A Probabilistic Surface Registration Framework with Applications to Partial Data Analysis

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Dennis Madsen

Basel, 2021

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Thomas Vetter, Universität Basel, Erstbetreuer
Prof. Dr. Volker Roth, Universität Basel, Zweitbetreuer
Prof. Dr. Tinashe Mutsvangwa, University of Cape Town, Externer Experte

Basel, den 22.06.2021

Prof. Dr. Marcel Mayor, Dekan

“Probability is expectation founded upon partial knowledge. A perfect acquaintance with all the circumstances affecting the occurrence of an event would change expectation into certainty, and leave neither room nor demand for a theory of probabilities."

— Boole, George, *An Investigation of the Law of Thought*

Preface

In front of you lays my dissertation - the cumulative work of the past four years. Before commencing on the PhD journey, I took part in a course about statistical shape modeling at the University of Basel. While the course was truly inspiring, it left me with many open questions. So I tried my utmost to answer these during my master thesis project. Even though a master's project is a journey by itself, many questions still remained unanswered. While I might not have found all the answers I was searching for, I am now, more than ever, aware of what I do not know and which of my questions are still unanswered in the research area of statistical shape models.

Even though the research topic pretty much speaks for itself, the people working in the area are truly inspiring. My greatest gratitude goes to the Scalismo evangelist Marcel Lüthi. I truly appreciate your patience, all our wonderful discussions and your immense insight. If it had not been for you, I would have never initiated this journey. Furthermore, during my master thesis and the first years of my PhD, I had the pleasure to enjoy coffee break talks and idea exchange with many people from the lab. Notable mentions from the GraVis group that made my time memorable: Sandro Schönborn, Ghazi Bouabene, Thomas Gerig, Bernhard Egger, Adam Kortylewski, Andreas Schneider and Patrick Kahr.

Special thanks go to: Dana Rahbani for her kindness, infectious optimism and for her idea to the ginger-sounding acronym. Jonathan Aellen for his curious nature and special eye for detail. Andreas Morel-Forster for always finding time to listen to my crazy ideas and co-authoring some of my papers.

I am grateful to all the people that helped me with the final proofreading:

Andreas Morel-Forster, Dana Rahbani, Jonathan Aellen, Rufus Lobo, Patrick Buder and Đorđe Relić.

Doing a PhD requires resilience to recover from the many dead-ends and dark places one is facing. I have though been fortunate to have the most supportive wife and daughter around. They have given me space when needed, kept me timely on track, always been there with their immense love and reminded me that there are more important things in life than work. I love you both.

Last but not least, I want to thank Prof. Thomas Vetter for his guidance, always insightful remarks and for trusting and believing in me. And so, this thesis, "das ist meine Kür."

Abstract

The usage of Point Distribution Models (PDMs) is very attractive, especially in medical imaging analysis. This is due to their simplicity, robustness, and the relatively low number of data items needed for their construction. When a PDM is used to analyze a novel image one can automatically extract measures from it, the likelihood of the target shape can be estimated or it can be used to complete partial observations. The most important aspect of constructing PDMs and analyzing novel data is how to establish point-to-point correspondence. Previous work mainly relies on the underlying assumption that exactly one optimal correspondence exists. Particularly when partial data is completed, the established correspondence has a big influence and can result in incorrect reconstructions.

In this thesis, we introduce a generalized probabilistic non-rigid registration framework. We build upon the Gaussian Process Morphable Model (GPMM) framework, which clearly separates prior deformation information from the general registration algorithm. We view non-rigid surface registration as a probabilistic inference problem. This leads us to a distribution of registrations from a specific target. More importantly, the distribution does not rely on any hard correspondence assumption. Instead of settling on one specific correspondence assumption, our method takes all the different correspondence pairs into account.

The thesis is divided into 3 main parts: first we introduce GiNGR, which is a generalized iterative non-rigid surface registration framework that unifies previous algorithms such as Iterative Closest Point (ICP) and Coherent Point Drift (CPD) under the same non-rigid registration framework. We then show

how we can use GiNGR to make existing registration algorithms probabilistic in comparison to their deterministic nature. We empirically show how a simple ICP method benefits from the probabilistic framework by becoming more robust and able to quantify the registration uncertainty. Finally, we show several applications of the probabilistic registration framework by analyzing partial data and how PDMs can be constructed from partial data with different properties using multiple imputations.

