

## CHAPTER 1

# Shape modeling using Gaussian Process Morphable Models

Marcel Lüthi<sup>a,\*</sup>, Andreas Forster<sup>\*</sup>, Thomas Gerig<sup>\*</sup> and Thomas Vetter<sup>\*</sup>

<sup>\*</sup> Department of Mathematics and Computer Science, University of Basel, Switzerland

<sup>a</sup> Corresponding: marcel.luethi@unibas.ch

### Abstract

In this article we discuss Gaussian Process Morphable Models (GPMM), which are a generalization of classical statistical shape models. GPMMs extend statistical shape models in two ways: First they are formulated using Gaussian processes instead of the usual multivariate normal distribution, which makes it possible to obtain continuous models of the shape variability. Second, and more importantly, they allow us to model the shape variability using analytically defined covariance functions. This makes it possible to use GPMMs in cases where we do not have sufficiently many representative training examples. In particular, we will show in this article how to use GPMMs for 1) establishing correspondence between shapes where the GPMM is used as a shape-prior in the registration algorithm, and 2) fitting statistical shape models to data with pathologies.

This article provides an introduction to the basic theory and concepts behind GPMMs as well as a more tutorial style discussion on how to apply GPMMs in practical applications. As a main theoretical contribution we will introduce a novel kernel, which enforces axial symmetry of shape deformations. This is useful for modeling anatomical structures that are nearly symmetric, such as the skull or the pelvis. In the tutorial part, we will study the problem of building a skull model from noisy example data. This problem is challenging due to the limited availability of training data, the frequently occurring holes and artifacts in the data, as well as the complex topology of the skull itself. We will show how we can use GPMMs to build different prior models of varying degree of sophistication. Our results confirm that the better the model represents the characteristics of the skull, the better are the registration results we obtain. We will also show how we can use GPMMs to improve statistical shape models, such that they can better explain the full shape variability, or fit pathological shapes. In this context we discuss a fitting method, which does not only provide an accurate fit to the data, but allows us to split the result into the anatomically normal shape and the pathological deformations.

All the modelling and visualization methods that we discuss in this article are publicly available as part of the open source software scalismo [19].

## 1. Introduction

Statistical shape models (SSMs) are a well established tool in medical image analysis [9]. The core idea is that a statistical shape model represents the normal shape variation of a class of shapes, which is then used as prior knowledge in an algorithm. The most important examples of statistical shape models are the Active Shape Model [5] and the Morphable model [4], which learn the shape variation from given training examples, and represent the shape variation using the leading principal components. These models are linear, parametric models with mathematically convenient properties. As they can only represent shapes that are in the linear span of the given training examples, they lead to algorithms that are robust towards artifacts and noise. The downside of this specificity is that to learn an expressive model (i.e. one that can express all possible target shapes), a lot of training data is needed. Lüthi et al. have recently introduced a generalization of these classical statistical shape models, called Gaussian Process Morphable Models (GPMM), which use the mathematical framework of Gaussian processes to model shape variations [13]. GPMMs make it possible to construct expressive shape priors using analytically defined covariance functions, even when there are no or only few example shapes available to learn the shape variations from.

The main goal of this article is to give an introduction to GPMMs and to show that with this not too complicated concept, we can solve many shape modelling problems in a unified way, which previously required specialized algorithms. As a theoretical contribution we introduce a novel covariance function, which allows us to formulate GPMMs that enforce axial symmetric shape deformations. We will illustrate the use of GPMMs in an extended practical use case, which addresses the problem of building a skull model from a set of noisy skull surfaces. Our intention is that this use case does not only illuminate the concepts, but also serves as a practical guide on how to do shape modelling using GPMMs. In particular we want to illustrate the main considerations that need to be taken into account when we design prior models, how we can visualize and quantitatively assess the model quality and how we can choose the free parameters.

In the final section of this article, we will also introduce a novel application of GPMMs, which is the analysis of pathological shapes using a statistical shape model. The main idea is that we combine a classical statistical shape model, which is learned from example data, with an analytically defined model for the pathology. This combined model is fitted to the pathological shape and subsequently divided into the anatomically normal part, represented by the classical shape model, and the pathological deformation. This approach does not only make it possible to use a statistical shape model to fit pathological shapes, but also provide us with the most likely anatomical normal shape.

## 2. Shape modeling using Gaussian processes

Gaussian process morphable models are a generalization of classical (i.e. PCA-based) statistical shape models. We therefore start by revisiting the basic concepts of statistical shape models, before discussing how these models can be interpreted using the formalism of Gaussian processes.

### 2.1. Classical statistical shape models revisited

PCA-based statistical shape models assume that the space of all possible shape deformations can be learned from a set of typical example shapes  $\{\Gamma_1, \dots, \Gamma_n\}$ . Each shape  $\Gamma_i$  is represented as a discrete set of landmark points, i.e.

$$\Gamma_i = \{x_k^i \mid x_k \in \mathbb{R}^3, k = 1, \dots, N\},$$

where  $N$  denotes the number of landmark points. In early approaches, the points typically denoted anatomical landmarks, and  $N$  was consequently small (in the tens). Most modern approaches use a dense set of points to represent the shapes. In this case, the number of points is typically in the thousands. The crucial assumption is that the points are in correspondence among the examples. This means that the  $k$ -th landmark point  $x_k^i$  and  $x_k^j$  of two shapes  $\Gamma_i$  and  $\Gamma_j$  represents the same anatomical point of the shape. These corresponding points are either defined manually, or automatically determined using a registration algorithm. To build the model, a shape  $\Gamma_i$  is represented as a vector  $\vec{s}_i \in \mathbb{R}^{3N}$ , where the  $x, y, z$ - components of each point are stacked into a large vector:

$$\vec{s}_i = (x_{1x}^i, x_{1y}^i, x_{1z}^i, \dots, x_{Nx}^i, x_{Ny}^i, x_{Nz}^i).$$

This vectorial representation makes it possible to apply standard multivariate statistics to model a probability distribution over shapes. The usual assumption is that the shape variations can be modeled using a normal distribution

$$\vec{s} \sim \mathcal{N}(\vec{\mu}, \Sigma)$$

where the mean  $\vec{\mu}$  and covariance matrix  $\Sigma$  are estimated from the example data:

$$\vec{\mu} = \bar{s} := \frac{1}{n} \sum_{i=1}^n \vec{s}_i \tag{1.1}$$

$$\Sigma = S := \frac{1}{n-1} \sum_{i=1}^n (\vec{s}_i - \bar{s})(\vec{s}_i - \bar{s})^T. \tag{1.2}$$

As the number of points  $N$  is usually large, the covariance matrix  $\Sigma$  cannot be represented explicitly. Fortunately, as it is determined completely by the  $n$  example datasets, it has at most rank  $n$  and can therefore be represented using at most  $n$  basis vectors.

The basis vectors are found by performing a Principal Component Analysis (PCA) [11]. In its probabilistic interpretation, PCA leads to a model of the form

$$\vec{s} = \bar{s} + \sum_{i=1}^n \alpha_i \sqrt{d_i} \vec{v}_i \alpha_i \sim N(0, 1) \quad (1.3)$$

where  $(\vec{v}_i, d_i)$ ,  $i = 1, \dots, n$ , are the eigenvectors and eigenvalues of the covariance matrix  $\Sigma$  [23]. It is easy to check that under these assumptions,  $\vec{s} \sim \mathcal{N}(\bar{s}, S)$ . Thus, we have a efficient, parametric representation of the distribution.

## 2.2. Gaussian process morphable models

The literature of PCA based statistical shape models usually emphasizes the point that it is the shape that is modeled. Equation 1.3 however, gives rise to a different interpretation: A statistical shape model is a model of deformations  $\vec{\phi} = \sum_{i=1}^n \alpha_i \sqrt{d_i} \vec{v}_i \sim \mathcal{N}(0, S)$ , with which the mean shape  $\bar{s}$  is deformed. The probability distribution is on the deformations. This is the interpretation we use when we generalize these models to define Gaussian Process Morphable Models. We define a probabilistic model directly on the deformations. To stress that we are modeling deformations (i.e. vector fields defined on the reference domain  $\Gamma_R$ ), and to become independent of the discretization, we model the deformations as a Gaussian process.

