Gaussian Process Morphable Models

Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter

Abstract—Models of shape variations have become a central component for the automated analysis of images. An important class of shape models are point distribution models (PDMs). These models represent a class of shapes as a normal distribution of point variations, whose parameters are estimated from example shapes. Principal component analysis (PCA) is applied to obtain a low-dimensional representation of the shape variation in terms of the leading principal components. In this paper, we propose a generalization of PDMs, which we refer to as *Gaussian Process Morphable Models* (GPMMs). We model the shape variations with a Gaussian process, which we represent using the leading components of its Karhunen-Loève expansion. To compute the expansion, we make use of an approximation scheme based on the Nyström method. The resulting model can be seen as a continuous analog of a standard PDM. However, while for PDMs the shape variation is restricted to the linear span of the example data, with GPMMs we can define the shape variation using any Gaussian process. For example, we can build shape models that correspond to classical spline models and thus do not require any example data. Furthermore, Gaussian processes make it possible to combine different models. For example, a PDM can be extended with a spline model, to obtain a model that incorporates learned shape characteristics but is flexible enough to explain shapes that cannot be represented by the PDM. We introduce a simple algorithm for fitting a GPMM to a surface or image. This results in a non-rigid registration approach whose regularization properties are defined by a GPMM. We show how we can obtain different registration schemes, including methods for multi-scale or hybrid registration, by constructing an appropriate GPMM. As our approach strictly separates modeling from the fitting process, this is all achieved without changes to the fitting algorithm. To demonstrate the applicability and versatility of GPMMs, we perform a set of experiments in typical usage scenarios in medical image analysis and computer vision: The model-based segmentation of 3D forearm images and the building of a statistical model of the face. To complement the paper, we have made all our methods available as open source.

Index Terms—Statistical shape modeling, Gaussian processes, image analysis, non-rigid registration

1 Introduction

The automatic interpretation and analysis of objects in an image is at the core of computer vision and medical image analysis. A popular approach is analysis by synthesis [1], which asserts that in order to explain an image, we need to be able to synthesize its content. This is achieved by fitting a probabilistic model to an image such that one-to-one correspondence between the model and the image is established. The image can then be explained using the model information. The better the model represents the structure of the objects to be analyzed, the easier it becomes to fit the model. For this reason statistical shape models have become very popular. An important class of statistical shape models are point distribution models (PDMs). PDMs represent object boundaries by their point positions. These are statistically analyzed using principal component analysis (PCA). Well known types of PDMs include the Active Shape Model [2] and the Morphable Model [3]. In its original formulation, the Active Shape Model represents the object boundary using a few manually defined landmark points. In the Morphable Model, the boundary is represented by a dense set of points, for which the correspondence is automatically determined by a registration algorithm. A variety of other statistical shape models focus on higher-order geometric features and/or analysis by statistical techniques that recognize the benefit of understanding shape relations in populations as being nonlinear [4]. We refer the reader to the survey paper of Heimann et al. [5] for a detailed overview of statistical shape models.

The focus of this work are point distribution models. PDMs are linear, parametric models and hence are mathematically convenient and easy to incorporate in image-analysis algorithms. Since they can represent only 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 a model that can express all possible target shapes, a lot of training data is needed.

The main contribution of this work is that we introduce a generalization of PDMs, which we refer to as Gaussian Process Morphable Models (GPMM). We model a shape as a deformation u from a reference shape $\Gamma_R \subset \mathbb{R}^3$; i.e., a shape s can be represented as

$$s = \{x + u(x) | x \in \Gamma_R\},$$

for some deformation $u : \Omega \to \mathbb{R}^3$, with $\Omega \supseteq \Gamma_R$. We model the deformations as a Gaussian process $u \sim GP(\mu, k)$ where $\mu : \Omega \to \mathbb{R}^3$ is a mean deformation and $k : \Omega \times \Omega \to \mathbb{R}^{3\times3}$ a

- *M. Lüthi, T. Gerig, and T. Vetter are with the Department of Mathematics and Computer Science, University of Basel, Basel 4001, Switzerland. E-mail: {marcel.luethi, thomas.gerig, thomas.vetter}@unibas.ch.*
- *C. Jud is with the Medical Image Analysis Center, University Hospital Basel, Basel 4056, Switzerland. E-mail: christoph.jud@unibas.ch.*

Manuscript received 4 Apr. 2016; revised 20 July 2017; accepted 24 July 2017. Date of publication 14 Aug. 2017; date of current version 11 July 2018. (Corresponding author: Marcel Lüthi.) Recommended for acceptance by E. G. Learned-Miller. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TPAMI.2017.2739743

covariance function or kernel. Note that in contrast to classical PDMs, our definition allows for the possibility to define the boundary Γ_R continuously. The core idea behind our approach is that we obtain a parametric, low-dimensional model by representing the Gaussian process using the r leading basis function $\phi_i : \Omega \to \mathbb{R}^3$ of its Karhunen-Loève expansion

$$u = \mu + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i} \phi_i, \ \alpha_i \in \mathcal{N}(0,1), \tag{1}$$

(here, λ_i is the variance associated with each basis function ϕ_i). As we usually assume strong smoothness of the deformations when modeling shapes, it is often possible to achieve good approximations using only a few leading basis functions, which makes the representation practical. The main difficulty of this approach is to efficiently compute the leading eigenfunction/eigenvalue pairs. To this end, we propose to use a Nyström approximation and make use of a recently introduced computational approach which is able to use a large number of input points for computing the approximation [6].

The biggest advantage of GPMMs compared to PDMs is that we have much more freedom in defining the covariance function. As a second main contribution we will show in Section 3 how expressive prior models for registration can be derived by leveraging the modeling power of Gaussian processes. By estimating the covariances from example data our method becomes a continuous version of a PDM. When we have no or only little training data available, arbitrary kernel functions can be used to define the covariances. In particular, we can define models of smooth deformations using spline models or radial basis functions, which are frequently used in registration approaches. We show how a simple registration approach, whose regularization properties are defined in terms of a GPMM, allows us to use these models for actual surface and image registration. Besides these simple models, GPMMs also make it possible to combine different covariance functions (or kernels) to mimic more sophisticated registration schemes. We show how to construct priors that have multi-scale properties or can incorporate landmark constraints. We will also show how to combine models learned from training data with analytically defined covariance functions in order to increase the flexibility of PDMs in cases where not sufficient training data is available. Although in contrast to PDMs, GPMMs model deformations defined on a continuous domain, we can always discretize it to obtain a model that is mathematically equivalent to a PDM. This makes it possible to leverage the modeling flexibility of GPMMs also in classical shape modeling algorithms, such as for example the Active Shape Model fitting [2] algorithm or the coherent point drift method [7].

To show the versatility and effectiveness of GPMMs we performed experiments in two typical application scenarios of GPMMs in medical image analysis and computer vision. The first application scenario considers the model-based segmentation and registration of CT images of the forearm. We discuss how to build a model which is specifically tailored to the task of forearm registration. In a second experiment we performed Active Shape Model fitting and show how combining the learned model with an analytically defined prior can improve the segmentation accuracy. Further, we present an application of GPMMs for 3D image-to-image registration and compare the result to the popular B-spline registration method implemented in Elastix [8]. In the second application scenario, we discuss how GPMMs give rise to a new strategy for building face models, which can make better use of the available data and can reduce the need for tedious manual annotations. Finally, we show on a qualitative example how GPMMs can improve an existing face models to better fit faces that are not represented in the original model.

As supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2017.2739743, to this article we provide a study of the approximation properties of our numerical methods and discussion how the approximation quality is influenced by different choices of covariance functions. Furthermore, we discuss how to choose the parameters of our method in order to reach a given approximation quality. All our methods are implemented as part of the open source software Statismo [9] and Scalismo [10].

