Fig. 13 we display the range of ±3 standard deviations for the first five principal components, for 5 patients. Patient 1 is the same patient as in Fig. 9, and Patient 2 the same as in Fig. 10. These examples lead us to the following conclusions:

1. All of these patients do suffer from trochlea dysplasia.
2. In all of these cases a deepening of the trochlea groove seems to be indicated. In most cases an small augmentation of the trochlea's edges should be considered.
3. The remaining flexibility is quite small, i.e. the range of possible reconstructions is most likely lower than the precision attainable in the operation. This means that the surgeon does not need to take any alternatives to the most probable reconstruction into account.

All the results we have presented so far provide a surgeon with a visual representation of the situation. Obviously it would be desirable in the future to transfer this directly into the operating room, for instance with an augmented reality system or even implementing the reconstruction automatically with a robot.

For now, to gain a more quantitative view of the reconstruction, it is possible to measure the exact distance between the patient's trochlea and the proposed reconstruction in each MRI slice. In Fig. 14a, we have plotted this distance on the proposed shape. The largest distance is seen in the trochlea groove and on its edges. Such a color code can guide the surgeon with respect to the depth and extent of the proposed trochlear groove. Fig. 14b shows a histogram of the distance values. The green bars represent the distance at the given points $\mathbf{s}_g$, showing that at these points the shapes essentially coincide. In red, we see the histogram of the trochlear region, which shows, again, that at many of these points, the proposed shape differs from the original one, proposing a deepening of the trochlear goove and a possible augmentation of its edges.