Let  $\Gamma_R \subset \mathbb{R}^3$  be a reference shape and denote by  $\Omega \subset \mathbb{R}^3$  a domain, such that  $\Gamma_R \subseteq \Omega$ . We define a Gaussian process  $u \in \text{GP}(\mu, k)$  with mean function  $\mu : \Omega \rightarrow \mathbb{R}^3$  and covariance function  $k : \Omega \times \Omega \rightarrow \mathbb{R}^{3 \times 3}$ . Note that any deformation  $\hat{u}$  sampled from  $\text{GP}(\mu, k)$ , gives rise to a new shape  $\Gamma$  by warping the reference shape  $\Gamma_R$ :

$$\Gamma = \{x + \hat{u}(x) \mid x \in \Gamma_R\}.$$

Similar to the PCA representation of a statistical shape model used in (Equation 1.3), a Gaussian process  $\text{GP}(\mu, k)$  can be represented in terms of an orthogonal set of basis functions  $\{\phi_i\}_{i=1}^{\infty}$

$$u(x) \sim \mu(x) + \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} \phi_i(x), \alpha_i \in \mathcal{N}(0, 1), \quad (1.4)$$

where  $(\lambda_i, \phi_i)$  are the eigenvalue/eigenfunction pairs of the integral operator

$$\mathcal{T}_k f(\cdot) := \int_{\Omega} k(x, \cdot) f(x) d\rho(x),$$

where  $\rho(x)$  denotes a measure. The representation (1.4) is known as the Karhunen-Loève expansion of the Gaussian process [3]. Since the random coefficients  $\alpha_i$  are uncorrelated, the variance of  $u$  is given by the sum of the variances of the individual components. Consequently, the eigenvalue  $\lambda_i$  corresponds to the variance explained

by the  $i$ -th component. This suggests that if the  $\lambda_i$  decay sufficiently quickly, we can accurately approximate the process using the first  $r$  components only:

$$\tilde{u}(x) \sim \mu(x) + \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} \phi_i(x). \quad (1.5)$$

The expected error of this approximation is given by the tail sum

$$\sum_{i=r+1}^{\infty} \lambda_i.$$

The resulting model is a finite dimensional, parametric model, similar to a standard statistical shape model. Note, however, that there is no restriction that the covariance function  $k$  needs to be the sample covariance matrix. Any valid positive definite covariance function can be used.

The question remains how we can compute the eigenfunction/eigenvalue pairs in (1.5). Although for some kernel functions analytic solutions are available (see e.g. [2, 8]) for most interesting models we need to resort to numeric approximations. Fortunately, this problem has been widely studied in the literature (See e.g. Flannery et al. Chapter 18 [6]). A classical method, which we use in our implementation, is the Nyström method [17]. This method is discussed in detail in Appendix A.1.

### 2.3. Defining Gaussian processes

To define a Gaussian process  $\text{GP}(\mu, k)$  for modelling shape deformations, we need to define the mean function  $\mu : \Omega \rightarrow \mathbb{R}^3$  and a covariance function  $k : \Omega \times \Omega \rightarrow \mathbb{R}^{3 \times 3}$ . In the following, we will discuss two important strategies for choosing these parameters. As we are only interested in modelling the deformations on the reference surface  $\Gamma_R$ , we always let  $\Omega = \Gamma_R$ .

#### 2.3.1. Learning deformations from data

Similar to classical statistical shape models, we can learn the mean and covariance structure of the models from example data. Let  $\{\Gamma_1, \dots, \Gamma_n\}$  be the example surfaces and  $\{u_1, \dots, u_n\}$ ,  $u_i : \Gamma_R \rightarrow \mathbb{R}^d$  denote the corresponding deformation fields, which establish correspondence between the reference  $\Gamma_R$  and the respective surface, i.e.

$$\Gamma_i = \{x + u(x) | x \in \Gamma_R\}.$$

We can define the mean at every point  $x$  as the sample mean

$$\mu_{\text{SSM}}(x) = \frac{1}{n} \sum_{i=1}^n u_i(x) \quad (1.6)$$

and the covariance function at the points  $x$  and  $x'$  by the sample covariance

$$k_{\text{SSM}}(x, x') = \frac{1}{n-1} \sum_{i=1}^n (u_i(x) - \mu_{\text{SSM}}(x))(u_i(x') - \mu_{\text{SSM}}(x'))^T. \quad (1.7)$$

We refer to this kernel  $k_{\text{SSM}}$  as the *sample covariance kernel* or empirical kernel.

### 2.3.2. Modeling smooth deformations

We can also define a Gaussian process model when we have no example shapes. In this case, we usually assume that the chosen reference shape is close to the (hypothetical) average shape of the modeled class of shapes and let the mean deformation be zero

$$\mu(x) = (0, 0, 0)^T.$$

In absence of other prior knowledge, the most basic assumption when modeling shape deformations is that shape deformations vary smoothly. A well known covariance function that enforces smooth functions is the scalar-valued *Gaussian kernel*, defined by

$$g(x, x') = \exp(-\|x - x'\|^2 / \sigma^2),$$

where  $\sigma^2$  defines the range over which the function values at the points  $x$  and  $x'$  are correlated. The larger the values of  $\sigma$ , the more smoothly varying the resulting deformations fields will be. In order to use this scalar-valued kernel for modeling deformations, we can define a matrix valued Gaussian kernel as

$$k(x, x') = s \cdot \text{diag}(g) := s \begin{pmatrix} g(x, x') & 0 & 0 \\ 0 & g(x, x') & 0 \\ 0 & 0 & g(x, x') \end{pmatrix}.$$

The diagonal structure of the resulting  $3 \times 3$  matrix means that the  $x, y, z$  component of the modeled vector field are independent. The parameter  $s \in \mathbb{R}$  determines the variance (i.e. scale) of a deformation vector. This construction can be generalized by defining the matrix valued kernel  $k$  as

$$k(x, x') = A^T g(x, x') A, \quad A \in \mathbb{R}^{3 \times 3}, \quad (1.8)$$

which would allow us to introduce anisotropic scaling and correlations between the components [14].

### 2.3.3. Combining kernels

From a mathematical point of view, the only requirement that is needed to define a valid Gaussian process is that the covariance function is a symmetric and positive semi-definite kernel [1]. It is well known that the Gaussian kernel satisfies this property [17]. While it is in general rather difficult to prove that a symmetric function is

positive semi-definite, the following rules can be used to construct kernels that have this property [21, 14]:

**Theorem 1.** *Let  $g, h : \Omega \times \Omega \rightarrow \mathbb{R}$  be two symmetric positive semi-definite kernels and  $f : \Omega \rightarrow \mathbb{R}$  an arbitrary function. Then the following rules can be used to generate new positive semi-definite kernels:*

1.  $k(x, x) = g(x, x) + h(x, x)$
2.  $k(x, x) = \alpha g(x, x), \alpha \in \mathbb{R}_+$
3.  $k(x, x') = g(x, x')h(x, x')$
4.  $k(x, x') = f(x)f(x')$
5.  $B^T h(x, x')B, B \in \mathbb{R}^{r \times n}$
6.  $k(x, x') = k_3(\phi(x), \phi(x')) k_3 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \phi : \Omega \rightarrow \mathbb{R}^n$

Using these simple rules we can easily proof that the covariance functions that we defined above are positive semi-definite.<sup>1</sup> More importantly, however, they allow us to combine existing kernels to build more complicated models. We will make extensive use of these rules in Section 4 and 5, in order to define application specific shape priors.

### 2.3.4. Modeling symmetric variations

Many biological shapes are mirror symmetrical. It turns out that we can also encode this property directly into a covariance function, and thus obtain shape priors that enforces such symmetries. Without loss of generality, we assume that the mirror plane is orthogonal to the first dimension of the domain over which the Gaussian process is defined and that this plane is positioned at the origin. The mirrored position of a given point can then be expressed by flipping the sign of the first coordinate. We can define a symmetry-kernel over  $\mathbb{R}^3$  from a valid scalar valued kernel  $k(\cdot, \cdot)$  as

$$k_{sym}(x, x') = Ik(x, x') + \bar{I}k(x, \bar{x}'), \quad (1.9)$$

where  $I$  is the  $3 \times 3$  identity matrix and

$$\bar{I} = \begin{Bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix}, \quad \bar{x}' = \begin{Bmatrix} -x'_1 \\ x'_2 \\ x'_3 \end{Bmatrix}. \quad (1.10)$$

<sup>1</sup>Combine rule 4, 1 and 2 for deriving the empirical covariance kernel, and rule 5 for the matrix-valued Gaussian kernel.

The flipped sign of  $\bar{x}'$  reflects the point at the above mentioned mirror plane.<sup>2</sup> The negative sign for a specific diagonal entry in  $\bar{I}$  favors opposed oriented deformations for symmetric points along this dimension. A proof that this kernel is positive semi-definite is given in [15].