1.1 Related Work

Our work can be seen as the unification of two different concepts: On one hand, we extend PDMs such that they become more expressive. On the other hand we model prior distributions for surface and image registration. There are works from both the shape modeling and the registration community which are conceptually similar or have the same goals as we pursuit with our approach. Most notably, the work of Wang and Staib [11], which aims for extending the flexibility of shape models, and the work by Grenander et al. [12], who use Gaussian processes as priors for registration are very close in spirit to our model. The idea of Wang and Staib is to extend the flexibility of a PDM by combining a learned covariance matrix used in a point distribution model with covariance matrices that represent other, synthetic deformations. This corresponds exactly to our idea for combining covariance functions in the GP setting. However, their method requires that the full covariance matrix can be represented, which is only feasible for very coarsely discretized shapes. In contrast, our method yields a continuous representation and allows for an arbitrarily fine discretization once the prior is evaluated in the final registration procedure. On the registration side, the use of Gaussian processes for image registration has been extensively studied in the 90s by Grenander et al. (see the overview article [12] and references therein). Similar to our approach, they propose to use a basis function representation to span the model space. However, in all these works the basis functions have to be known analytically [13], or the initial model needs to be of finite rank [14]. In our method we use the Nyström approximation to numerically approximate the leading eigenfunctions, which makes it possible to approximate any Gaussian process and thus to allow us to use arbitrary combinations of kernels in our models. We believe that this modeling flexibility is what makes this approach so powerful.

Besides the above mentioned works that aim at unifying the concepts of priors for registration and shape modeling, there is a huge body of literature on non-rigid registration,

to which our method is directly relevant. Non-rigid registration is a fundamental problem in computer graphics, computer vision and medical image analysis [15], [16], [17]. A comprehensive overview of recent approaches for non-rigid registration used in medical image analysis is given by Sotiras et al. [16]. In computer vision, very similar registration methods to those used in medical image analysis are used in optical flow computation [15]. Registration approaches are also of importance in other areas of computer vision, such as for example in stereo matching [18] or finding corresponding matching points in images [19]. An important part of all these methods is to formulate appropriate prior assumption over the possible deformation fields, and hence the ideas proposed in this paper are directly applicable.

The space of admissible deformations represented by GPMMs is defined using a covariance function (or kernel). Mathematically this space corresponds to a Reproducing Kernel Hilbert Space (RKHS). There are many other works that propose to model the admissible deformations for non-rigid registration by means of a kernel and RKHS. Especially for landmark based registration, spline based models and radial basis functions have been widely used [20]. The algorithm for solving a standard spline-based landmark registration problem corresponds to the MAP solution in Gaussian process regression [21]. Using Gaussian process regression for image registration has been proposed by Zhu et al. [22]. A similar framework for surface registration, where kernels are used for specifying the admissible deformation was proposed by Steinke et al. [23]. While they do not provide a probabilistic interpretation of the problem, their approach results in the same final registration formulation as our approach. The use of Reproducing Kernel Hilbert Spaces for modeling admissible deformation also plays an important role for diffeomorphic image registration (see, e.g., [24], Chapter 9). In this context, it has also been proposed to combine basic kernels for multi-scale [25], [26] and spatially-varying models [27] for registration. However, the work focuses more on the mathematical and algorithmic aspects of enforcing diffeomorphic mappings rather than the modeling aspect.

Besides the work of Wang and Staib [11] there have been many other works for extending the flexibility of PDMs. This is typically achieved by adding artificial training data [28] or by segmenting the model either spatially [3], [29] or in the frequency domain [30], [31]. The use of Gaussian processes to model the covariance structure is much more general and subsumes all these methods. Another set of work gives shape model based algorithms more flexibility for explaining a target solution [32], [33], [34]. Compared to our model, these approaches have the disadvantage that the model is not generative anymore and does not admit a clear probabilistic interpretation.

This paper is a summary and extension of our previous conference publications [35], [36], [37]. It extends our previous work in several ways: 1) It provides an improved presentation of the basic method and in particular its numeric implementation. 2) It provides an analysis of the approximation properties of this scheme. 3) It proposes new combinations of kernels to combine point distribution models with GPMMs based on analytically defined kernels. 4) It features a more detailed validation including surface and image registration, as well as Active Shape Model fitting.

2 Gaussian Process Morphable Models

Before describing GPMMs, we summarize the main concepts behind point distribution models, on which we will build up our work.

2.1 Point Distribution Models

The main assumption behind point distribution models is that the space of all possible shape deformations can be learned from a set of typical example shapes $\{\Gamma_1, \ldots, \Gamma_n\}$. Each shape Γ_i is represented as a discrete set of landmark points; i.e.,

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

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 kth landmark point x_k^i and x_k^j of two shapes Γ_i and Γ_j represent 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 Γ_i is represented as a vector $s_i \in \mathbb{R}^{3N}$, where the $x, y, z-$ components of each point are stacked onto each other:

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

This vectorial representation makes it possible to apply the 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

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

where the mean μ and covariance matrix Σ are estimated from the example data

$$\mu = \overline{s} := \frac{1}{n}\sum_{i=1}^{n} \vec{s_i} \tag{2}$$

$$\Sigma = S := \frac{1}{n-1}\sum_{i=1}^{n} (\vec{s_i} - \overline{s})(\vec{s_i} - \overline{s})^T. \tag{3}$$

As the number of points N is usually large, the covariance matrix Σ cannot be represented explicitly. Fortunately, as it is determined completely by the n example data-sets, it has at most rank n and can therefore be represented using n basis vectors. This is achieved by performing a Principal Component Analysis [38]. In its probabilistic interpretation, PCA leads to a model of the form

$$s = \overline{s} + \sum_{i=1}^{n} \alpha_i \sqrt{d_i}\vec{u_i}, \tag{4}$$

where (u_i, d_i), $i = 1, \ldots, n$, are the eigenvectors and eigenvalues of the covariance matrix S. Assuming that $\alpha_i \sim \mathcal{N}(0, 1)$ in (4), it is easy to check that $s \sim \mathcal{N}(\overline{s}, S)$. Thus, we have a efficient, parametric representation of the distribution.

2.2 Gaussian Process Morphable Models

The literature on PDMs usually emphasizes the shapes that are modeled. Equation (4) however, gives rise to a different interpretation: A point distribution model is a model of deformations $\vec{\phi} = \sum_{i=1}^{n} \alpha_i \sqrt{d_i}\vec{u}_i \sim \mathcal{N}(0, S)$ which are added to a mean shape $\overline{s}$. 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 Γ_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 $u \in \mathcal{GP}(\mu, k)$ with mean function $\mu : \Omega \to \mathbb{R}^3$ and covariance function $k : \Omega \times \Omega \to \mathbb{R}^{3\times 3}$. Note that any deformation $\hat{u}$ sampled from $\mathcal{GP}(\mu, k)$ gives rise to a new shape by warping the reference shape Γ_R

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

Similar to the PCA representation of the PDM used in (Equation (4)), a Gaussian process $\mathcal{GP}(\mu, k)$ can be represented in terms of an orthogonal set of basis functions $\{\phi_i\}_{i=1}^{\infty}$

$$u(x) \sim \mu(x) + \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i}\phi_i(x),\ \alpha_i \in \mathcal{N}(0, 1), \tag{5}$$

where (λ_i, ϕ_i) are the eigenvalue/eigenfunction pairs of the integral operator

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

and $\rho(x)$ denotes a measure. The representation (5) is known as the Karhunen-Loève expansion of the Gaussian process [39]. Since the random coefficients α_i are uncorrelated, the variance of u is given by the sum of the variances of the individual components. Consequently, the eigenvalue λ_i corresponds to the variance explained by the ith component. This suggests that if the λ_i decay sufficiently quickly, we can use the low-rank approximation

$$\tilde{u}(x) \sim \mu(x) + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\phi_i(x), \tag{6}$$

to represent the process. 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 model. Note however, that there is no restriction that the covariance function k needs to be learned from examples, as is required for the sample covariance matrix in (2). Any valid positive definite covariance function can be used. As we will show in Section 3 this makes it possible to define powerful prior models, even when there is little or no example data available.