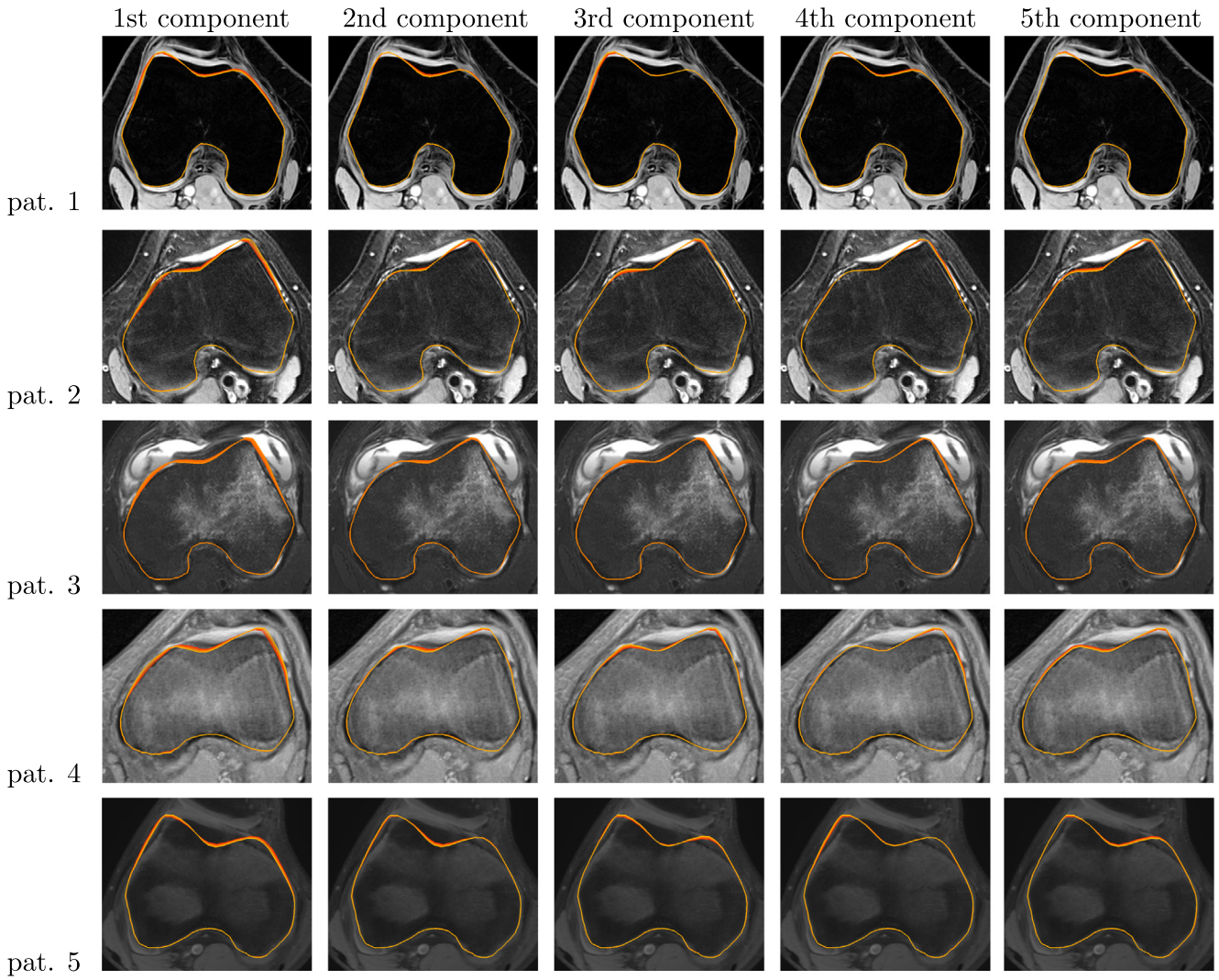


Fig. 13. The first five principal components of the posterior shape model for five different patients. While there is some remaining flexibility in the model, it is quite small, showing that no significantly different reconstructions other than the most probable reconstruction need to be considered.

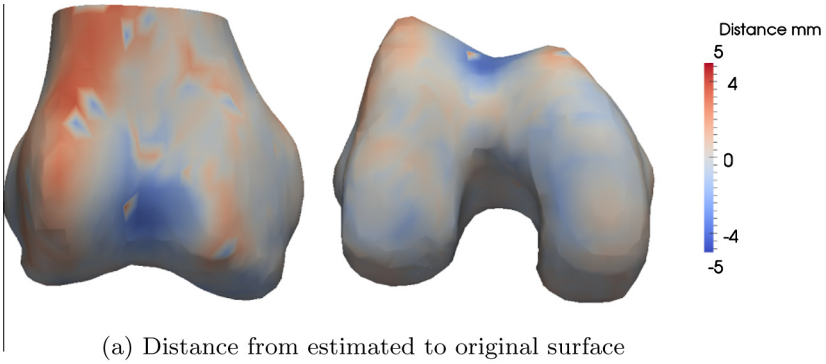
7. Conclusion

We have introduced a statistically sound method to include known partial data to methods involving statistical shape models by conditioning the statistical model on the given data. Because this posterior shape model is itself again a statistical shape model, it can be used without alterations in all algorithms relying on statistical shape models. As we have shown that the posterior model can be efficiently computed by a small $n \times n$ singular value decomposition, its use comes at almost no additional computational cost. We therefore believe that our method should replace existing ad hoc methods used to combine shape models and given data.

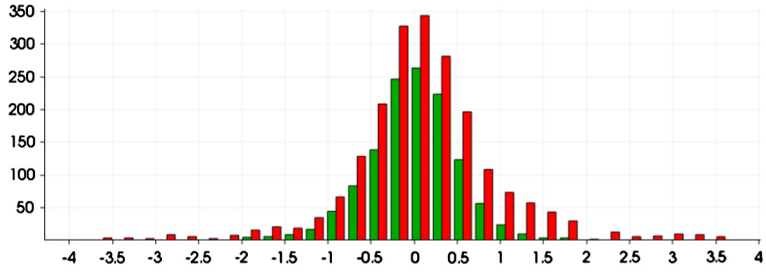
To show the usefulness of our approach, we have applied the posterior shape model to two problems from medical image analysis that naturally depend on the use of statistical shape models as prior knowledge. Since their invention, statistical shape models have been used for image segmentation. But the possibility to include user interaction to correct mistakes of fully automatic methods has so far received little attention. With our posterior models, user-supplied landmarks can be included into any algorithm that uses a statistical shape model. For our experiment, we have used a straight-forward strict shape constraint. While this does not warrant a perfect segmentation, only a good approximation within the model, we were able to show that the results benefit from the inclusion of landmarks. We see one of the great strengths of the posterior model in the fact that they are not limited to this specific segmentation model, but can be included like any other shape model in other, possibly more involved, algorithms. The second experiment we presented is also a standard application of statistical shape models: to predict one part of the shape from another. In our case we predict the malformed trochlea region of patients suffering from trochlear dysplasia from the remaining intact part of the distal femur. With the posterior shape model we can compute not only the most probable reconstruction, but a full shape model of possible reconstructions that can help a surgeon make an informed decision on how to perform a reconstructive surgery.