It turns out that defining a GPMM using the symmetric kernel  $k_{sym}$  leads to a model with increased variance compared to the original covariance function  $k$ . First the total variance changes for points close to the symmetry plane. Second the variance parallel to the symmetry plane increases while it is reduced in the perpendicular direction. Exactly on the symmetry plane the perpendicular variance is zero. While the change in perpendicular direction is expected, we correct the increased variance using a correction function  $S(x, x')$ , which reduces the variance depending on the distance to the symmetry plane:

$$S(x, x') = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & f(x)f(x') & 0 \\ 0 & 0 & f(x)f(x') \end{Bmatrix}, \quad f(x) = \sqrt{\frac{1.0}{2.0 - k(x, x) + k(x, \bar{x})}}. \quad (1.11)$$

The final symmetric kernel is given as:

$$k_{sym}(x, x') = k(x, x') + S(x, x')\bar{I}k(x, \bar{x}'). \quad (1.12)$$

### 3. Non-rigid registration using Gaussian process priors

In this section we show how we can use GPMMs as prior models for non-rigid surface registration. The idea is that we define a GPMM of possible shape deformations  $u \sim GP(\mu, k)$  to model the shape variability with respect to a reference shape  $\Gamma_R$ . Our main assumption is that we can identify for each point  $x_R \in \Gamma_R$ , the corresponding point  $x_T \in \Gamma_T$  of the target surface  $\Gamma_T$ . The goal of the registration problem is to recover the deformation field  $u$  that relates the two surfaces.

To this end, we formulate the problem as a MAP estimate:

$$\arg \max_u p(u|\Gamma_T, \Gamma_R) = \arg \max_u p(u)p(\Gamma_T|\Gamma_R, u), \quad (1.13)$$

where  $p(u) \sim GP(\mu, k)$  is a Gaussian process prior over deformation fields and the likelihood function is given by  $p(\Gamma_T|\Gamma_R, u) = \frac{1}{Z} \exp(-\eta^{-1} \mathcal{D}[\Gamma_T, \Gamma_R, u])$ , where  $\mathcal{D}$  is a metric that measures the similarity of two surfaces,  $\eta \in \mathbb{R}$  is a weighting parameter and  $Z$  a normalization constant.

In order to find the MAP solution, we reformulate the registration problem as an energy minimization problem. Taking logs in (1.13) we arrive at the equivalent mini-

<sup>2</sup>This corresponds to the sum over the orbit method for constructing a kernel following this mirror invariance (see [7]).



mization problem

$$\arg \min_u \mathcal{D}[\Gamma_R, \Gamma_T, u] - \eta \ln p(u) \quad (1.14)$$

Using the low-rank approximation (Equation (1.5)) we can restate the problem in the parametric form

$$\arg \min_{\alpha} \mathcal{D}[\Gamma_R, \Gamma_T, \mu + \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} \phi_i] + \eta' \|\alpha\|^2. \quad (1.15)$$

Here we used that the coefficients  $\alpha$  in (1.5) are independent and hence  $p(u) \propto \exp(-\|\alpha\|^2)$ . By denoting the model by

$$\mathcal{M}[\alpha](x) = \mu(x) + \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} \phi_i(x)$$

we can write this in the simpler form:

$$\arg \min_{\alpha} \mathcal{D}[\mathcal{M}[\alpha], \Gamma_T] + \eta' \|\alpha\|^2. \quad (1.16)$$

The final registration formulation (1.16) is a parametric optimization problem, which can be approached using standard optimization algorithms. This formulation also makes it clear that the registration problem is really a problem of model fitting, and that all the prior assumptions about possible deformations are represented in the GPMM  $\mathcal{M}[\alpha]$ .

### 3.1. Hybrid landmark and surface registration

Using Gaussian processes as a prior for surface registration also gives rise to a simple way to include landmark constraints. Let  $L_R = \{l_R^1, \dots, l_R^m\}$  and  $L_T = \{l_T^1, \dots, l_T^m\}$  be two sets of corresponding landmarks. Assume that a user has defined a number of landmark points  $L_R = \{l_R^1, \dots, l_R^m\}$  on a reference shape together with the matching points  $L_T = \{l_T^1, \dots, l_T^m\}$  on a target surface. These landmarks provide us with known deformations at the matching points, i.e.

$$\begin{aligned} L &= \{(l_R^1, l_T^1 - l_R^1), \dots, (l_R^m, l_T^m - l_R^m)\} \\ &=: \{(l_R^1, \hat{u}^1), \dots, (l_R^m, \hat{u}^m)\}. \end{aligned}$$

If we choose the likelihood function as

$$p(L|u) = \prod_{i=1}^m \mathcal{N}(u(l_R^i) | \hat{u}^i, \sigma^2 \mathcal{I}_d),$$

which asserts that the inaccuracies of the landmarks can be modeled as independent Gaussian noise, then the problem is an instance of Gaussian process regression and the

posterior distribution  $p(u|L)$  is known in closed form [17]. It turns out that the distribution  $p(u|L)$  is again a Gaussian process  $GP(\mu_p, k_p)$ , and its mean  $\mu_p$  and covariance  $k_p$  are given by

$$\mu_p(x) = \mu(x) + K_X(x)^T (K_{XX} + \sigma^2 I)^{-1} \hat{U} \quad (1.17)$$

$$k_p(x, x') = k(x, x') - K_X(x)^T (K_{XX} + \sigma^2 I)^{-1} K_X(x'). \quad (1.18)$$

Here, we defined

$$K_X(x) = (k(x, x_i))_{i=1}^n \in \mathbb{R}^{3n \times 3},$$

$$K_{XX} = (k(x_i, x_j))_{i,j=1}^n \in \mathbb{R}^{3n \times 3n}$$

and

$$\hat{U} = (\hat{u}^1 - \mu(x), \dots, \hat{u}^n - \mu(x))^T \in \mathbb{R}^{3n}.$$

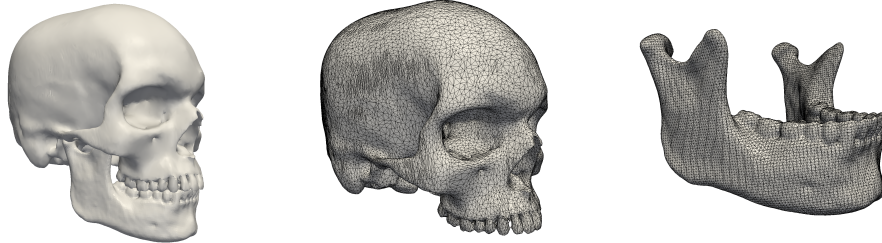
This implies that the posterior  $p(u|L)$  is again a valid GPMM (that is, a shape model), which already includes information about a given target surface. Thus we can use this posterior as the prior in the registration formulation (1.13). This results in a hybrid registration approach for surface registration, where the landmark information is implicitly enforced by the shape prior.

#### 4. Case study: Building a statistical shape model of the skull

In this section we show how Gaussian process morphable models can be used in practice. We consider the task of building a statistical shape model of the human skull. There are several aspects that make this task challenging: 1) The skull is a large and complex organ, with many structures that are difficult to model. 2) Due to the limited CT resolution, data segmented from CT usually contains holes and missing data. 3) Most data that we get in practice are images of elderly people, and hence teeth are often missing and dental fillings lead to large metal artifacts. 4) Finally, it is difficult to acquire sufficiently many example shapes to estimate the full shape variability of skulls.

The most important step in building a statistical model of the skull is to establish correspondence between the example shapes, which is performed using surface registration. In order to make this process robust to artifacts in the data, we already need a strong shape prior in this step. We will, of course, use a GPMM for this purpose.

In a first step we build a GPMM and discuss how we specifically tailored the prior to the problem of skull registration, using only analytically known covariance functions. In the second step we use this prior to establish correspondence between surfaces in a non-rigid registration task. In the final step, we learn a statistical shape model from the registered data, and demonstrated how to overcome the problem of not having a sufficiently large dataset to capture all shape variation.



**Figure 1.1** The reference skull used to build the skull models. Cranium and mandible are modeled separately. The cranium mesh has approximately 46’000 vertices and the mandible mesh approximately 26’000.