2.3 Computing the Eigenfunctions

The low-rank approximation (6) can only be performed if we are able to compute the eigenfunction/eigenvalue pairs $(\phi_i, \lambda_i)_{i=1}^{r}$. Although for some kernel functions analytic solutions are available (see, e.g., [12], [13]), for most interesting models we need to resort to numeric approximations. A classical method, which has recently received renewed attention from the machine learning community, is the Nyström method [21]. The goal of the Nyström method is to obtain a numerical estimate for the eigenfunctions/eigenvalues of the integral operator

$$\mathcal{T}_k f(\cdot) := \int_{\Omega} k(x, \cdot) f(x) d\rho(x), \tag{7}$$

i.e., the pairs (ϕ_i, λ_i), satisfying the equation

$$\lambda_i \phi_i(x') = \int_{\Omega} k(x, x')\phi_i(x)\, d\rho(x),\ \forall x' \in \Omega, \tag{8}$$

are sought. The Nyström method is intended to approximate the integral in (8). 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 Ω, and to randomly sample points $X = \{x_1, \ldots, x_n\}, x_l$ according to p. The samples $(x_l)_{l=1,\ldots,n}$ for x' in (8) lead to the matrix eigenvalue problem

$$Ku_i = \lambda_i^{mat} u_i, \tag{9}$$

where $K_{il} = k(x_i, x_l)$ is the kernel matrix, u_i denotes the ith eigenvector and λ_i^{mat} the corresponding eigenvalue. The eigenvalue λ_i^{mat} approximates λ_i, while the eigenfunction ϕ_i in turn is approximated with

$$\tilde{\phi}_i(x) = \frac{\sqrt{n}}{\lambda_i^{mat}} k_X(x) u_i \approx \phi_i(x), \tag{10}$$

where $k_X(x) = (k(x_1, x), \ldots, k(x_n, x))$. Note that since the kernel is matrix valued ($k : \Omega \times \Omega \to \mathbb{R}^{d\times d}$), the matrices K and k_X are block matrices: $K \in \mathbb{R}^{nd\times nd}$ and $k_X \in \mathbb{R}^{nd\times d}$.

Clearly, the quality of this approximation improves with the number of points n that are sampled (see the supplementary material, available online, for a detailed discussion). As n becomes larger (i.e., exceeds a few thousand points), deriving the eigenvalue problem (9) might still be computationally infeasible. Following Li et al. [6], we therefore apply a random SVD [40] for efficiently approximating the first eigenvalues/eigenvectors without having to compute the eigenvalues of the full $N \times N$ matrix. Theoretical bounds of the method [40], as well as its application for the Nyström approximation [6], show that it leads to accurate approximation for kernels with a fast decaying spectrum. For our application, the error induced by the random SVD is negligible compared to the approximation error caused by the low-rank approximation and the Nyström method.

2.3.1 Accuracy of the Low-Rank Approximation

It is clear that our method depends crucially on the quality of the low-rank approximation. Ideally, we would like to see the low-rank model as a convenient reparametrization of the original process, which would not affect the shape variations that are spanned by our model. This is indeed the case when there are strong correlations in the deformations, which is

Fig. 1. The 3D face surface used to illustrate the effect of different models.

Fig. 2. Samples using a Gaussian kernel with scale factor $s = 10$ and bandwidth $\sigma = 100$ mm.

for example the case when the modeled deformations are smoothly varying over the domain. Fortunately, in shape modeling it is usually justified to make strong smoothness assumptions, and thus the approximation works well in practice. However, if we want to model very small, local variations, the quality of the low-rank approximation starts to deteriorate. In order not to digress from the main theme of the paper, we have put a detailed discussion of these issues into the supplementary material, available online.

2.3.2 Computational Complexity and Practical Implementation

From Equation (6) we see that for evaluating a deformation u at a point x, the sum over the r eigenfunctions evaluated at x needs to be computed. This in turn requires n evaluations of the kernel function to compute the $nd-$vector $k_X(x)$ in (10). For a fixed covariance function the complexity is clearly linear in r and n. In shape modeling applications, we often need to compute the deformations for all the points of a densely represented surface, which might have hundred-thousands of points. Depending on the kernel and the number of eigenfunctions to be computed, the computation might take several minutes. An effective strategy in such cases is to discretize the mean function μ and eigenfunctions $\phi_i, i = 1, \ldots, r$ in an offline step for all the points of interest. We note that for a fixed discretization $\tilde{\Omega} = \{x_1, \ldots, x_N\}$ we can define the vectors

$$\overline{s} := (\mu(x_1)_x, \mu(x_1)_y, \mu(x_1)_z, \ldots, \mu(x_N)_x, \mu(x_N)_y, \mu(x_N)_z)^T,$$

and

$$u_i := (\phi_i(x_1)_x, \phi_i(x_1)_y, \phi_i(x_1)_z, \ldots, \phi_i(x_N)_x, \phi_i(x_N)_y, \phi_i(x_N)_z)^T,$$

and arrive at a model of the same mathematical form as the classical point distribution model defined in (4). Consequently, it can be used as a replacement in any algorithm that uses classical point distribution models, and it has exactly the same runtime complexity.

3 Modeling with Kernels

The formalism of Gaussian processes provides us with a rich language to model shape variations. In this section we explore some of these modeling possibilities, with a focus on models that we find most useful in our work on model-based image analysis and in particular surface and image registration. Many more possibilities for modeling with Gaussian processes have been explored in the machine learning community (see, e.g., Duvenaud, Chapter 2 [41]).

To visualize the shape variations represented by a model, we define a GPMM on the face surface (see Fig. 1) and show the effect that randomly sampled deformation from this model have on the face surface.[1] Using the face for visualizing shape variations has the advantage that we can judge how anatomically valid a given shape deformation is.

3.1 Models of Smooth Deformations

A simple Gaussian process model is a zero mean Gaussian process that enforces smooth deformations. The assumption of a zero mean is typically made in registration tasks. It implies that the reference surface is a representative shape for the class of shapes which are modeled or, in other words, that the shape is close to a (hypothetical) mean shape. A particularly simple kernel that enforces smoothness is the Gaussian kernel defined by

$$k_g(x, y) = \exp(-\|x - y\|^2/\sigma^2),$$

where σ^2 defines the range over which deformations are correlated. Hence the larger the values of σ, the more smoothly varying the resulting deformations fields will be. In order to use this scalar-valued kernel for registration, we can define a matrix valued kernel as

$$k(x, y) = s \cdot I_{3\times3} k_g(x, y),$$

where the identity matrix $I_{3\times3}$ signifies 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. Fig. 2 shows random samples from the model for two different values of σ. This construction can be generalized by defining the matrix valued kernel as

$$k(x, y) = A k_g(x, y) A^T,\ A \in \mathbb{R}^{3\times3}, \tag{11}$$

which allows us to introduce anisotropic scaling and correlations between the components.

Besides Gaussian kernels, there are many different kernels that are known to lead to smooth functions. For registration purposes, spline models, Elastic-Body Splines [43] B-Splines [44] or Thin Plate Splines [45] are maybe the most commonly used ones.

3.2 Point Distribution Models

An ideal prior for the registration of faces would only allow valid face shapes. This is the motivation behind PDMs [2], [3]. The characteristic deformations are learned from a set of typical examples surfaces $\Gamma_1, \ldots, \Gamma_n$. More precisely, by establishing correspondence, between a reference Γ_R and each of the

1. This face is the average face of the publicly available Basel Face Model [42].

Fig. 3. Samples using a sample covariance kernel, which is learned from 200 training faces. All the random samples look like valid faces.

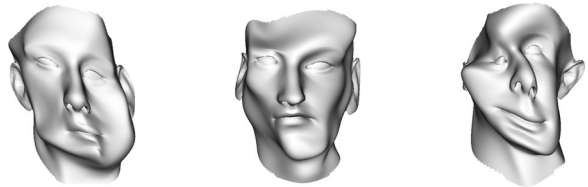


Fig. 4. Samples using a kernel defined on multiple scales. The random sample show large deformations which change the overall face shape, as well as local, detailed shape variations.

training surfaces, we obtain a set of deformation fields $\{u_1, \ldots, u_n\}, u_i : \Omega \to \mathbb{R}^d$, where $u_i(x)$ denotes a deformation field that maps a point on the reference $x \in \Gamma_R$ to its corresponding point $u_i(x)$ on the ith training surface. A Gaussian process $\mathcal{GP}(\mu_{\text{PDM}}, k_{\text{PDM}})$ that models these characteristic deformations is obtained by estimating the empirical mean

$$\mu_{\text{PDM}}(x) = \frac{1}{n}\sum_{i=1}^{n} u_i(x),$$

and covariance function

$$k_{\text{PDM}}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(u_i(x) - \mu_{\text{PDM}}(x))(u_i(y) - \mu_{\text{PDM}}(y))^T. \tag{12}$$

We refer to the kernel k_{PDM} as the *sample covariance kernel* or empirical kernel. Samples from such a model are depicted in Fig. 3, where the variation was estimated from 200 face surfaces from the Basel Face Model [42]. In contrast to the smoothness priors, all the sampled face surfaces represent anatomically plausible faces. The model that we obtain using this sample covariance kernel is a continuous analog to a PCA based shape model.