Outlook. An obvious way to extend our model is to adopt a more specific noise model, in case anything further is known about the accuracy or noise affliction of the given data. This could be a more complicated normal distribution or a different distribution altogether.

Like all shape models, our model depends heavily on the correct estimation of correspondence, both for the building of the original



(a) Distance from estimated to original surface



(b) Histogram of the distance values

Fig. 14. Quantitative evaluation of the distance between original and proposed femur shape. (a) shows the distance color-coded on the proposed shape. In (b) the green bars show the histogram of the distance values for the given points, the red bars the histogram of the estimated points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

statistical model and for the correspondence of the given data. But once the correspondence is correctly estimated, we have a closed-form expression for a posterior shape model that can be applied in any algorithm that uses a statistical shape model.

Acknowledgements

This work has been funded by the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) supported by the Swiss National Science Foundation as well as Synthes GmbH, Oberdorf, Switzerland.

Appendix A. Alignment, registration, and correspondence

In statistical shape models, shapes are assumed to be vectors, i.e. it is possible to form linear combinations of shapes. However, in practice, shapes are often given as triangulated surfaces, with a different number and ordering of triangles for each shape and it is not clear how linear combinations of these surfaces can be formed. The common approach for this problem is to *register* the shapes. In short, registration can be described as finding vector fields ϕ_i such that a set of shapes $\Gamma_1, \ldots, \Gamma_n$ can be represented as a deformation of a common reference Γ:

$$\Gamma_i = \{x + \phi_i(x) | x \in \Gamma\}, \quad i = 1, \ldots, n. \tag{A.1}$$

If Γ is a triangulated surface with m vertices, each shape can then be represented by the shape vector $\mathbf{s}_i \in \mathbb{R}^p$ with $p = 3m$ as:

$$\mathbf{s}_i = \left(v^x_{i,1}, v^y_{i,1}, v^z_{i,1}, \ldots, v^x_{i,m}, v^y_{i,m}, v^z_{i,m}\right)^T, \tag{A.2}$$

where the vector $v_{i,j} = \left(v^x_{i,j}, v^y_{i,j}, v^z_{i,j}\right)$ represents the x, y, z coordinates of the jth vertex of Γ_i. It is this representation that allows treating shapes like vectors and form linear combinations in the statistical shape model. The same principle can be applied to other discretization methods or representations of shapes. Finding the vector fields ϕ_i is a central problem in medical imaging and computer vision and is referred to as the *registration* or *correspondence* problem. The vertices $v_{i,j}$ should represent corresponding points in each surface Γ_i. Many algorithms for surface registration have been described in the literature (see e.g. Audette et al. (2000) for an overview).

Typically, statistical shape models model only variations in shape and not variations in position. The most straight-forward way to achieve this is to *align* the shapes before registration. Then, the vector fields ϕ_i reflect only the changes in shape, as the position and orientation of the surfaces is normalized. In practice, this means that for each shape Γ_i a rigid motion aligning it with the reference has to be computed. This is referred to as "rigid registration". Similarly, if an existing model is to be fitted to a novel data set, the two have to be aligned first, so that the shape model only needs to account for changes in shape.