#### 4.1. Data and experimental setup

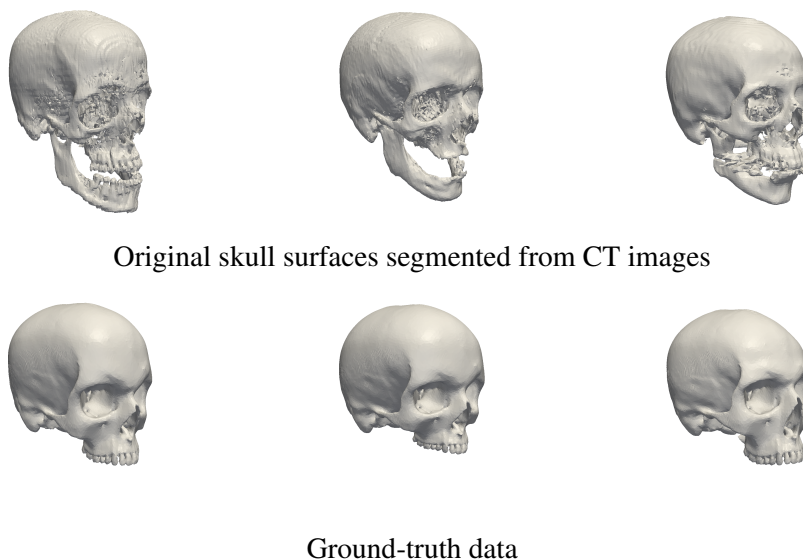
As a basis for all our modeling tasks we use a manually segmented, anatomically normal reference skull, which we denote by  $\Gamma_R$ . This reference skull is depicted in Figure 1.1. An issue that a good skull model should address is that the mouth can open and close. As this is not conveniently modeled using Gaussian processes, we have separated the cranium from the mandible and built separate models for each part. In this section we will only discuss the model built for the cranium, which is the more complicated of the two structures. The same considerations apply also when building a model of the mandible.

As example data, we use a set of automatically segmented CT images of the skull. Due to the poor quality of the example surfaces, we cannot use them directly to evaluate the quality of our models. In order to generate a ground-truth dataset, we have performed a surface-to-surface registration, which was guided by 16 manually defined landmark points. We then warped the reference skull with the resulting deformation field. A detailed description of how registration was done is given in Appendix A.2. Figure 1.2 shows typical example surfaces together with our generated ground truth.

For computing the low-rank approximation we choose 700 points on the surface of the reference skull  $\Gamma_R$ , which are used in the Nyström approximation (Cf. Appendix A.1). For all the experiments we approximate the original Gaussian process using the 200 leading eigenfunctions.

#### 4.2. Modeling shape priors for the skull

The first step is to build a shape prior that can be used in surface registration. In this section we will develop several different prior models, starting from simple smoothness priors to more sophisticated models that also enforce symmetry. For all the models we use a zero mean Gaussian process  $GP(\vec{0}, k)$ , where  $\vec{0} = (0, 0, 0)^T$  is the zero vector and  $k : \Gamma_R \times \Gamma_R \rightarrow \mathbb{R}^{3 \times 3}$  is a matrix-valued covariance function. As discussed previously, the assumption that the process is zero-mean, implies that we believe that



**Figure 1.2** Three examples of skull surfaces, which were segmented from CT data, and the corresponding ground-truth shapes, which we use for our evaluation.

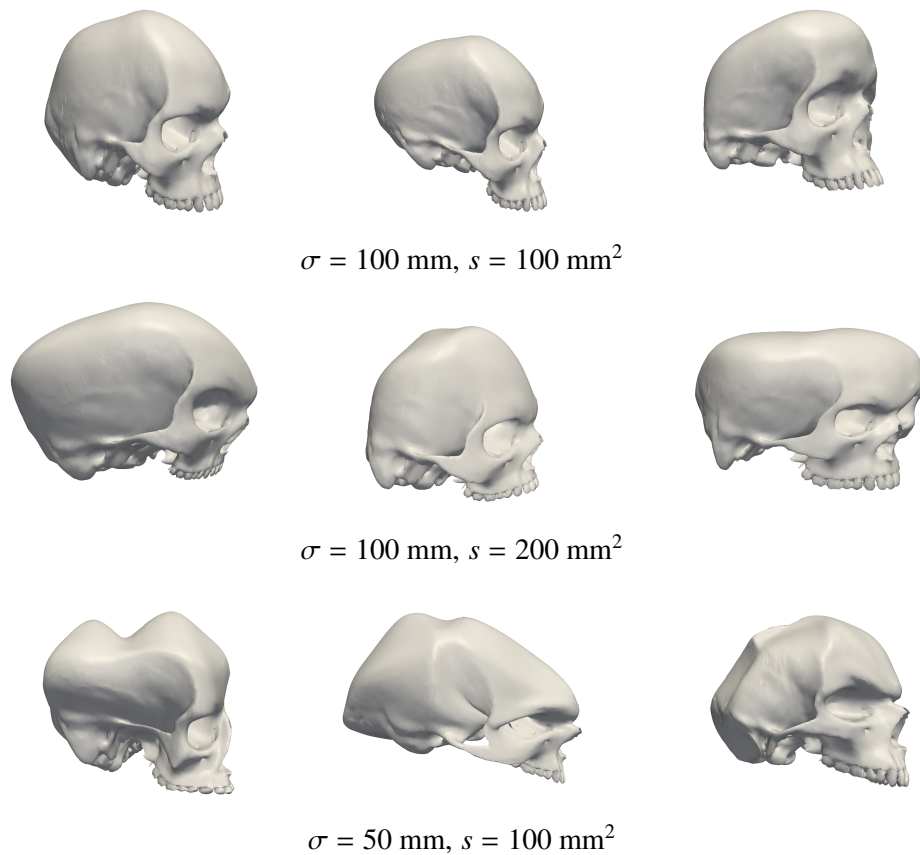
the shape of the reference skull is approximately average.

#### 4.2.1. Modeling smooth deformation using a single Gaussian kernel

The most basic modelling assumption we can make is that the deformations that relate two shapes of the same shape family are smooth. As discussed in Section 2.3.2 smooth deformations can, for example, be modeled using a Gaussian kernel for each component.

$$k_g^{(s,\sigma)}(x, x') = s \operatorname{diag}\left(\exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)\right).$$

There are two parameters that we have to define: the smoothness  $\sigma$  and the scale of the deformation  $s$ . Both parameters have a natural interpretation and a unit (usually  $mm$ ). The scale directly translates to the variance of the size of the deformation. Hence it is measured in  $mm^2$ . Taking into account that with high probability the deformation component are not larger than 3 standard deviations, we have an indication of how to choose the parameter. The parameter  $\sigma^2$  determines the strength of the correlation between two points  $x$  and  $x'$ . We know that for the Gaussian kernel, that values further away than  $2\sigma$  from the mean are already small, and hence almost uncorrelated. We should therefore choose the parameter such that  $2\sigma$  corresponds to approximately the range of correlations (in  $mm$ ) that we expect.



**Figure 1.3** Random samples from a GPMMs defined using a Gaussian kernel with different values for scale parameter  $s$  and smoothness parameter  $\sigma$ .

An important feature of Gaussian process models is that they are generative, and thus we can visualize samples from the prior to see if our assumptions are reasonable. To visualize the samples, we draw a random deformation field  $u_i$  from the Gaussian process model and warp the reference surface with this deformation. More precisely, the new surface is defined by warping all the points of the reference with the deformation field. Figure 1.3 shows random samples for different parameters of  $s$  and  $\sigma$ .

From visual inspection, we clearly see that choosing  $s = 200\text{mm}^2$  leads to unnaturally big deformations. We therefore prefer to use the value  $s = 100\text{mm}^2$  in our model, which seems approximately right. For the smoothness parameter  $\sigma^2$ , the samples with  $\sigma = 100\text{mm}$  look more natural than those where  $\sigma = 50\text{mm}$ . On the other hand, such a strong smoothness assumption also implies that the registration algorithm will not be able to explain more detailed shape variations.

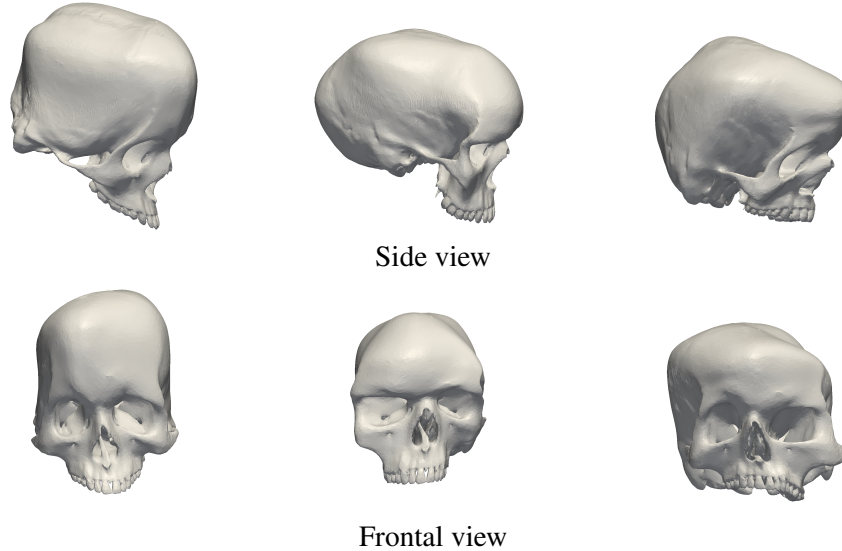


Figure 1.4 Three random samples from a GPMM defined using a multiscale kernel.