3.3 Combining Kernels

The real power of modeling with Gaussian processes comes to bear if the "simple" kernels are combined to define new kernels, making use of a rich algebra that kernels admit. In the following, we present basic combinations of kernels to give the reader a taste of what can be achieved. For a more thorough discussion of how positive definite kernels can be combined, we refer the reader to Shawe-Taylor et al. [46] (Chapter 3, Proposition 3.22).

3.3.1 Multiscale Models

If $k_1, \ldots, k_n : \Omega \times \Omega \to \mathbb{R}^{d\times d}$ are positive definite kernels, then the linear combination

$$k_S(x,x') = \sum_{i=1}^{l} \alpha_i k_i(x,x'),\ \alpha_i \in \mathbb{R},$$

is positive definite as well. This provides a simple means of modeling deformations on multiple scale levels by summing kernels that model smooth deformations with kernels for more local, detailed deformation. A particularly simple implementation of such a strategy is to sum up Gaussian kernels, with decreasing scale and bandwidth

$$k_{MS}(x,x') = \sum_{i=1}^{l} \frac{s}{i} I_{3\times 3} \exp\left(-\frac{\|x-x'\|^2}{(\sigma/i)^2}\right),$$

where s determines the base scale and σ the smoothness and l the number of levels. As shown in Fig. 4, this simple approach already leads to a multiscale structure that models both large scale deformations as well as local details.

3.3.2 Reducing the Bias in Point Distribution Models

Due to the limited availability of training examples, point distribution models are often not able to represent the full shape space accurately and thus introduce a bias towards the training shapes into model-based methods [34]. One possibility to avoid this problem is to provide an explicit bias model, which is added to the point distribution model. Denote by $k_{\text{PDM}} : \Omega \times \Omega \to \mathbb{R}^{d\times d}$ the sample covariance kernel and let $k_g : \Omega \times \Omega \to \mathbb{R}$ be a Gaussian kernel with bandwidth parameter σ. A simple model to reduce the bias would be to use a Gaussian kernel with a large bandwidth; i.e., we define

$$k_b(x,x') = k_{\text{PDM}}(x,x') + sI_{3\times 3}k_g(x,x'),$$

where the parameter s defines the scale of the average error. This parameter could, for example, be estimated using crossvalidation. This simple model assumes that the error is spatially correlated; i.e., if a model cannot explain the structure at a certain point, its neighboring points are likely to also show the same error.

3.3.3 Localizing Point Distribution Models

Another possibility to obtain more flexible models is to make models more local by breaking the global correlations. Recalling that the kernel function $k(x,x')$ models the correlation between the points x and x'. Setting the correlation $k(x,x')$ to 0 for $x \neq x'$ decouples the points and hence increases the flexibility of a model. Such an effect can be achieved by a multiplication of two kernel functions, which again results in a positive definite kernel. A simple example of a local model is obtained by multiplying a kernel with a Gaussian kernel with small bandwith σ. For example, by defining

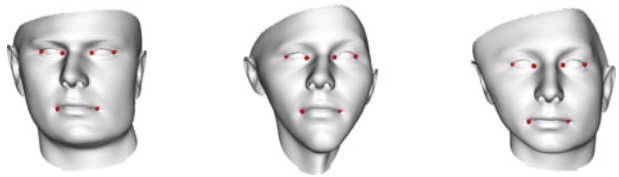
$$k_l(x,x') = k_{\text{PDM}}(x,x') \odot I_{3\times 3}k_g(x,x'),$$

(where $\odot$ defines element-wise multiplication), we obtain a localized version of a point distribution model. Samples from such a model are shown in Fig. 5. We observe that the samples locally look like valid faces, but globally, the kernel still allows for more flexible variations, which could not be described by the model, and which may not constitute an anatomically valid face.

3.4 Posterior Models

In many applications we have not only information about the correlations but know for certain points exactly how

Fig. 5. Random samples form a localized point distribution model. Whereas the variations look anatomically valid locally, there are no global correlations anymore, which makes the model more flexible.



(a) $s = 10, \sigma = 30mm$

Fig. 6. Random samples from a posterior model, which has been obtained by taking the Gaussian process model shown in Fig. 2, and applying Gaussian process regression to keep the points shown in red fixed.

they should be mapped. Assume for instance, that a user has clicked a number of landmark points on a reference shape $L_R = \{l_R^1, \ldots, l_R^n)\}$ together with the matching points on a target surface $L_T = \{l_T^1, \ldots, l_T^n)\}$. From the landmarks we can compute the deformation $\hat{u}^i$ for each point l_R^i of the reference shape

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

Let $u \sim GP(\mu, k)$ be a Gaussian process model and assume further that the observations $\hat{u}^i$ are subject to Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma I_{3\times 3})$. It turns out that the distribution $u|l_R^1, \ldots, l_R^n, \hat{u}^1, \ldots, \hat{u}^n$ is again a Gaussian process $GP(\mu_p, k_p)$ whose mean μ_p and covariance k_p are known in closed form. This construction is known as Gaussian process regression. We refer to [21], Chapter 2, and [35] for the mathematical details. Fig. 6 shows random samples from such prior, where the points shown in red were fixed by setting $(\hat{u}^i = (0,0,0)^T), i = 1, \ldots, n)$.

4 Registration Using Gaussian Process Morphable Models

In this section we show how we can use GPMMs as prior models for surface and image registration. The idea is that we define a model for the variability of a given object $O_R \subset \mathbb{R}^d$ and fit this model to a target object $O_T \subset \mathbb{R}^d$, which is either represented as a surface or an image. Our main assumption is that we can identify for each point $x_R \in O_R$, a corresponding point $x_T \in O_T$ of the target object O_T. More formally, it is assumed that there exists a deformation $u : \Omega \to \mathbb{R}^d$ such that

$$O_T = \{x + u(x) | x \in O_R\}.$$

The goal of the registration problem is to recover the deformation field u, which relates the two objects. To this end, we formulate the problem as a MAP estimate

$$\arg\max_u p(u) p(O_T | O_R, u), \tag{13}$$

where $p(u) \sim \mathcal{GP}(\mu, k)$ is a Gaussian process prior over the admissible deformation fields and the likelihood function $p(O_T|O_R, u)$ is defined as

$$p(O_T|O_R, u) = \frac{1}{Z} \exp(-\eta^{-1} \mathcal{D}[O_R, O_T, u]).$$

Here $\mathcal{D}$ is a metric that measures the similarity of the objects O_T and O_R, $\eta \in \mathbb{R}$ 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 (13) we arrive at the equivalent minimization problem

$$\arg\min_u \mathcal{D}[O_R, O_T, u] - \eta \ln p(u). \tag{14}$$

Using the low-rank approximation (Equation (6)) we can restate the problem in the parametric form

$$\arg\min_{\alpha_1, \ldots, \alpha_r} \mathcal{D}\left[O_R, O_T, \mu + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i} \phi_i\right] + \eta' \sum_{i=1}^{r} \alpha_i^2, \tag{15}$$

where we used that the coefficients α in (6) are independent and hence $p(u) \propto \exp(-\sum_{i=1}^r \alpha_i^2)$.

The final registration formulation (15) is highly appealing. All the assumptions are represented by the eigenfunctions $\phi_i, i = 1 \ldots, r$, which in turn are determined by the Gaussian process model. Thus, we have split the registration problem into three separate problems:

(1) *Modeling:* Specify a model for the deformations by defining a Gaussian process model for the deformations $u \sim \mathcal{GP}(\mu, k)$
(2) *Approximation:* Approximate the model by replacing in parametric form $\tilde{u} = \mu + \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} \phi_i \sim \mathcal{GP}(\mu, \tilde{k})$, in terms of its eigendecomposition.
(3) *Fitting:* Fit the model to the data by minimizing the optimization problem (15).

The separation of the modeling and the fitting step is most important, as it allows us to treat the conceptual work of modeling our prior assumptions independently from the search of a good algorithm to actually perform the registration. Indeed, in this paper we will use the same, simple fitting approach for both surface and image fitting, which we detail in the following.

4.1 Model Fitting for Surface and Image Registration

To turn the conceptual problem (15) into a practical one, we need to specify the representations of the reference and target object O_R, O_T and define a distance measure $\mathcal{D}$ between them.