A.1. Building the statistical deformation model

To build a deformation model, we use a database of manually segmented anatomically normal femur shapes. We choose an arbitrary shape as the reference, which we denote by $\Gamma_R \subset \mathbb{R}$. On this reference, we define a number of landmark points $L_R = \{p^R_1, \ldots, p^R_n\} \subset \Gamma_R$ (Fig. 2). The same landmarks are defined on all the other example surfaces from the database, and we perform a rigid alignment Horn (1987) to exclude rotational and translational components from our model. We denote the aligned surfaces by $\Gamma_1, \ldots, \Gamma_n$. As a next step we establish correspondence between the aligned surfaces by performing image registration on level set representation of the surfaces, as proposed by Paragios et al. (2002). More specifically, we register the signed distance images of the shape with the diffeomorphic Demons algorithm (Vercauteren et al., 2007). The result is a set of deformation fields $u_1, \ldots, u_n$, $u_i : \Omega \to \mathbb{R}^3$. We build the *statistical deformation model* by applying Principal Component Analysis on these deformation fields (Rueckert et al., 2003). Then, each instance of the statistical deformation model is a deformation field $u[\boldsymbol{\alpha}] : \Omega \to \mathbb{R}^d$ that can

be represented as a linear combination of principal components q_i with the coefficients given by the parameter vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$:

$$x \mapsto \mu(x) + \sum_{i=1}^{n} \alpha_i \, q_i(x), \tag{A.3}$$

where we used the function notation introduced in Section 4.2.1.

A.2. Image segmentation with a statistical deformation model

Let C_T be a new CT-Image of a patient, for which we want to perform the segmentation. By using a simple threshold segmentation, we can obtain a rough segmentation of the femur bone, which we denote as $\widetilde{B}_T$.

We annotate $\widetilde{B}_T$ with the same landmarks as were used in the registration and perform the rigid alignment to the reference image. We compute a distance image I_T: $\Omega \to R$ from the segmentation $\widetilde{B}_T$. For the reference, we have a high-quality manual segmentation B_R, from which we compute the distance image I_R. To find the deformation that best relates the reference with the target, we use again a registration approach of these distance images. We seek the deformation $u[\boldsymbol{\alpha}]$, which minimizes the error between the reference I_R and the target I_T and which is at the same time a likely instance of the model (i.e. $\|\boldsymbol{\alpha}\|^2$ is small):

$$\int_{\{x | I_R(x) < 0\}} (I_R(x) - I_T(x + u[\boldsymbol{\alpha}](x)))^2 dx + \lambda \|\boldsymbol{\alpha}\|^2 \to \min_{\alpha \in \mathbb{R}^n}. \tag{A.4}$$

The parameter $\lambda \in \mathbb{R}$ determines how strongly unlikely deformations are penalized.

The final segmentation B_T of the image C_T is obtained by transferring the high quality reference segmentation B_R onto the CT image. More precisely, we define the map $\phi(x) = x + u[\boldsymbol{\alpha}](x)$, $x \in \Omega$ and obtain the segmentation B_T as

$$B_T(x) = B_R(\phi^{-1}(x)).$$

While the inverse $\phi^{-1}(x)$ is typically not available explicitly, we can always compute an approximate inverse using Chen et al. (2008).

Appendix B. Equivalence between PCA and Gaussian process formulation

Proposition 1. *The Gaussian process* $\mathcal{GP}(\mu_p, \Sigma_p)$ *from* Eq. (28) *describes the same multivariate normal distribution as the conditional distribution* $p(\mathbf{s}|\mathbf{s}_g)$ *from* Eq. (11).

Proof. The Gaussian process interprets shape deformations as vector-valued functions defined on a finite domain $\Omega = (x_1, \ldots, x_N)$. But because these points are the points of our discrete surface model, each function $f : \Omega \to \mathbb{R}^d$ can be identified with a it p = Nd-dimensional vector of the values the vector components take at all the points. In this sense, the mean function $\mu(x_i)$ from Eq. (28) is the same as the mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$ from Eq. (11). Accordingly, the matrix Σ_{gg} corresponds to the matrix $\mathbf{Q}_g\mathbf{Q}_g^T$, and $\boldsymbol{\Sigma}_g(x_i)$ to $\mathbf{Q}_g\mathbf{Q}^T$. Thus, the mean μ_c of the Gaussian process $\mathcal{GP}(\mu_c, \Sigma_c)$ can be rewritten as:

$$\boldsymbol{\mu} + \mathbf{Q}\ \mathbf{Q}_g^T\left(\mathbf{Q}_g\mathbf{Q}_g^T + \sigma^2\mathbf{I}_q\right)^{-1}\ (\mathbf{s}_g - \boldsymbol{\mu}_g) \tag{B.1}$$