#### 4.2.2. Modeling deformations on multiple scale levels

The dilemma that small deformations are needed to explain detailed variations, but that these seem to produce unnaturally looking samples can be resolved by combining smaller and larger deformations in one model. The rules for combining kernels defined in Section 2.3 makes it possible to mix properties of different models. In particular, we can sum two covariance functions  $g, h : \Gamma \times \Gamma \rightarrow \mathbb{R}^{3,3}$  to obtain a new valid covariance function  $k(x, x') = g(x, x') + h(x, x')$ . Under this new covariance function  $k$ , two deformations  $u(x), u(x')$  are correlated, if they are either correlated by  $g(x, x')$  or  $h(x, x')$ . The variance of the deformation is the sum of the deformations defined by  $g$  and  $h$ . For modeling deformations for our skull prior, we define the kernel

$$k_{\text{ms}}(x, x') = k_g^{(70,100)}(x, x') + k_g^{(30,50)}(x, x').$$

This *multiscale kernel* assumes that large shape deformations are smoothly varying, but still allows for more local, but smaller changes.<sup>3</sup> Note that while the smoothness parameter is a combination of the above defined models, it has the same total variance as the kernels  $k_g^{(100,50)}$  and  $k_g^{(100,100)}$  defined above. From the random samples in Figure 1.4 we see that this model produces more natural shape variations, but still allows for flexible deformations.

<sup>3</sup>This idea of defining multiscale kernels is discussed in more detailed by Opfer et al. [16]

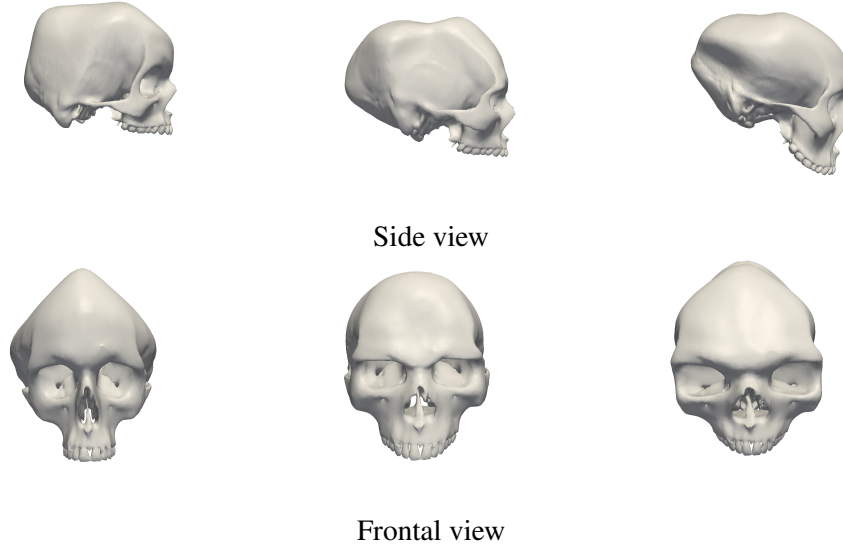


Figure 1.5 Three random samples from a GPMM defined a symmetric kernel.

#### 4.2.3. Incorporating symmetries into the prior

One important aspect of the skull is not yet represented in our model. Skulls are nearly symmetric about the sagittal plane. From the generated samples we see, however, that the models we built so far do not enforce such a symmetry. To incorporate symmetry into the model, we use the method discussed in 2.3.4 and define the symmetric kernel  $k_{\text{psym}}$  from the multiscale kernel  $k_{ms}$ :<sup>4</sup>

$$k_{\text{psym}}(x, x') = k_{ms}(x, x') + S(x, x')\bar{I}k_{ms}(x, \bar{x}'),$$

where  $S$  and  $\bar{x}$  and  $\bar{I}$  are defined as in Section 2.3.4. The kernel  $k_{\text{psym}}$  models perfect symmetry, which is usually not given in natural shapes. We therefore also allow for small asymmetric deformations, which we again model using a Gaussian kernel. The final kernel has the form

$$k_{\text{sym}}(x, x') = k_{\text{psym}}(x, x') + \text{diag}(k_g^{(2,100)}(x, x'))$$

Sampling from this model already leads to much more realistically looking skull shapes, as shown in Figure 1.5.

<sup>4</sup>To define the symmetry plane we used 3 manually clicked points on the sagittal plane. We then use this information to rigidly transform the skull so that the symmetry plane aligns with the y-z plane at the coordinate axis.

#### 4.2.4. Quantitative comparison of different prior models

So far we have selected the model using purely qualitative arguments and with the help of the visualized samples. In this section we support our choices by providing a quantitative evaluation of our models. For this purpose we use the standard metrics of generalization, specificity as well as a compactness introduced by Styner et al. [22].

The *generalization* ability measures how accurately the model, denoted by  $\mathcal{M}[\alpha]$ , can fit the datasets from the modeled class of shapes, which is here represented by the set of ground truth datasets  $\{\Gamma_1, \dots, \Gamma_m\}$ . More precisely, we define generalization for a model  $\mathcal{M}$  as:

$$\text{gen}(\mathcal{M}) = \frac{1}{m} \sum_{i=1}^m \min_{\alpha \in \mathbb{R}^n} D[\mathcal{M}[\alpha], \Gamma_i].$$

As a distance measure we use the averaged squared Euclidean distance of the corresponding points. As the model is in correspondence with the ground-truth dataset, we can compute the minimization in closed form.<sup>5</sup>

*Specificity* is a measure of how well randomly generated samples from the model resemble valid instances from this class of shapes. It is evaluated by comparing the distance of  $k$  randomly generated sample shapes (determined by the set of model parameters  $\{\tilde{\alpha}_1, \dots, \tilde{\alpha}_k\}$ ) to the closest shape of a set of valid example shapes

$$\text{spec}(\mathcal{M}) = \frac{1}{k} \sum_{j=1}^k \min_i D[\mathcal{M}[\tilde{\alpha}_j], \Gamma_i].$$

Finally, *compactness* indicates how much variance the model represents when it is represented using only the first  $r$  basis functions. This variance is given by the sum of the leading  $r$  eigenvalues:

$$\text{var}(\mathcal{M}) = \sum_{i=1}^r \lambda_i.$$

We report here two measures of compactness: 1) how many basis functions are needed to represent 99% of the total variance of the model and 2) how much of the total variance of the process (given by  $\int_{\Gamma_R} k(x, x) dx$ ) is approximated using the first 200 eigenfunctions). Compactness can be seen as a measure of complexity. The more complex the model (i.e. the more flexible it is to represent many different shapes accurately) the more eigenfunctions are needed to accurately represent the model.

Table 1.1 shows the 3 measures applied to the different prior models. We see that the numbers confirm our visual impression that the symmetric model is clearly

<sup>5</sup>To compute the minimum in practice, we use Gaussian process regression with an anisotropic noise of 0.1 mm, on a 1000 uniformly sampled points.



Model	Generalization	Specificity	99 % Var.	Approx. Var
$k_g^{(100,100)}$	0.30 mm	14.6 mm	48	0.999
$k_g^{(200,100)}$	0.30 mm	20.6 mm	48	0.999
$k_g^{(100,50)}$	0.28 mm	15.5 mm	183	0.978
Multiscale	0.24 mm	14.7 mm	126	0.994
Symmetric	0.21 mm	12.5 mm	73	0.999

**Table 1.1** Quantitative evaluation of the model quality for different models. The last two columns show the compactness of the model. The second last column shows how many basis functions were required to approximate 99 % of the variance, while the last column shows how much of the models total variance was approximated using 200 basis functions.

the best, followed by the multi-scale model. For both models we managed to keep the compactness and specificity low by incorporating more knowledge about the true structure of the shape variability. That the symmetric model can at the same time be the most specific and the most general is a consequence of the strong prior assumptions. The leading 200 eigenfunctions, which are used to represent the model, capture most of the characteristic variations in the class of skull shapes. In contrast, the leading 200 eigenfunctions of the more generic models include variations that are not occurring in this class.