We start with the case where the object O_R, O_T correspond to surfaces $\Gamma_R, \Gamma_T \subset \mathbb{R}^3$. A simple measure $\mathcal{D}$ is the mean squared distance from the reference to the closest target point, i.e.,

$$D[\Gamma_R, \Gamma_T, u] = \int_{\Gamma_R} (CP_{\Gamma_T}(x + u(x)))^2 \, dx,$$

where CP_{Γ_T} is the distance function defined by

$$\mathrm{CP}_{\Gamma_T}(x) = \|x - \arg\min_{x' \in \Gamma_T} \|x - x'\|\|.$$

Hence, for the case of surface registration, the registration problem (15) becomes

$$\arg\min_{\alpha} \int_{\Gamma_r} \mathrm{CP}_{\Gamma_T}\left(x + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\phi_i(x)\right) dx + \eta \sum_{i=1}^{r} \alpha_i^2. \quad (16)$$

Note that for surface registration, we are only interested in deformations defined on Γ_R. It is therefore sufficient to compute the Nyström approximation using only points sampled from the reference Γ_R.

The second important case is image registration. Let $I_R, I_T : \Omega \to \mathbb{R}$ be two images defined on the image domain Ω. In this case, we usually choose $\mathcal{D}$ such that it integrates some function of the image intensities over the two images (see, e.g., [16] for an overview of different similarity measures). In the simplest case, we can use the squared distance of the intensities. The image registration problem becomes

$$\arg\min_{\alpha} \int_{\Omega} \left[I_R(x) - I_T\left(x + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\phi_i(x)\right)\right]^2 dx + \eta \sum_{i=1}^{r} \alpha_i^2. \quad (17)$$

Note that to be well defined, the Gaussian process needs to be defined on the full image domain Ω. Therefore, we sample points from the full image domain Ω to compute the Nyström approximation.

Independently of whether we do surface or image registration, we can easily obtain a hybrid registration scheme by including landmarks directly into the model using a posterior model (cf. Section 3.4). Furthermore, besides these straight-forward algorithms for surface and image registration, we can also directly make use of any algorithm that is designed to work with classical PDMs, such as for example the Active Shape Model fitting method [2]. This is possible because the model (6) is of the same form as a PDM, with the only difference that we have continuously defined basis function. As discussed in Section 2.3.2 we can obtain the same representation as a classical PDM by discretizing the basis functions for a given set of points.

5 Results

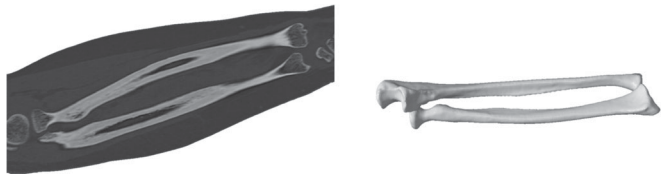
In this section we illustrate the use of GPMMs in typical application scenarios from medical image analysis and computer vision.

5.1 Model-Based Segmentation of Forearm CT Images

We start with a discussion on how to build an application-specific prior model of the ulna bone using analytically defined kernels. We use this model to perform surface registration in order to establish correspondence between a set of ulna-surfaces, and thus to be able to build a point distribution model. In a second experiment we use this model to perform Active Shape Model fitting and show how increasing the model's flexibility using a GPMM improves the results. Finally, we also show an application of GPMMs for image registration.

5.1.1 Experimental Setup

Our data consists of 36 segmented images of the right forearm bones (ulna and radius). For 27 of these bones we have the original CT image. Using the 36 given segmentations, we extracted the ulna surface using the marching cubes algorithm [47]. We chose an arbitrary data-set as a reference and defined on each ulna surface 4 landmark points (two on the proximal, two on the distal part of the ulna), which we used to rigidly aligned the original images, the segmentation as well as the extracted surfaces to the reference image [48]. Fig. 7 shows a typical CT image and the forearm bones.



(a) CT Image Slice (b) Surface

Fig. 7. A slice through a CT image of the forearm (left) and the extracted bone surface from a ground-truth segmentation.

We integrated GPMMs in the open source shape modeling software libraries Scalismo [10] and Statismo [9]. We used Scalismo for model-building, surface registration and Active shape model fitting. For performing the image registration experiments, we used Statismo, together with the Elastix toolbox for non-rigid image registration [8].

5.1.2 Building Prior Models

The first step in any application of GPMMs is building a suitable model. In the first two examples, we concentrate on the ulna. We know from prior experience that the deformations are smooth. We capture this by building our models using a Gaussian kernel

$$k_g^{(s,\sigma)}(x, x') = s\mathcal{I}_{3\times 3}\exp(-\|x - x'\|^2/\sigma^2),$$

where s determines the scale of the deformations and σ the smoothness. The simplest model we build is an isotropic Gaussian model defined using only a single kernel $k_g^{(100,100)}(x, x')$. The next, more complex model is an (isotropic) multi-scale model that models deformations on different scale levels

$$k_{ms}(x, x') = \sum_{i=1}^{3} k_g^{(100/i,100/i)}(x, x').$$

In the third model, we include the prior knowledge that for the long bones, the dominant shape variation corresponds to the length of the bone. Using the construction given in Equation (11) we can define the anisotropic covariance function

$$k_{ams}(x, x') = RSk_g^{(150,100)}(x, x')S^T R^T + k_g^{(50,50)}(x, x') + k_g^{(30,30)}(x, x'),$$

where $R \in \mathbb{R}^{3\times 3}$ is the matrix of the main principal axis of the reference and $S = \mathrm{diag}(1, 0.1, 0.1) \in \mathbb{R}^{3\times 3}$ is a scaling matrix. Multiplying with the matrix SR has the effect that the scale of the deformations in the direction of the main principal axis (i.e., the length axis) is amplified 10 times compared to the deformations in the other space directions. We compute for each model the low-rank approximation, where we choose the number of basis functions such that 99 percent of the total variance of the model is approximated.

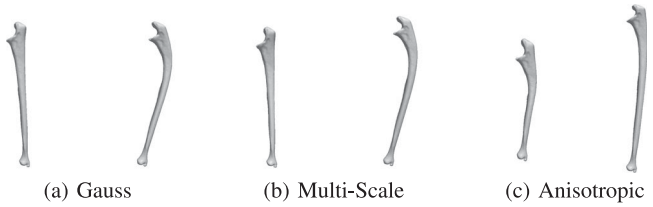


Fig. 8. The effect of varying the first two modes of variation for each model.

Fig. 8 shows the first mode of variation of the three models. We observe that for the anisotropic model, the main variation is almost a pure scale variation in the length axis, while in the other models it goes along with a bending of the bone.

Following Styner et al. we evaluate these three models using the standard criteria generalization, specificity and compactness [49]. Generalization refers to the model's ability to accurately represent all valid instances of the modeled class. We will discuss it in the next section. Specificity refers to the model's ability to only represent valid instances of the modeled bone. It is evaluated by randomly sampling instances of the model and then determining their distance to the closest example of a set of anatomically normal training examples. Compactness is the accumulated variance for a fixed number of components. This reflects the fact that if two models have the same generalization ability, we would prefer the one with less variance. Table 1 summarizes the specificity and compactness for these models. We evaluated both measures once consider only the first component, and once with the full model. We see the anisotropic model is more specific and more compact than the other models, which means that it should lead to more robust results in practical applications.

5.1.3 Surface Registration

To evaluate the generalization ability, we need to determine how well the model can represent valid target shapes, by fitting the model to typical shape surfaces. To fit the model, we use the surface registration algorithm presented in Section 4.1. Fig. 9 shows a boxplot with the generalization results. We also see that the multi-scale and the anisotropic model lead to similar results, but both outperform model where only a simple Gaussian kernel was used. That the anisotropic model can fit the models with the same accuracy as the multi-scale model, despite being much more compact, means that it is clearly better targeted to the given application. We we will see in the last experiment, this is a big advantage in more complicated registration tasks, such as image to image registration.

TABLE 1
The Specificity and Compactness Values Computed for Each of the Three Models

Model	Specificity		Compactness	
	1st PC	Full model	1st PC	Full model
Gauss	2.6	5.8	50.6	299.1
Isotropic Multi-Scale	2.3	6.1	51.1	317.0
Anisotropic Multi-Scale	1.9	2.9	51.1	137.1

The lower the specificity and compactness the better.