On the other hand, we can write out the mean $\boldsymbol{\mu}_c$ of the conditional distribution $p(\mathbf{s}|\mathbf{s}_g)$ as:

$$\boldsymbol{\mu} + \mathbf{Q}\left(\mathbf{Q}_g^T\mathbf{Q}_g + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{Q}_g^T(\mathbf{s}_g - \boldsymbol{\mu}_g) \tag{B.2}$$

While they are similar, these expressions are not the same, as the inner terms seem like rearranged versions of each other. However the following Lemma 1 proves that they are indeed equivalent.

Similarly, the covariance function Σ_c of the Gaussian process can be rewritten as a matrix:

$$\mathbf{Q}\mathbf{Q}^T - \mathbf{Q}\mathbf{Q}_g^T\left(\mathbf{Q}_g\mathbf{Q}_g^T + \sigma^2\mathbf{I}_q\right)^{-1}\mathbf{Q}_g\mathbf{Q}^T \tag{B.3}$$

while the covariance matrix $\boldsymbol{\Sigma}_c$ of the conditional distribution is written out as:

$$\sigma^2\mathbf{Q}\left(\mathbf{Q}_g^T\mathbf{Q}_g + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{Q}^T. \tag{B.4}$$

By factoring out the matrices $\mathbf{Q}$ and $\mathbf{Q}^T$ and rearranging the last expression we see that the two expressions are equal if and only if:

$$\left(\mathbf{I}_n + \sigma^{-2}\mathbf{Q}_g^T\mathbf{Q}_g\right)^{-1} = \mathbf{I}_q - \mathbf{Q}_g^T\left(\mathbf{Q}_g\mathbf{Q}_g^T + \sigma^2\mathbf{I}_q\right)^{-1}\mathbf{Q}_g. \tag{B.5}$$

This identity can be proved by setting $\mathbf{A} = \mathbf{I}_n$, $\mathbf{U} = \mathbf{Q}_g^T$, $\mathbf{B} = \sigma^{-2}\mathbf{I}_q$, $\mathbf{V} = \mathbf{Q}_g$ in the Binomial Inverse Theorem (Lemma 2). □

Lemma 1. *The following identity holds for any matrix* $\mathbf{A} \in \mathbb{R}^{q \times n}$*:*

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \sigma^2\mathbf{I}_q)^{-1} = (\mathbf{A}^T\mathbf{A} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{A}^T. \tag{B.6}$$

Proof. First, we consider the case of $q \geqslant n$. Let $\mathbf{A} = \mathbf{U}\,\mathbf{W}\mathbf{V}^T$ be a singular value decomposition of $\mathbf{A} \in \mathbb{R}^{q \times n}$. For this proof, we assume that this is a not an "economy SVD" but a full singular value decomposition with matrix sizes $\mathbf{U} \in \mathbb{R}^{q \times q}$, $\mathbf{W} \in \mathbb{R}^{q \times n}$, and $\mathbf{V} \in \mathbb{R}^{n \times n}$. Bearing in mind that the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal, we have:

$$\begin{aligned}
&\mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \sigma_q^2\mathbf{I})^{-1} && \text{(B.7)}\\
&= \mathbf{V}\mathbf{W}^T\mathbf{U}^T(\mathbf{U}\mathbf{W}\mathbf{W}^T\mathbf{U}^T + \mathbf{U}\ \sigma^2\mathbf{I}_q\ \mathbf{U}^T)^{-1} \\
&= \mathbf{V}\mathbf{W}^T\mathbf{U}^T\ \mathbf{U}(\mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}_q)^{-1}\mathbf{U}^T && \text{(B.8)}\\
&= \mathbf{V}\mathbf{W}^T(\mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}_q)^{-1}\mathbf{U}^T,
\end{aligned}$$