### 4.3. Skull-registration using model fitting

Once we have the prior model defined, we can perform the actual surface registration. Recall from Section 3 that to perform the registration we minimize the functional

$$\arg \min_{\alpha} \mathcal{D}[\Gamma_T, \mathcal{M}[\alpha]] + \eta' \|\alpha\|^2. \quad (1.19)$$

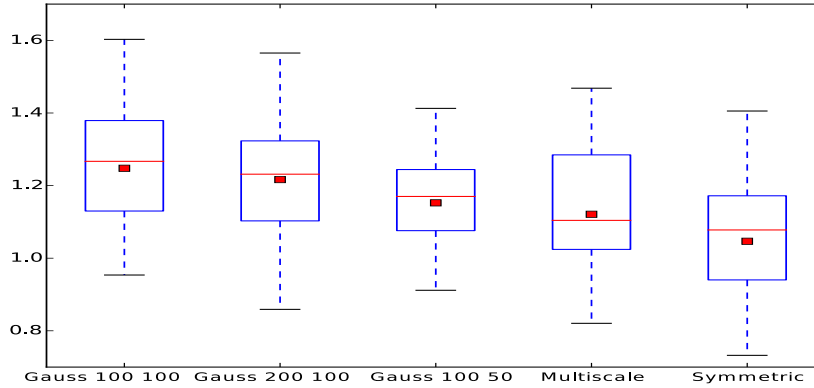
For the distance measure  $\mathcal{D}$ , we choose a robust distance measure

$$\mathcal{D}[\Gamma_1, \Gamma_2] = \int_{\Gamma_1} \rho(x - CP_{\Gamma_2}(x)) \quad (1.20)$$

where  $CP_{\Gamma}(x) = \arg \min_{x' \in \Gamma} (\|x - x'\|)$  finds the closest points to  $x$  on the surface  $\Gamma$  and  $\rho$  is the Huber loss function [10] defined by

$$\phi(x) = \begin{cases} x^2/2 \\ (|x| - k/2) \end{cases} \quad k = 1.345.$$

Figure 1.6 show the accuracy of registration results evaluated using the ground-truth data. We see that the results correspond to what we expected from the model metrics. The symmetric kernel clearly yields the best results, followed by the multiscale kernel. Figure 1.7 shows a typical registration result, where the red surface shows the registration solution and the white (transparent) surface the ground truth.



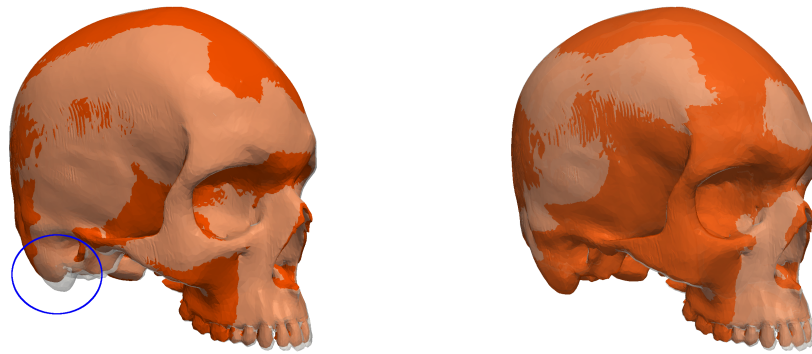
**Figure 1.6** Registration results for the different models. The numbers represent the Procrustes distance (in mm) between the registered surface and the ground truth.

We see that the shape is well matched in general. Upon close inspection, however, we can spot places where there are still larger errors (indicated by the arrow). We could now start to tune the parameters of our model in the hope that we can find values which reduce this error. A more direct and efficient solution is to define landmark points to enforce the correct deformations manually and to include these as known observations into the model. As discussed in Section 3.1, the resulting posterior model is again a Gaussian process, and thus can itself be used as the prior model in the registration functional (1.19). Figure 1.7 (right) shows the same registration, but this time with landmarks. We see that the error is greatly reduced when we enforce the deformation using a landmark. That adding landmark points will reduce the error for all models is shown in Figure 1.8, where we used 6 manually defined landmarks on all the skulls.

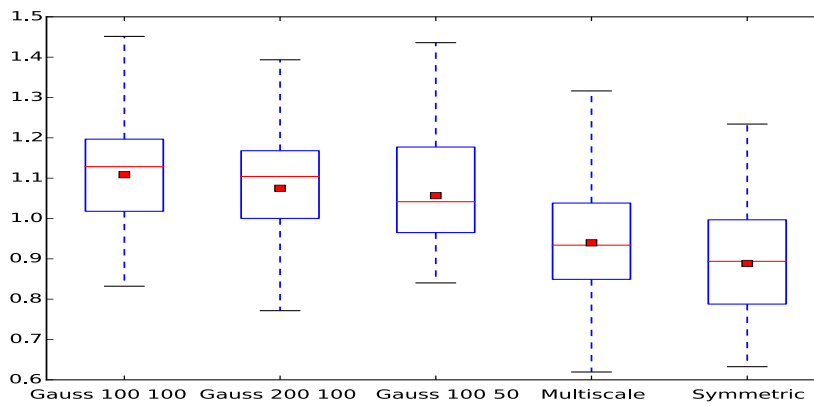
#### 4.4. Computing the statistical shape model

In the final step we can use the registered data to compute a statistical shape model [5, 4]. Let  $\{u_1, \dots, u_n\}, u_i : \Gamma_R \rightarrow \mathbb{R}^d$  denote the deformation field resulting from the registration, which establishes correspondence between the surfaces  $\Gamma_R$  and  $\Gamma_i$ . As discussed above, in the Gaussian process framework, a statistical shape model is simply a Gaussian process model  $\text{GP}(\mu_{\text{SSM}}, k_{\text{SSM}})$ , where the mean and covariance functions are estimated from the data using the formulas for the sample mean and sample covariance (Equation (1.6) and (1.7)). It is interesting to compute the quantitative measures from the previous section (Cf. Table 1.1) also for this model.<sup>6</sup> In Table 1.2 we see

<sup>6</sup>In this case we perform a leave-one-out procedure to estimate the generalization error.



**Figure 1.7** Left: A typical registration result. The shape is in generally well matched, but some local features still show some registration error. Right: The same registration result, but when the correct matching is enforced using a landmark that is placed in the erroneous region.



**Figure 1.8** Registration results for the different models when 5 landmarks were used to guide the registration. The numbers represent the Procrustes distance (in mm) between the registered surface and the ground truth.

that the generalization error is comparably large, but the model is much more specific and also more compact than the previously defined models. The good specificity value constitute the big advantage of statistical shape models.

The relatively large generalization error of 0.89mm is due to the small number of example shapes. It shows that 46 examples are not sufficient to cover the full shape variation of the skull. Fortunately, combining different kernels also yields a simple solution to this problem. The idea is to model the missing variability as a small, but smoothly varying deformation using e.g. a Gaussian kernel  $k_g(100, 1)$ . Combining this

Model	Generalization	Specificity	99 % Var.	Approx. Var
SSM	0.89 mm	3.6 mm	24	1.000
SSM (augmented)	0.23 mm	4.6 mm	23	0.999

**Table 1.2** Quantitative evaluation of the model quality for the classical statistical shape models and the augmented model.

with the statistical shape model leads to the *augmented* model

$$k_{\text{aug}}(x, x') = k_{\text{SSM}}(x, x') + k_g(100, 1),$$

where we assumed that the scale of the error is around 1 mm and is smoothly varying. Computing the usual metrics reveals that this final model generalizes much better, while it still provides excellent specificity and compactness (see Table 1.2). Hence, in practical applications this model is likely to be superior to a classical statistical shape model.

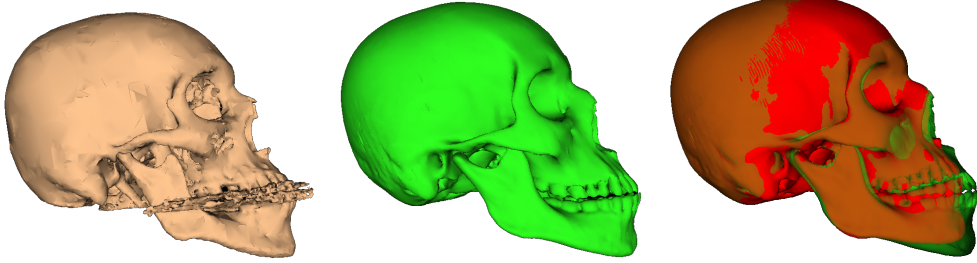
## 5. Modeling and analyzing pathologies

That GPMMs allow us to increase the variability of a statistical shape model makes it possible to use statistical shape models for fitting pathological shapes. In this section we discuss this important topic on the exemplar case study of fitting a skull model to a CT scan of a patient with an overbite. The idea is to use the statistical skull model in order to devise a surgical plan to attain a desired aesthetic facial profile [20].