Table of Contents

Acronyms

BFM Basel Face Model

BMI Body Mass Index

BCPD Bayesian Coherent Point Drift

CBCT Cone-beam computed tomography

CP Correspondence-Proposal

CPD Coherent Point Drift

CT Computerized Tomography

EM Expectation–maximization

GiNGR Generalized Iterative Non-Rigid Point Cloud and Surface Registration Using Gaussian Process Regression

GP Gaussian Process

GPMM Gaussian Process Morphable Model

GPR Gaussian Process Regression

ICP Iterative Closest Point

MAP Maximum a posteriori probability

MCMC Markov Chain Monte Carlo

MH Metropolis-Hastings

MOOC Massive Open Online Course

MRI Magnetic Resonance Imaging

PCA Principal Component Analysis

pdf probability density function

PDM Point Distribution Model

P-GiNGR Probabilistic GiNGR

SLAM Simultaneous Localization and Mapping

SSM Statistical Shape Model

VBI Variational Bayesian Inference

1 Introduction

The goal of computer vision is to automate basic tasks that the human visual system can do. For photographs of everyday scenes and objects, it is trivial for humans to classify most objects in an image. We can point out where the objects in an image are located and even partition the photo into multiple segments to clearly separate the objects and the background. However, for medical images, a similar task often requires extensive training of experts. This training can be very time-consuming and is prone to a large variability in the outcome, depending on the task at hand. By utilizing computer vision systems within medical image analysis, we can make systems that are robust and scalable.

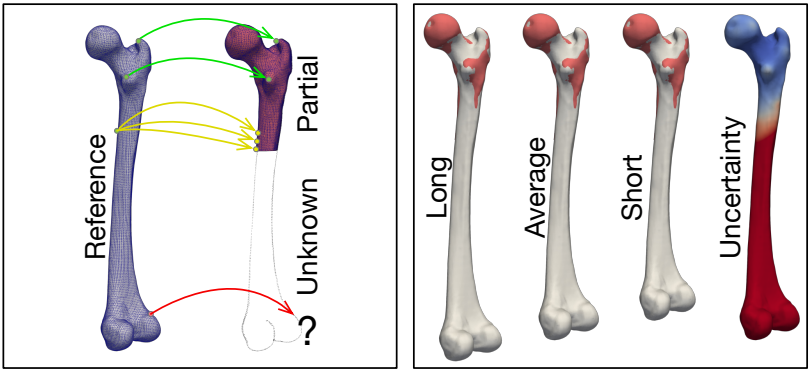
A Statistical Shape Model (SSM) provides an intuitive way for both surgeons and patients to assess shape data. An SSM is a generative model that describes the shape variability within a shape family. A target image or a surface can be analyzed with an SSM through analysis-by-synthesis, i.e. by utilizing the SSM to generate an image or a surface that represents the target as close as possible. An SSM can help to automatically quantify how likely a given shape is. It can also be used to automatically extract measures, such as target volume or distances between anatomical landmarks. An SSM is often used to reconstruct the complete target shape from partial observations [5, 102, 104], which

is useful for forensic investigation, reconstructive surgeries or patient-specific implant design [74, 91]. Another popular application of SSMs is image segmentation, for which an overview is given in [33]. In this work, we focus on Point Distribution Models (PDMs) which are a type of SSMs. PDMs provide inherent correspondence which we consider to be an important property for a lot of automatic analyses beyond segmentation. Other models without a point-correspondence assumption are e.g. SSMs based on level sets [92] or non-parametric shape priors [45]. The classical PDM is formulated as a distribution over shapes. It is therefore possible to derive a posterior model if only part of a target is given [1] or if information such as weight, sex, or the age of a patient is known [7].