where in (B.8) we have used that $\mathbf{U}$ diagonalizes the inverse matrix in the bracket. The matrix $\mathbf{W}^T$ has a block structure $\mathbf{W}^T = \begin{bmatrix}\widetilde{\mathbf{W}} & \mathbf{0}\end{bmatrix}$, where $\widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ is the diagonal matrix of singular values of $\mathbf{A}$ and $\mathbf{0} \in \mathbb{R}^{n \times (q-n)}$ fills up $\mathbf{W}^T \in \mathbb{R}^{n \times q}$ with zeroes. We can therefore continue:

$$\begin{aligned}
&= \mathbf{V}\begin{bmatrix}\widetilde{\mathbf{W}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}(\widetilde{\mathbf{W}}^2 + \sigma^2\mathbf{I}_n)^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma^{-2}\mathbf{I}_{q-n}\end{bmatrix}\mathbf{U}^T \\
&= \mathbf{V}\begin{bmatrix}\widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}^2 + \sigma^2\mathbf{I}_n)^{-1} & \mathbf{0}\end{bmatrix}\mathbf{U}^T \\
&= \mathbf{V}\begin{bmatrix}(\widetilde{\mathbf{W}}^2 + \sigma^2\mathbf{I}_n)^{-1}\end{bmatrix}\begin{bmatrix}\widetilde{\mathbf{W}} & \mathbf{0}\end{bmatrix}\mathbf{U}^T \\
&= \mathbf{V}(\mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{W}^T\mathbf{U}^T \\
&= \mathbf{V}(\mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{V}^T\mathbf{V}\ \mathbf{W}^T\mathbf{U}^T \\
&= (\mathbf{V}\mathbf{W}^T\mathbf{W}\mathbf{V}^T + \mathbf{V}\sigma^2\mathbf{I}_n\mathbf{V}^T)^{-1}\mathbf{V}\ \mathbf{W}^T\mathbf{U}^T \\
&= \left(\mathbf{A}^T\mathbf{A} + \sigma_n^2\mathbf{I}\right)^{-1}\ \mathbf{A}^T.
\end{aligned}$$

This concludes the proof for $q \geqslant n$. The case $q < n$ can be proved by transposing both sides of (B.6). □

Lemma 2 (*Binomial Inverse Theorem*). *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{U} \in \mathbb{R}^{n\times q}$, $\mathbf{B} \in \mathbb{R}^{q\times q}$, $\mathbf{V} \in \mathbb{R}^{q\times n}$. *If* $\mathbf{B}$ *is invertible, we have:*

$$(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}. \quad \text{(B.9)}$$

Proof.

This is the Binomial Inverse Theorem for the case that $\mathbf{B}$ is invertible. For a proof, see e.g. Strang (2003). □

References

Albrecht, T., Knothe, R., Vetter, T., 2008. Modeling the remaining flexibility of partially fixed statistical shape models. In: Workshop on the Mathematical Foundations of Computational Anatomy, MFCA'08, New York, USA.

Audette, M.A., Ferrie, F.P., Peters, T.M., 2000. An algorithmic overview of surface registration techniques for medical imaging. Medical Image Analysis 4, 201–217.

Baka, N., de Bruijne, M., Reiber, J., Niessen, W., Lelieveldt, B., 2010. Confidence of model based shape reconstruction from sparse data. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010. IEEE, pp. 1077–1080.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.

Blanc, R., Reyes, M., Seiler, C., Szekely, G., 2009. Conditional variability of statistical shape models based on surrogate variables. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Blanc, R., Seiler, C., SzTkely, G., Nolte, L.P., Reyes, M., 2012. Statistical model based shape prediction from a combination of direct observations and various surrogates: application to orthopaedic research. Medical Image Analysis 16, 1156–1166.

Blanc, R., Szekely, G., 2012. Confidence regions for statistical model based shape prediction from sparse observations. IEEE Transactions on Medical Imaging 31, 1300–1310.

Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3d faces. In: SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. ACM Press, pp. 187–194.

Blanz, V., Vetter, T., 2002. Reconstructing the complete 3d shape of faces from partial information. Informationstechnik und Technische Informatik 44, 1–8.

Chen, M., Lu, W., Chen, Q., Ruchala, K., Olivera, G., 2008. A simple fixed-point approach to invert a deformation field. Medical Physics 35, 81.