Fitting a statistical shape model built from healthy examples only does not work in this case. The model would trade-off the fitting accuracy in the healthy regions against a better explanation of the pathological deformations. To overcome this difficulty, we explicitly model the pathological part using an analytically defined Gaussian process. While we could use exactly the same approach, as for extending the model variation in a statistical skull model (Cf. Section 4.4), we use here a variant of this approach, which does not compute a low-rank approximation of the full model, but approximates the statistical shape model and the model for the pathological part separately. This allow us to separate the healthy part of the deformation from the combined deformation including also the pathological part after the fitting, as illustrated in Figure 1.9 (right).

### 5.1. Simultaneous fitting of healthy and pathological deformations

Let  $u_{\text{SSM}} \sim GP(\mu_{\text{SSM}}, k_{\text{SSM}})$  and  $u_p \sim GP(0, k_g(x, x'))$  be two GPMMS which explain the healthy and the pathological part of a shape deformation. Assuming that we have



**Figure 1.9 Registration of a pathological skull with overbite** On the *left* the pathological target skull is illustrated. The *middle* shows the fitting result using the statistical model combined with the pathological model. On the *right* the fitting result in green as overlay on top of the SSM part only in red, predicting the healthy shape.

for both models computed the low-rank models (Cf. Section 2.2),

$$\mathcal{M}_{\text{SSM}}[\alpha^{\text{SSM}}] = \mu_{\text{SSM}} + \sum_{i=1}^r \alpha_i^{\text{SSM}} \sqrt{\lambda_i^{\text{SSM}}} \phi_i^{\text{SSM}} \quad (1.21)$$

$$\mathcal{M}_{\text{p}}[\alpha^{\text{p}}] = \sum_{j=1}^{r'} \alpha_j^{\text{p}} \sqrt{\lambda_j^{\text{p}}} \phi_j^{\text{p}}, \quad (1.22)$$

we can write any shape  $\Gamma$  as

$$\Gamma = \{x + \mathcal{M}_{\text{SSM}}[\alpha^{\text{SSM}}](x) + \mathcal{M}_{\text{p}}[\alpha^{\text{p}}](x) | x \in \Gamma_R\}$$

for some sets of parameters  $\alpha^{\text{SSM}}$  and  $\alpha^{\text{p}}$ .

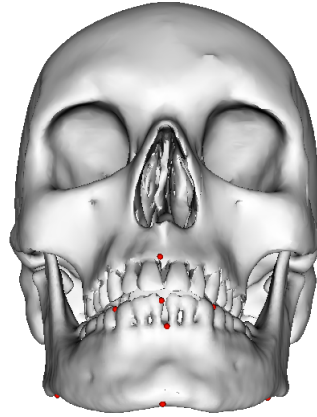
Thus, we can write the combined registration problem as

$$\arg \min_{\alpha^{\text{SSM}}, \alpha^{\text{p}}} \mathcal{D}[\mathcal{M}_{\text{SSM}}[\alpha^{\text{SSM}}] + \mathcal{M}_{\text{p}}[\alpha^{\text{p}}], \Gamma_T] + \eta \|\alpha^{\text{SSM}}\|^2 + \eta \|\alpha^{\text{p}}\|^2,$$

where  $\mathcal{D}$  is the distance measure defined in (1.20). The optimal parameter-sets  $\alpha^{\text{SSM}}$  and  $\alpha^{\text{p}}$ , represent the explanations of the different deformation models. The anatomically normal skull shape is thus given by the model  $\mathcal{M}_{\text{SSM}}$  and the associated model parameters  $\alpha^{\text{SSM}}$ .

## 5.2. Building a model for pathologies

The main modes of deformation of the combined model are not necessarily orthogonal. This means that it might be possible to explain certain shape deformation with either of the models. To reduce this effect, the model explaining the pathological deformations should not account for deformations outside the region where we expect the shape to depict pathological variability. A model can be restricted to a predefined region of the full domain by using a spatially varying kernel  $k_{\text{sv}}(x, x')$ . The spatially



**Figure 1.10 Landmark placement on the reference shape** For this example we have placed 8 landmarks in the teeth and mandible region of the skull. The landmarks in the teeth region prevent the model from adapting to the metal artifacts in the teeth region.

varying property of the kernel is defined through a function  $a(x)$  that damps the size of the deformations a kernel allows based on the location of the points. We define the spatially-varying kernel as

$$k_{sv}(x, x', k) = a(x)k(x, x')a(x')$$

where  $k$  itself is an arbitrarily chosen kernel and  $a(x)$  is the function that damps the kernel functions magnitude except in a defined region.

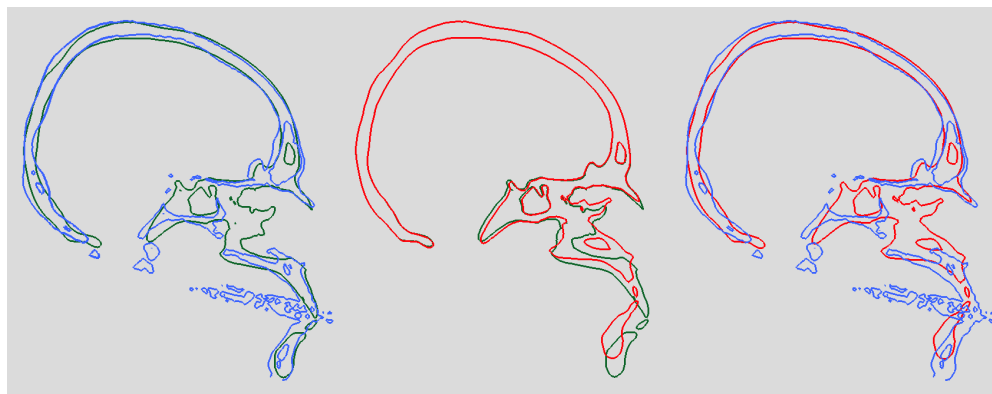
In our example we expect the pathological deformation to appear in the region of the mandible. Therefore we defined the bias model only in the region of the mandible, where we expect the pathology and hence need additional flexibility. We use a function

$$a(x) = \exp(-\|x - x_c\|^2/\sigma^2)$$

defining a region around the mandible with  $x_c$  as the closest point to  $x$  on the mandible. The parameter  $\sigma$  controls the size of the region around the mandible with function values of 1.0 for points that are close to the mandible and smoothly go to 0.0 for points that are further away.

### 5.3. Experiment

We applied this procedure to fit the skull shown in Figure 1.9. As there are large metal artifacts in the teeth region we decided to place 8 landmarks to increase the fitting accuracy (see Figure 1.10). The fitting results are depicted in Figure 1.9. The green surface (*middle/right*) is our final fitting result while on the *left* the original target skull



**Figure 1.11 Sliced mesh representation of the fitting result** In all the illustrations we can see the registration target in blue, the fitting result with bias model in green and the part explained by the statistical shape model in red. On the *left* we see the fitting result to the blue target. The region of the mandible the shape is well aligned. The *middle* shows the parts explained by the statistical shape model in red and the additional part from the bias model in green. It can be seen that the bias model allows for additional flexibility in the mandible region, which originates from the spatially varying kernel. On the *right* we see that most of the target skull is explained well by the statistical shape model. However, in the mandible region it is helpful to add more flexibility with the bias model.

is depicted. On the *right* the part explained by the statistical shape model only is depicted in red. The visible difference, of the fitting result of the combined model and the statistical shape model only reconstruction, demonstrates the ability to separate the pathological deformation using the combined model approach. The results are also shown in the profile view for 2D slices through the skull in Figure 1.11. The target is shown in blue, the fitting result in green and the part explained exclusively by the statistical shape model in red. Again, it can be seen that the statistical shape model does not explain the overbite, but the pathological part of the deformation is explained by the added spatially varying model.

## 6. Conclusion

In this article we have discussed GPMMs, a generalization of PCA-based statistical models. With GPMMs we are not restricted to estimate all the covariance structure from training examples, but can additionally use analytically defined kernels to model our prior assumptions. Typically, these analytically defined kernels encode smoothness assumptions of the deformation. However, also more complex constraints, such as that the model should match observed deformations can be formulated, or that the model shape variation is symmetric around a given axis. The rules for constructing kernels, allow us to combine learned and analytically defined kernels, and thus to build

expressive shape priors even when there are not sufficiently many example surfaces available to learn the full shape variability.

We have presented an extended use case, where we discussed how GPMMs can be used for the problem of building a model of the skull. We have shown how we can start with simple prior assumptions, and gradually incorporate more knowledge into the model. We have seen that a practical way to assess if our prior is reasonable is by visualizing random samples from the model. Our quantitative evaluation of the different models confirmed that incorporating more prior knowledge into the model does not only lead to a better model as measured by the model metrics, but that this translates directly to better registration results. Our experiments also confirmed the usefulness of combining a learned statistical shape model with an analytically defined prior. In cases when we do not have sufficiently many example surfaces to learn the shape variation, we showed that augmenting it with an analytically defined model greatly enhances the generalization ability, without sacrificing much of the specificity of the model. Indeed, the result suggest that whenever we have training data available, it is always beneficial to include these into the model. We have also shown how the combination of a statistical shape model and analytically defined model make it possible to fit shapes with deformities or pathologies. The result can even be divided up again, to obtain the most likely anatomical normal explanation of the data and an explanation of the pathological part.