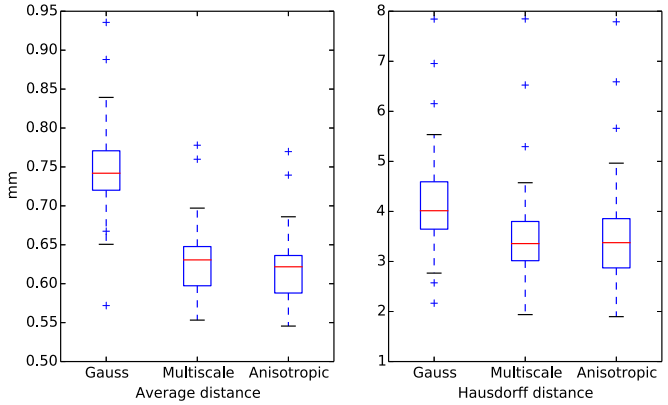


Fig. 9. Generalization ability measured by fitting the different models to all ulna surfaces.

5.1.4 Generalized Active Shape Model Fitting

The well known Active Shape Modeling approach [2] can be interpreted as a special case of Gaussian process registration as introduced in Section 4, where the model is a classical PDM (i.e., the sample mean and covariance kernel (12) are used) and an iterative algorithm is used to fit the model to the image. Active Shape Model fitting is a very successful technique for model-based segmentation. Its main drawback is that the solution is restricted to lie in the span of the underlying point distribution model, which might not be flexible enough to accurately represent the shape. In our case, where we have only 36 datasets of the ulna available, we expect this to be a major problem.

To build an Active Shape Model, we use the fitting results obtained in the previous section together with the original CT images as training data. Besides a standard ASM, we use the techniques for enlarging the flexibility of shape models discussed in Section 3, to build an extended model with additive smooth deformations (cf. Section 3.3.2), and a "localized" model (cf. Section 3.3.3). In the first case, we use a Gaussian kernel $k_g^{(3,100)}$ to model the unexplained part. Also for localization we choose a Gaussian kernel $k_g^{(1,100)}$, but this time with scale 1, in order not to change the variance of the original model. In both cases, we approximate the first 100 eigenfunctions. Fig. 10 shows the corresponding fitting result from a leave-one-out experiment. We see that both the extended and the localized model improve the results compared to the standard Active

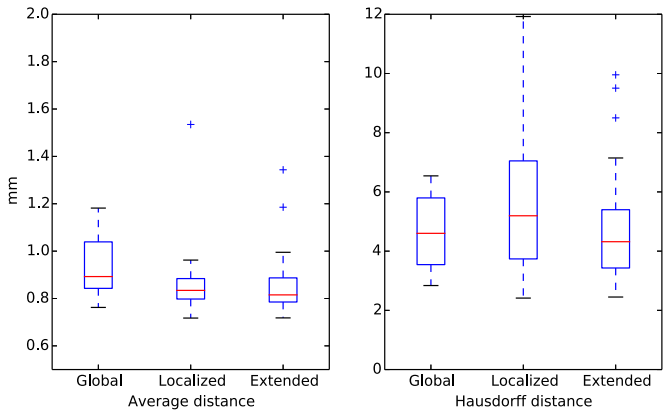


Fig. 10. Accuracy of the active shape model fitting algorithm for three different models.

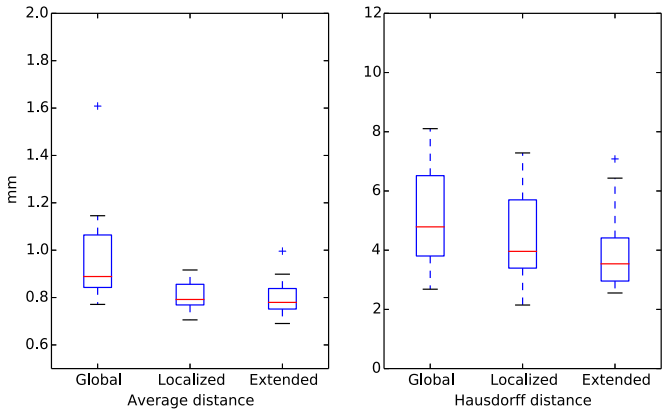


Fig. 11. Accuracy of the active shape model fitting algorithm for three different models, when 4 landmarks at the proximal and distal ends were used to make the models more robust.

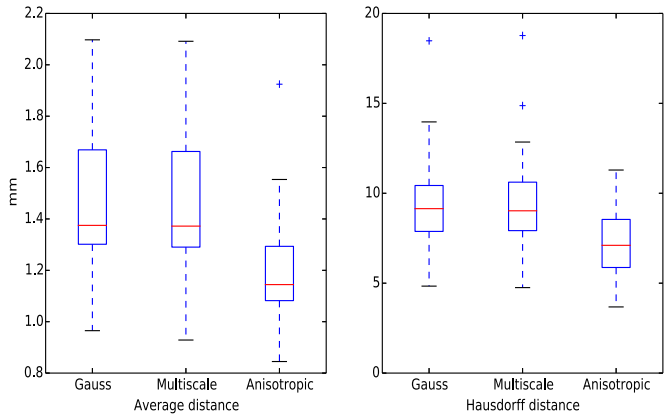


Fig. 12. Accuracy of image to image registration results performed with different models, compared on a ground-truth segmentation of the bones.

Shape Model. We can also observe that by adding flexibility, the model becomes less robust and the number of outliers (i.e., bad fitting results) increases.

We can remedy that effect by incorporating landmark constraints on the proximal and distal ends, by computing a posterior model (see Section 3.4). This has the effect of fixing the proximal and distal ends and prevents the model from moving away too far from the correct solution. Fig. 11 shows that this has the desired effect and the combination of including landmarks and increasing the model flexibility leads to clearly improved results.

5.1.5 Image to Image Registration

In the next experiment we show that our model can also be used to perform 3D image to image registration, using the full forearm CT images. We choose one image as a reference and build a GPMM on the full image domain. In the application of GPMMs to image registration, we have to be careful about the image borders, as the basis functions are global, and hence values at the boundary might strongly influence values in the interior. We therefore mask the images and optimize only on the bounding box of the bones. We use a simple mean squares metric and a stochastic gradient descent algorithm to optimize the registration functional (17). To evaluate the method, we warp the ground-truth segmentation of the forearm bones with the resulting deformation field and determine the distance between the corresponding surfaces. Fig. 12 shows the results for the same three models as used in the first experiment. In this example, where the optimization task is much more difficult, we see that the anisotropic model, which is much more targeted to the application, has clear advantages.

We also compared our method to a standard B-Spline registration method [50], which is the standard registration method used in Elastix. First, we use a B-spline that is only defined on a single scale level. As expected, since B-Splines are not application-specific, the registrations are less robust and the accuracy is worse on average (see Fig. 12). In its standard setting, Elastix uses a multi-resolution approach, where it refines the B-Spline grid, in every resolution level. This corresponds roughly to our multi-scale approach, but with the important difference that new scale levels are added for each resolution level. This strategy makes the approach much more stable and, thanks to the convenient numerical properties of B-Splines, allows for arbitrarily fine deformations. As shown in Fig. 13 in this multi-resolution setting, the B-Spline registration yields more accurate results on average than our method, but, as expected, is less robust. It is interesting to compare the two strategies in more detail. While our model has 500 parameters, the final result of the B-Spline registration has 37,926 parameters. Thanks to the convenient numerical properties of B-Splines, more could be added if to increase the model's flexibility even further. This explains why the B-Spline approach can yield more accurate solutions than GPMMs. With GPMMs, the number of parameters is limited by the number of eigenfunctions we can accurately approximate. If the image domain is large compared to the scale of the features we need to match, this quickly becomes a limitation. We refer to the supplementary material, available online, for a more detailed discussion of the approximation quality.