In this thesis, we focus on point-to-point correspondence which is the underlying principle in PDMs. Point-to-point correspondence, point correspondence, or simply correspondence are often used interchangeably. In this thesis we will mainly be using point-correspondence. To construct a PDM we need a dataset of point-sets in point-correspondence. Likewise, when using a PDM to analyze a target we also need to establish point-correspondence between the PDM and the target. Establishing point-correspondence between two shapes means we find point identifiers on the two shapes that have the same anatomical meaning. For instance, when analyzing faces this could the tip of the nose on the two faces or when analyzing hands the tip of the thumb. In [16] when PDMs were introduced, the corresponding points were found by manually labeling a sparse set of points, which meant that every point had an anatomical meaning. In more recent times, correspondence is found densely over surfaces and varies from a few thousand to millions of points. The initial usage of sparsely manually annotated anatomical landmarks has, in other words, shifted to densely located landmarks which also involve points in between anatomical points, known as mathematical landmarks. In order to densely establish point-correspondence between a reference (sometimes referred to as template or source) and a target point-set we rely on automated processes known as point-set registration. The process of registering two point-sets is also known as pairwise registration [103].

So far we have been discussing shapes as surfaces and as point-sets. In the literature, registration methods are usually published separately for special surface objects or for point-sets. With Gaussian Process Morphable Models (GPMMs) introduced in [52] we can unify these two branches of research. A PDM is a distribution over shapes, whereas a GPMM is a distribution over deformations from a reference surface. The reference can either be a continuous surface or it can be a point-set discretization of a surface. For the remainder of the thesis we will mainly be focusing on surfaces with point-sets being a special instance under the GPMM formulation. On a technical note, we mainly use triangulated meshes to represent surfaces. The move away from a sparse set of landmarks with a clear anatomical meaning has the consequence that identifying the same landmark identifier among two or more surfaces can be difficult, especially in large smooth areas. The problem becomes even more prominent when registering partial data as shown in Fig. 1.1a. As the mathematical landmarks are spanned between anatomical landmarks, it is impossible to know which part of the complete femur shaft corresponds to the partial femur shaft. In such a setting, with traditional non-rigid registration methods, we can only obtain a single *best* correspondence which is highly dependent on the initial starting point of the reference. We will refer to hard correspondence whenever a method is used to find an exact one-to-one mapping between surfaces. Notice how this problem can be problematic when analyzing partial data as well as when wanting to construct a PDM from partial data. An example of using hard correspondence is shown in Fig. 1.1b where a partial femur is completed with a PDM using differently obtained correspondences. In the case of the femur, the correspondence has a large influence on the predicted length of the femur. The length may vary up to several centimeters, depending on the obtained correspondence (even while having the same registration accuracy).

The hard correspondence problem does not mean that we should avoid using PDMs. Despite their inherent flaws, PDMs have shown to be valuable in many applications. But it does mean that we should be aware of their weaknesses and underlying assumptions and find methods that avoid using the hard



(a) (b)

Figure 1.1: Hard correspondence problem. In (a) we show how it can be especially difficult to obtain correspondence to a partial object in areas where the surface is smooth and anatomical meaningful landmarks are therefore difficult to identify. The green arrows show landmarks that can uniquely be identified, the yellow arrows show landmarks that cannot be uniquely identified as they are located on a smooth surface and the red arrow shows a landmark that is not visible on the partial surface. The unknown outline shows a possible completion of the partial observation. In (b) we show different completions of the partial femur surface by using differently estimated correspondences. We see how the same partial surface might be reconstructed to having very different lengths, depending on the obtained correspondence. We also show the correspondence uncertainty which we can obtain using probabilistic registration (blue: small uncertainty, red: large uncertainty).

correspondence property.