Cootes, T., Taylor, C., 1992. Active shape models-'smart snakes'. In: Proc. British Machine Vision Conference 266275.

Cootes, T., Taylor, C., 2001. Constrained active appearance models. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, 2001 (ICCV 2001). IEEE, pp. 748–754.

Davatzikos, C., Liu, D., Shen, D., Herskovits, E., 2002. Spatial normalization of spine mr images for statistical correlation of lesions with clinical symptoms. Radiology 224, 919–926.

De Bruijne, M., Lund, M., Tank, L., Pettersen, P., Nielsen, M., 2007. Quantitative vertebral morphometry using neighbor-conditional shape models. Medical Image Analysis 11, 503–512.

Feldmar, J., Ayache, N., 1996. Rigid, affine and locally affine registration of free-form surfaces. International Journal of Computer Vision 18, 99–119.

Grenander, U., Miller, M., 1998. Computational anatomy: an emerging discipline. Quarterly of Applied Mathematics 56, 694.

Heimann, T., Meinzer, H., 2009. Statistical shape models for 3D medical image segmentation: a review. Medical Image Analysis.

Horn, B., 1987. Closed-form solution of absolute orientation using unit quaternions. JOSA A 4, 629–642.

Joshi, S., Miller, M., Grenander, U., 1997. On the geometry and shape of brain sub-manifolds. International Journal of Pattern Recognition and Artificial Intelligence 11, 1317–1343.

Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J., et al., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29, 196–205.

Liu, T., Shen, D., Davatzikos, C., 2004. Predictive modeling of anatomic structures using canonical correlation analysis. In: IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004. IEEE, pp. 1279–1282.

Lüthi, M., Albrecht, T., Vetter, T., 2009. Probabilistic modeling and visualization of the flexibility in morphable models. In: Proceedings of the 13th IMA International Conference on Mathematics of Surfaces XIII. Springer, p. 264.

Lüthi, M., Blanc, R., Albrecht, T., Gass, T., Goksel, O., Bnchler, P., Kistler, M., Bousleiman, H., Reyes, M., Cattin, P., Vetter, T., 2012. Statismo – a framework for PCA based statistical models.

Metz, C., Baka, N., Kirisli, H., Schaap, M., van Walsum, T., Klein, S., Neefjes, L., Mollet, N., Lelieveldt, B., de Bruijne, M., et al., 2010. Conditional shape models for cardiac motion estimation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, 452–459.

Paragios, N., Rousson, M., Ramesh, V., 2002. Matching distance functions: a shape-to-area variational approach for global-to-local registration. In: ECCV (2), pp. 775–789.

Petersen, K., Nielsen, M., Brandt, S., 2011. Conditional point distribution models. Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging, 1–10.

Pfirrmann, C., Zanetti, M., Romero, J., Hodler, J., 2000. Femoral trochlear dysplasia: Mr findings. Radiology 216, 858–864.

Pieper, S., Halle, M., Kikinis, R., 2004. 3D SLICER. In: IEEE International Symposium on Biomedical Imaging ISBI 2004.

Rasmussen, C., Williams, C., 2006. Gaussian Processes for Machine Learning. Springer.

Roweis, S., 1998. EM algorithms for PCA and SPCA. NIPS, 626–632.

Rueckert, D., Frangi, A., Schnabel, J., 2003. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. IEEE Transactions on Medical Imaging 22, 1014–1025.

Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical Applications in Genetics and Molecular Biology 4, 32.

Strang, G., 2003. Introduction to Linear Algebra. Wellesley Cambridge Pr.

Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society 61, 611–622.

Tomoshige, S., Oost, E., Shimizu, A., Watanabe, H., Kobatake, H., Nawano, S., 2012. Relaxed conditional statistical shape models and their application to non-contrast liver segmentation. Abdominal Imaging: Computational and Clinical Applications, 126–136.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2007. Non-parametric diffeomorphic image registration with the demons algorithm. Lecture Notes in Computer Science 4792, 319.

Verdonk, R., Jansegers, E., Stuyts, B., 2005. Trochleoplasty in dysplastic knee trochlea. Knee Surgery, Sports Traumatology, Arthroscopy 13, 529–533.