All the different applications of GPMMs that we showed in this article are based on the same mathematical principle and are made possible thanks to the great flexibility of Gaussian processes. We believe that being able to use a single, mathematically well established concept for all these different task, is the biggest advantage of GPMMs for shape modelling. It greatly reduces the algorithmic complexity and lets us focus on modelling the problem at hand. Moreover, all the methods that are needed for using GPMMs are available as open source, as part of the modelling software Scalismo [19] and Statismo [12].

## Appendix

### A.1. Approximating eigenfunctions using the Nyström method

The goal of the Nyström method is to obtain a numerical estimate for the eigenfunctions/eigenvalues of the integral operator

$$\mathcal{T}_k f(\cdot) := \int_{\Omega} k(x, \cdot) f(x) d\rho(x). \quad (.23)$$

The pairs  $(\phi_i, \lambda_i)$ , satisfying the equation

$$\lambda_i \phi_i(x') = \int_{\Omega} k(x, x') \phi_i(x) d\rho(x), \quad \forall x' \in \Omega \quad (.24)$$



are sought. The Nyström method is intended to approximate the integral in (.24). This can, for example, be achieved by letting  $d\rho(x) = p(x) dx$  where  $p(x)$  is a density function defined on the domain  $\Omega$ , and to randomly sample points  $X = \{x_1, \dots, x_n\}$  according to  $p$ . The samples  $(x_l)_{l=1, \dots, N}$  for  $x'$  in (.24) lead to the matrix eigenvalue problem

$$Ku_i = \lambda_i^{mat} u_i, \quad (.25)$$

where  $K_{il} = k(x_i, x_l)$  is the kernel matrix,  $u_i$  denotes the  $i$ -th eigenvector and  $\lambda_i^{mat}$  the corresponding eigenvalue. Note, that since the kernel is matrix valued ( $k : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$ ), the matrices  $K$  and  $k_X$  are block matrices:  $K \in \mathbb{R}^{nd \times nd}$  and  $k_X \in \mathbb{R}^{nd \times d}$ . The eigenvalue  $\lambda_i^{mat}$  approximates  $\lambda_i$ , while the eigenfunction  $\phi_i$  in turn is approximated with

$$\tilde{\phi}_i(x) = \frac{\sqrt{n}}{\lambda_i^{mat}} k_X(x) u_i \approx \phi_i(x), \quad (.26)$$

where  $k_X(x) = (k(x_1, x), \dots, k(x_n, x))$ .

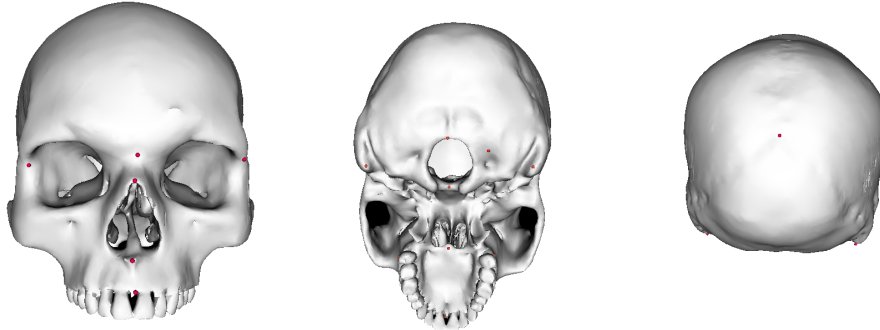
Clearly, the quality of this approximation improves with the number of points  $n$ , which are sampled. In our applications, we usually find that 500 to 1000 randomly sampled points are sufficient to obtain a good approximation. For a more thorough evaluation of how the number of points affects the approximation quality, we refer to Rosasco et al. [18] for a theoretical treatment, and Lüthi et al. for a practical evaluation for shape modelling [13].

## A.2. Obtaining the ground-truth

Due to the low quality of the available example shapes, which contains both missing parts (e.g. missing teeth), segmentation artifacts as well as metal artifacts around the teeth, we decided to perform our evaluations on a clean ground-truth dataset. The idea is to start from a clean reference shape (shown in Figure .12), which is used as an atlas, which is then registered on each data set. In this way it is ensured that our numerical comparison are not distorted by the large amount of artifacts in the data. For registration of the reference skull to the examples, we use the registration approach discussed in Section 3. As a shape prior we used a combination of a multi-scale kernel, which was symmetrized, plus two additional Gaussian kernel to allow for non-symmetric deformations.

$$k_{sym}(k_g^{50,100}(x, x') + k_g^{20,50}(x, x')) + k_g^{(2,100)}(x, x') + k_g^{(20,5)}(x, x'). \quad (.27)$$

The last kernel in Equation .27 allows for non-symmetric, local deformations and thus ensures that the model is not bias by the symmetric kernel and can accurately represent the target surface. To also retain this flexibility in the low-rank approximation, we represent the model using the leading 400 basis functions. The registration has



**Figure .12** 16 landmark points were defined on each skull, to enforce the correct correspondences for the ground-truth registrations. Integrating the landmark points improves the registration results especially for cases where there are large artifacts or holes in the data.

been performed using the same setup as discussed in Section 4.3. In order to ensure correct correspondences, we have in incorporated 16 landmarks, which are shown in Figure .12. In absence of meaningful registration metrics (which are not available as the original data is too noise) we ensured by visual inspection that the registration result accurately matched its target.

## REFERENCE

1. Petter Abrahamsen. *A review of Gaussian random fields and correlation functions*. Norsk Regnesentral/Norwegian Computing Center, 1997.
2. Yali Amit, Ulf Grenander, and Mauro Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86(414):376–387, 1991.
3. Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*, volume 3. Springer, 2004.
4. Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press, 1999.
5. T. F Cootes, C. J Taylor, D. H Cooper, J. Graham, and others. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1), 1995.
6. Brian P. Flannery, William H. Press, Saul A. Teukolsky, and William Vetterling. *Numerical recipes in c. Press Syndicate of the University of Cambridge, New York*, 1992.
7. Bay Xavier Roustant Olivier Carraro Laurent Ginsbourger, David. Argumentwise invariant kernels for the approximation of invariant functions. *Annales de la facult des sciences de Toulouse Mathematiques*, 21(3):501–527, 4 2012.
8. Ulf Grenander and Michael I Miller. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*, 56(4):617–694, 1998.
9. Tobias Heimann, Bram Van Ginneken, Martin Styner, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, György Bekes, et al. Comparison and evaluation of methods for liver segmentation from ct datasets. *Medical Imaging, IEEE Transactions on*, 28(8):1251–1265, 2009.
10. Peter J Huber. *Robust statistics*. Springer, 2011.
11. Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
12. Marcel Lüthi, Remi Blanc, Thomas Albrecht, Tobias Gass, Orcun Goksel, Philippe Bchler, Michael Kistler, Habib Bousleiman, Mauricio Reyes, Philippe C Cattin, and others. Statismo—a framework for PCA based statistical models. *Insight Journal*, 2012.
13. Marcel Lüthi, Christoph Jud, Thomas Gerig, and Thomas Vetter. Gaussian process morphable models. *arXiv preprint arXiv:1603.07254*, 2016.
14. C. A Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, 2005.
15. Andreas Morel-Forster. *Generative Shape and Image Analysis by Combining Gaussian Processes and MCMC Sampling*. PhD thesis, Universität Basel, 2016.
16. R. Opfer. Multiscale kernels. *Advances in Computational Mathematics*, 25(4):357–380, 2006.
17. C. E Rasmussen and C. K.I Williams. *Gaussian processes for machine learning*. Springer, 2006.
18. Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *The Journal of Machine Learning Research*, 11:905–934, 2010.
19. Scalismo - scalable image analysis and shape modelling. <http://github.com/unibas-gravis/scalismo>.
20. Kamal Shahim, Philipp Jürgens, Philippe C Cattin, Lutz-P Nolte, and Mauricio Reyes. Prediction of cranio-maxillofacial surgical planning using an inverse soft tissue modelling approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 18–25. Springer, 2013.
21. John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.
22. Martin A Styner, Kumar T Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gábor Székely, Christopher J Taylor, and Rhodri H Davies. Evaluation of 3d correspondence methods for model building. In *Information processing in medical imaging*, pages 63–75. Springer, 2003.
23. Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.