1.1 Contribution and Overview

In this thesis, we revisit point-correspondence, which is one of the fundamental principles behind PDMs. Our main contribution is the formulation of a generalized probabilistic registration framework that makes use of a soft correspondence property. Instead of finding a single best registration, we instead compute a full posterior distribution of possible registrations from a single tar-

get. In Fig. 1.2 an overview of the thesis is given. The thesis is structured in a bottom-up fashion. In Chapter 2, we introduce Gaussian Process (GP), Gaussian Process Regression (GPR) and GPMMs which are fundamental building blocks in our registration framework. In Chapter 3 we introduce Generalized Iterative Non-Rigid Point Cloud and Surface Registration Using Gaussian Process Regression (GiNGR), which is a generalization of existing non-rigid registration algorithms such as Iterative Closest Point (ICP) and Coherent Point Drift (CPD). The reformulation of many existing algorithms into the modular GiNGR framework has more advantages than just making the hyper-parameters interpretable and comparable. This being, multi-resolution fitting to speed up registration of large point-sets, controllable efficiency through the GPMMs kernel approximation, a clear strategy to include expert annotations and the possibility to combine analytical and statistical learned deformations. In Chapter 4 we introduce the generalized probabilistic registration framework Probabilistic GiNGR (P-GiNGR), which is a fully probabilistic method for non-rigid registration, that makes use of the Metropolis-Hastings (MH) algorithm. Furthermore, we show how GiNGR is just a special instance of P-GiNGR which means that by reformulating existing algorithms into GiNGR we can have them utilize the probabilistic registration properties of P-GiNGR

Finally, in Chapter 5 we show different applications of P-GiNGR. We show how P-GiNGR can be used to analyze partial data and how PDMs can be constructed from a variety of partial data observations.

1.1.1 A Word of Caution

In contrast to the traditional mentioning of PDMs, we do not scale all example data to the same size, as we usually want to quantify the size of a shape. This is the notion being used within the GPMM community where a shape is defined as being invariant under some transformation. In GPMM this transformation is purely rigid, whereas traditional shape analysis often is invariant under a similarity transformation.

In this work, we focus on non-rigid registration between a single reference and a target. The reference and the target can be either point-sets or surfaces.

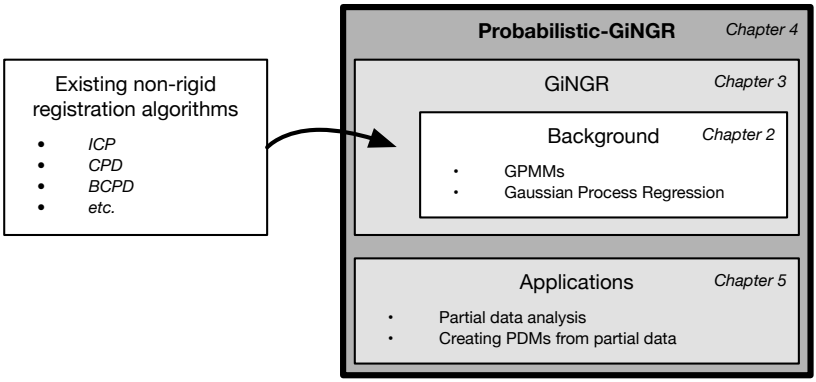


Figure 1.2: Overview of the thesis structure and how the different chapters are connected.

Unless otherwise specified, registration refers to finding a non-rigid transformation to deform the reference onto the target. Throughout the thesis, we will be referring to their non-rigid versions when mentioning ICP and CPD unless otherwise is specified.

1.2 Related work

Mentioning that no single best registration exists for modern 3D point-sets with millions of vertices has previously been done in [40]. The authors suggest to overcome the hard correspondence problem by computing correspondence probabilities between points on the reference and the target. This procedure is similar to the method used in CPD to obtain point-correspondence which we will look more into in Section 3.2.1. With the correspondence probabilities in [40], they compute the mean shape based on a set of correspondence probabilities instead of a set of shapes in correspondence. In order to perform Principal Component Analysis (PCA) and obtain the variation in the dataset, they use *virtual correspondence points* based on the correspondence probabilities. In Section 5.4 we show how this can be done directly from a distribution of different correspondence pairs.

Additional related work is found in the individual chapters. In Section 3.3 we discuss existing non-rigid registration methods. In Section 4.5 we discuss related work in probabilistic registration methods as well as methods for quantifying uncertainty in registration methods. In the introduction for the individual applications in Chapter 5 we have added related work to the specific applications.