5.2 Face Modeling Using GPMMs

Point distribution models of the face, such as the Basel Face Model [51] or the Large Scale Facial Model [52], are of great importance in computer vision. As humans perceive even tiny errors in the registration as unnatural, these models are usually learned from manually cleaned face scans, where all the salient points are annotated with landmarks. In this experiment we show a strategy for model building using

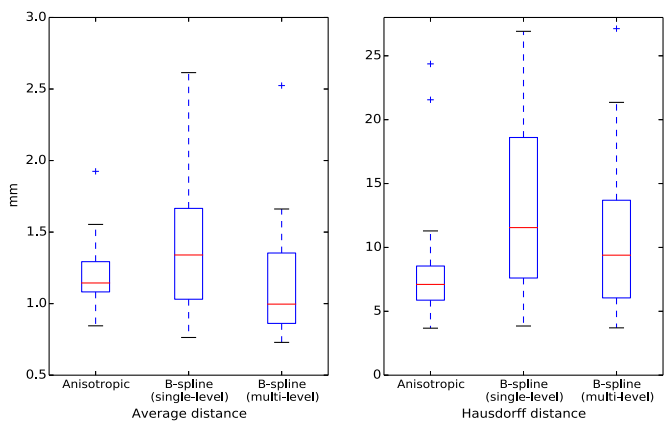
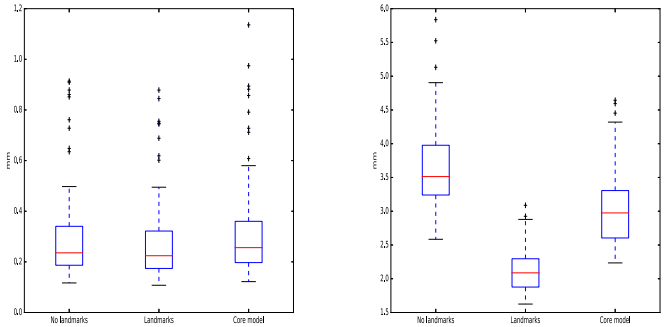


Fig. 13. Accuracy of image to image registration results performed with our best model, compared to a single-level and multi-level b-spline registration method implemented in Elastix.

(a) Reference (b) Example scan (c) Example scan

Fig. 14. (a) A reference face with annotated landmarks and a masked out region (in red). The mask region determines an area where the data quality is particularly bad, and special treatment in the registration algorithm is required. (b) and (c) Two example scans of the BU3D with annotated landmarks. Note that the number of annotated landmarks in the scans varies depending on the data quality.



(a) Average distance to ground-truth surface (b) Correspondence error on landmarks

Fig. 15. (a) shows that all three models accurately fit the target surface. (b) The correspondence, however, is improved if we use a *core model*.

GPMMs, which can make better use of the available data and requires less annotation. The idea is to build a point distribution model already once a few scans are registered. This *core model* is extended with an analytically defined prior and used in subsequent registration. Since this new prior already contains scans where the right correspondences has been enforced, we expect that it leads to better correspondence.

To test this hypothesis, we use 100 face scans from the publicly available Bu3D database [53]. The reference face and two example scans are shown in Fig. 14. The scans are manually annotated with landmarks. The number of landmarks differs, depending on the quality of the scan, but all of them have at least the lips, nose tip and eye corners annotated. We compare three different strategies for registration. In the first, we perform a basic registration using no landmarks. In the second, we use all the annotations. In the third, we use 10 of the registration results that were obtained using the landmarks, to build the core model. The basis for all three registrations is the following prior model, which is a sum of Gaussian kernels defined on different scale levels

$$k_{\mathrm{ms}}(x,x') = \sum_{i=0}^{3} \frac{30}{2^i} \exp\left(-\frac{\|x-x'\|^2}{(80/2^i)^2}\right) \mathcal{I}_{3\times 3}. \qquad (18)$$

To include the landmarks in the second strategy, a posterior model is built (Cf. Section 3.4). The extended model is built by adding to the sample covariance kernel k_{PDM}, computed from the registered surface scans, a scaled version of the multi-scale kernel sk_{ms}. The scaling factor s was determined by computing the maximal reconstruction error ϵ over all the scans and then choosing s such that $3\sqrt{sk_{\mathrm{ms}}(x,x)} = \epsilon$. The reference shape (Fig. 14a) is registered onto the target scans using the surface registration algorithm described in Section 4.1.[2]

Fig. 15 shows the average error of the registration for the three different strategies. All three strategies lead to an accurate fit with an average surface distance to the ground-truth scan of less than 0.3 mm. The quality of the correspondence however, greatly differs. Fig. 15b shows that the distance at the annotated landmarks points. As expected, the registration that explicitly includes all the landmark constraints is clearly best. The average error in this case corresponds to the modeled uncertainty for the landmarks. More interestingly, we see that including the registered example scans into the prior clearly leads to improved correspondences.

As a final experiment, we show how increasing the flexibility of the model can be used to work around limitations of already existing high quality face models, such as for example the Basel Face model (BFM) [51]. The BFM was built using mainly young people, and hence does not generalize well when registering older faces. To work around this limitation, we extend the shape variations modeled by the BFM with smooth deformations given by a Gaussian kernel with a very small scale:

$$k(x,x') = k_{\mathrm{BFM}}(x,x') + 1.0\exp\left(-\frac{\|x-x'\|^2}{30^2}\right),$$

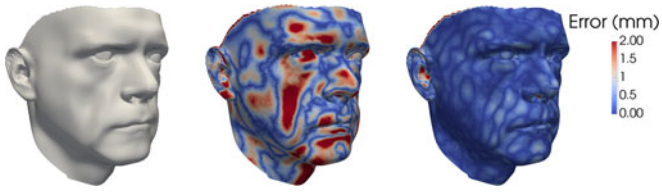
Since we choose the scale of the added deformation small, the model variation is dominated by the variation in the Basel face model and hence samples of the new model still look realistic (Fig. 16). Yet, the model is much more expressive and better generalizes to unseen faces. This is illustrated in Fig. 17 where the best reconstruction obtain with the BFM is compared to the best reconstruction obtained using the extended model.

6 Conclusion

We have presented Gaussian Process Morphable Models, a generalization of classical point distribution models. GPMMs extend standard PDMs based on object boundary

Fig. 16. Three sample from the Basel Face Model, extended with additional variability using a Gaussian kernel with small variance. Even though synthetic variation was added, the samples still correspond to anatomically valid face shapes.

2. To make the optimization robust to the missing data in the scans (Figs. 14b and 14c), we exclude points in the ear region (marked in red) and those whose closest point in a target shape is a boundary point of the target mesh.



(a) Target Face (b) Reconstruction BFM (c) Reconstruction extended BFM

Fig. 17. Best reconstruction of a target face (a) with the Basel face model (b) and the extended model (c).

positions in two ways: First GPMMs are defined by a Gaussian process, which makes them inherently continuous and do not force an early discretization. More importantly, rather than only estimating the covariances from example datasets, GPMMs can be specified using arbitrary positive definite kernels. This makes it possible to build complex shape priors, even in the case where we do not have many example dataset to learn a point distribution model. Similar to a PDM, a GPMM is a low-dimensional, parametric model. It can be brought into the exact same mathematical form as a PDM by discretizing the domain on which the model is defined. Hence our generalized shape models can be used in any algorithm that uses a standard shape model. To make our method easily accessible, we have made the full implementation available as part of the open source framework Statismo [9] and Scalismo [10].

Our experiments have confirmed that GPMMs are well suited for modeling prior shape knowledge in registration problems. As all prior assumptions about shape deformations are encoded as part of the GPMM, our approach achieves a clear separation between the modeling and optimization. This separation makes it possible to use the same numerical methods with many different priors. Furthermore, as a GPMM is generative, we can assess the validity of our prior assumptions by sampling from the model. We have shown how the same registration method can be adapted to a wide variety of different applications by simply changing the prior model. Indeed, Gaussian processes give us a very rich modeling language to define this prior, leading to registration methods that can include many types of different prior knowledge, including models learned from examples. From a practitioner's point of view, the straight-forward integration of landmarks may also be a valuable contribution, since it enables to develop efficient interactive registration schemes.

The most important assumption behind our models is that the shape variations can be well approximated using only a moderate number of leading basis functions. As shape deformations between objects of the same class are usually smooth and hence the deformations between neighboring points highly correlated, this assumption is usually satisfied. Furthermore for most anatomical shapes, fine detailed deformations only occur in parts of the shape. GPMMs give us the modeling power to model these fine deformations only where they are needed. Our method reaches its limitations when very fine deformations need to be modeled over a large domain, as it is sometimes required in image registration. In this case the approximation scheme becomes inefficient and the approximations inaccurate. An interesting extension for future work would be to devise a hierarchical, multi-resolution approach, which would partition the domain in order and perform separate approximation on smaller sub-domain. In this way, the modeling power of GPMMs could be exploited to model good priors for image registration, while still offering all the flexibility of classical image registration approaches.