2
Background

In this chapter we formalize the non-rigid registration task and introduce both GPR and the concept of GPMMs.

2.1 Non-rigid Registration

The goal of registration is to find an optimal spatial transformation of a reference surface $\Gamma_R \subseteq \Omega$, onto a target surface $\Gamma_T \subset \mathbb{R}^d$ with $\Omega \subset \mathbb{R}^d$ being the domain and d the domain dimension, which is typically $d = 2$ or $d = 3$. Formally we can write

$$\arg\min_{\mathcal{T}} \mathbb{d}\left[\mathcal{T}(\Gamma_R), \Gamma_T\right]. \tag{2.1}$$

The output of registration is then the optimal transformation $\mathcal{T}$ according to the distance function $\mathbb{d}(\cdot,\cdot)$. The deformations that are generally involved with surface and point-set registration are, scaling $\mathbb{s} \in \mathbb{R}$, rotation $R \in SO(d)$, translation $\boldsymbol{t} \in \mathbb{R}^d$ and non-rigid deformations $u : \Omega \to \mathbb{R}^d$. We can then formulate the registration problem as

$$\tilde{\mathcal{T}} = \arg\min_{\mathbb{s},R,\boldsymbol{t},u} \{\mathbb{d}\left[\mathbb{s}R(x + u(x)) + \boldsymbol{t}, \Gamma_T\right] + \lambda\mathcal{R}(u) | x \in \Gamma_R\}, \tag{2.2}$$

with $\tilde{\mathcal{T}}$ being the optimal transformation which is regularized by $\mathcal{R}$ to restrict the non-rigid deformations to be smooth and λ a hyper-parameter to control the amount of regularization.

For the majority of this thesis, we will focus on settings where the scaling is constantly one and the possible global rigid difference between the surfaces have been minimized by solving a least square problem based on a few landmarks on each of the surfaces as in [94]. The transformations that we are interested in are therefore u, which are smooth local deformations, that should let the shape stay within its shape class. The regularized local deformations are then found with

$$\tilde{u} = \arg\min_{u} \{\mathbb{d}\left[x + u(x), \Gamma_T\right] + \lambda \mathcal{R}(u) | x \in \Gamma_R\}. \tag{2.3}$$

To model the local deformations u, we make use of GPMMs.

2.2 Gaussian Process Morphable Models

We model the surface deformations with a GP model, which serves as a continuous function $u \sim \mathcal{GP}(\mu, k)$ on the domain. The GP is completely defined by its mean function $\mu : \Omega \to \mathbb{R}^d$ and a kernel function $k : \Omega \times \Omega \to \mathbb{R}^{d \times d}$. In the book by Rasmussen and Williams [76] an extensive overview of the area of GPs is given. Lüthi et al. [52] introduced GPMMs where GPs are used to describe deformations from a reference surface. Their formalization is a generalization of the classical PDM [16], which in contrast to PDMs also allows for analytically defined deformations by specifying the mean and the kernel function of the GP. The GP is defined such that $\mu : \Gamma_R \to \mathbb{R}^d$ and $k : \Gamma_R \times \Gamma_R \to \mathbb{R}^{d \times d}$. The deformations that the reference shape can undergo are distributed according to the defined GP, $u \sim \mathcal{GP}(\mu, k)$ with a surface defined such that

$$\Gamma = \{x + u(x) | x \in \Gamma_R\} \tag{2.4}$$

Using a GP as the model leads to a possible inifinite-dimensional non-parametric model. With the introduction of GPMMs, they propose to use the truncated Karhunen-Loève expansion to obtain a low-rank approximation of

the GP. With this, the GP is approximated as

$$u[\boldsymbol{\alpha}](x) = \mu(x) + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\phi_i(x), \alpha_i \sim \mathcal{N}(0,1) \tag{2.5}$$

where r is the number of basis functions used in the approximation and λ_i and ϕ_i are the i-th eigenvalue and eigenfunction of the covariance operator associated with the kernel function k. Consequently, any deformation u is uniquely determined by a coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^r$ with which we can specify the resulting surface as

$$\Gamma[\boldsymbol{\alpha}] = \{x + \mu(x) + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\phi_i(x) | x \in \Gamma_R\}. \tag{2.6}$$