We hope with this work to bridge the gap between the so far distinct world of classical shape modeling, where all the modeled shape variations are a linear combination of the training shapes, and the word of registration, where usually overly simple smoothness priors are used. We believe that it is the middle ground between these two extremes, where shape modeling can do most for helping to devise robust and practical applications.

Acknowledgments

This work has been funded as part of the Swiss National Science foundation project in the context of the project SNF153297. We thank Sandro Schönborn and Volker Roth for interesting and enlightening discussion. A special thanks goes to Ghazi Bouabene and Christoph Langguth, for their work on the Scalismo software, in which all the methods are implemented.

References

[1] U. Grenander and M. I. Miller, *Pattern Theory: From Representation to Inference*, vol. 1. Oxford, U.K.: Oxford Univ. Press, 2007.
[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
[3] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Comput. Graph. Interactive Techn.*, 1999, pp. 187–194.
[4] S. M. Pizer and J. Marron, "Chapter 6 - object statistics on curved manifolds," in *Statistical Shape and Deformation Analysis*, G. Zheng, S. Li, and G. Székely, Eds. New York, NY, USA: Academic, 2017, pp. 137–164. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128104934000079
[5] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.
[6] M. Li and J. T.-Y. Kwok, "Making large-scale Nystrom approximation possible," in *Proc. 27th Int. Conf. Mach. Learning*, 2010, p. 631.
[7] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[8] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.
[9] M. Lüthi, et al., "Statismo-a framework for PCA based statistical models," *Insight J.*, 2012.
[10] Scalismo - scalable image analysis and shape modelling. (2013). [Online]. Available: http://github.com/unibas-gravis/scalismo
[11] Y. Wang and L. H. Staib, "Boundary finding with prior shape and smoothness models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 738–743, Jul. 2000.
[12] U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quart. Appl. Mathematics*, vol. 56, no. 4, pp. 617–694, 1998.
[13] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. Amer. Statist. Assoc.*, vol. 86, no. 414, pp. 376–387, 1991.
[14] S. C. Joshi, et al., "Gaussian random fields on sub-manifolds for characterizing brain surfaces," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer, 1997, pp. 381–386.
[15] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: A survey," *Comput. Vis. Image Understanding*, vol. 134, pp. 1–21, 2015.

[16] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.
[17] A. Tagliasacchi, S. Bouaziz, M. Pauly, and H. Li, "Modern techniques and applications for real-time non-rigid registration," in *Proc. SIGGRAPH Asia Courses*, 2016, Art. no. 11.
[18] M. Galun, T. Amir, T. Hassner, R. Basri, and Y. Lipman, "Wide baseline stereo matching with convex bounded distortion constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2228–2236.
[19] Y. Lipman, S. Yagev, R. Poranne, D. W. Jacobs, and R. Basri, "Feature matching with bounded distortion," *ACM Trans. Graph.*, vol. 33, no. 3, 2014, Art. no. 26.
[20] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Trans. Med. Imag.*, vol. 27, no. 1, pp. 111–128, Jan. 2008.
[21] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Berlin, Germany: Springer, 2006.
[22] J. Zhu, S. C. Hoi, and M. R. Lyu, "Nonrigid shape recovery by gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1319–1326.
[23] B. Schölkopf, F. Steinke, and V. Blanz, "Object correspondence as a machine learning problem," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 776–783.
[24] L. Younes, *Shapes and Diffeomorphisms*, vol. 171. Berlin, Germany: Springer, 2010.
[25] M. Bruveris, L. Risser, and F.-X. Vialard, "Mixture of kernels and iterated semidirect product of diffeomorphisms groups," *Multiscale Model. Simul.*, vol. 10, no. 4, pp. 1344–1368, 2012.
[26] S. Sommer, F. Lauze, M. Nielsen, and X. Pennec, "Sparse multiscale diffeomorphic registration: The kernel bundle framework," *J. Math. Imag. Vis.*, vol. 46, no. 3, pp. 292–308, 2013.
[27] T. Schmah, L. Risser, and F.-X. Vialard, "Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation," in *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer, 2013, pp. 203–210.
[28] T. F. Cootes and C. J. Taylor, "Combining point distribution models with shape models based on finite element analysis," *Image Vis. Comput.*, vol. 13, no. 5, pp. 403–409, 1995.
[29] Z. Zhao, S. R. Aylward, and E. K. Teoh, "A novel 3D partitioned active shape model for segmentation of brain MR images," in *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer, 2005, pp. 221–228.
[30] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 414–423, Mar. 2003.
[31] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum, "Multiscale 3-D shape representation and segmentation using spherical wavelets," *IEEE Trans. Med. Imag.*, vol. 26, no. 4, pp. 598–618, Apr. 2007.
[32] T. Albrecht, M. Lüthi, and T. Vetter, "A statistical deformation prior for non-rigid image and shape registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[33] D. Kainmueller, H. Lamecker, M. O. Heller, B. Weber, H.-C. Hege, and S. Zachow, "Omnidirectional displacements for deformable surfaces," *Med. Image Anal.*, vol. 17, no. 4, pp. 429–441, 2013.
[34] Y. Le, U. Kurkure, and I. Kakadiaris, "PDM-ENLOR: Learning ensemble of local PDM-based regressions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1878–1885.
[35] M. Lüthi, C. Jud, and T. Vetter, "Using landmarks as a deformation prior for hybrid image registration," in *Proc. Joint Pattern Recognit. Symp.*, 2011, pp. 196–205.
[36] T. Gerig, K. Shahim, M. Reyes, T. Vetter, and M. Lüthi, "Spatially varying registration using Gaussian processes," in *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer, 2014, pp. 413–420.
[37] M. Lüthi, C. Jud, and T. Vetter, "A unified approach to shape model fitting and non-rigid registration," in *Machine Learning in Medical Imaging*. Berlin, Germany: Springer, 2013, pp. 66–73.
[38] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.
[39] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, vol. 3. Berlin, Germany: Springer, 2004.
[40] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/090771806
[41] D. Duvenaud, "Automatic model construction with Gaussian processes," Ph.D. dissertation, Pembroke College, Univ. Cambridge, Cambridge, U.K., 2014.
[42] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 296–301.
[43] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms, "Elastic body splines: A physics based approach to coordinate transformation in medical image matching," in *Proc. 8th IEEE Symp. Comput.-Based Med. Syst.*, 1995, pp. 81–88.
[44] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1427–1442, Nov. 2003.
[45] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 526–534, Jun. 2001.
[46] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[47] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.
[48] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 376–380, Apr. 1991.
[49] M. A. Styner, et al., "Evaluation of 3D correspondence methods for model building," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer, 2003, pp. 63–75.
[50] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3D statistical deformation models using non-rigid registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2001, pp. 77–84.
[51] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 296–301.
[52] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3D morphable model learnt from 10,000 faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5543–5552.
[53] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 211–216.

Marcel Lüthi received the bachelor's degree in computer science from the University of Applied Science Bern, Switzerland, the master's degree in engineering mathematics from the Chalmers University of Technology, Sweden, and the PhD degree from the University of Basel, in 2010. He is a researcher with the University of Basel. His research interests include statistical shape modeling and its application to medical image analysis.

Thomas Gerig received the master's degree in computer science from the Department of Mathematics and Computer Science, University of Basel, Switzerland, in 2013. He is working toward the PhD degree in the Department of Mathematics and Computer Science, University of Basel, Switzerland. He is interested in shape modeling, image analysis, algorithmic implementations, and numerical methods.

Christoph Jud received the BE degree in information technology from the Zurich University of Applied Sciences, in 2007, and the master's and PhD degrees from the Department of Mathematics and Computer Science, University of Basel, Switzerland. He is a postdoctoral researcher in the Medical Image Analysis Group, University Hospital Basel. His research interests include medical image analysis, statistical shape models, and continuous optimization.

Thomas Vetter studied mathematics and physics and received the PhD degree in biophysics from the University of Ulm, Germany. As a postdoctoral researcher at the Center for Biological and Computational Learning, Massachusetts Institute of Technology, Cambridge, he started his research on computer vision. In 1993, he moved to the Max-Planck-Institute, Tübingen, Germany, and, in 1999, he became a professor for computer graphics with the University of Freiburg, Germany. Since 2002, he has been a professor of applied computer science with the University of Basel, Switzerland. His current research is on image understanding. He combines methods from machine learning, computer graphics and computer vision to implement analysis-by-synthesis systems for an automated image perception.