We are able to reduce the computation costs of fitting by choosing a small r. As the $\boldsymbol{\alpha}$ vector contains independent normally distributed values, we also have an associated probability for each deformation

$$p(\Gamma[\boldsymbol{\alpha}]) = p(\boldsymbol{\alpha}) = (2\pi)^{-\frac{r}{2}} \exp(-\frac{\left\|\boldsymbol{\alpha}^2\right\|}{2}). \tag{2.7}$$

Different choices of kernel functions lead to different well-known deformation models, such as radial basis functions, b-splines or thin-plate splines. In order to model smooth deformations, we most commonly choose a zero-mean GP with a Gaussian kernel

$$g(x, x') = \exp\left(\frac{-||x - x'||^2}{2\beta^2}\right), \tag{2.8}$$

with β being the width of the Gaussian kernel and the full kernel function in d dimension is

$$k(x, x') = I_d \cdot g(x, x')s \tag{2.9}$$

where I_d is the d dimensional identity matrix and s is a kernel scaling parameter.

2.2.1 Point Distribution Models

Besides being able to analytically define the kernel function, the deformation prior can also be learned from a set of N example shapes in correspon-

dence. We can calculate the deformations u from a reference shape at each corresponding point x and model them by a GP with mean:

$$\mu_{PDM}(x) = \frac{1}{N}\sum_{i=1}^{n} u_i(x) \tag{2.10}$$

and the kernel function:

$$k_{PDM}(x, x') = \frac{1}{N-1}\sum_{i=1}^{n}(u_i(x) - \mu_{PDM}(x))(u_i(x') - \mu_{PDM}(x'))^T. \tag{2.11}$$

This makes the deformations modeled by the GP similar to the shape variance which is modeled in traditional PDMs [16] as also pointed out in [52]. Throughout the thesis, we will be referring to a GPMM with a kernel learned from examples using PCA as a PDM and a model using analytically defined kernels as a GPMM.

2.2.2 Standard Registration with GPMM

With the GPMM notation, we can rewrite Eq. (2.3) to minimize the distance between the two surfaces with the deformation field u which is approximated by the coefficient vector $\boldsymbol{\alpha}$

$$\tilde{\alpha} = \underset{\alpha_1,\ldots,\alpha_r}{\arg\min}\, \mathbb{d}\left[\Gamma[\boldsymbol{\alpha}], \Gamma_T\right] + \lambda\mathcal{R}(\boldsymbol{\alpha}), \tag{2.12}$$

with $\mathcal{R} = \sum_{i=1}^{r} \alpha_i^2$ regularizing the deformation fields. In [52] the authors define the distance function to be the mean squared distance from the deformed reference to the target surface. The optimization of Eq. (2.12) is then performed using a black-box optimization algorithm. In Chapter 3 we introduce GiNGR which instead of a black-box optimizer makes use of GPR to iteratively deform the reference surface. With GPR we can iteratively recompute and fix the correspondence assumption for each iteration. In comparison to a black-box optimizer, this works especially well for partial surfaces. As we will discuss in Section 3.3, GPR is very flexible in how the correspondence can be obtained and has an intuitive interpretation of the regularization term.

2.2.3 Analytical Posterior Models (GPR)

GPMMs make it simple and efficient to constrain a model to match known correspondences, such as user annotations or the estimated correspondence from taking the closest point. Indeed, the corresponding *analytical posterior model* is again a GP, whose parameters are known in closed form. This is also what in [76] is referred to as GPR.

Let $u \sim GP(\mu, k)$ be a GPMM and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_d)$, be the certainty of each known landmark. Landmarks l_R on the reference surface can then be matched with their corresponding landmarks l_T on the target. The set L consists of the n_{lm} reference landmarks and its expected deformation to match the target

$$L = \{(l_R^1, l_T^1 - l_R^1), \ldots, (l_R^{n_{lm}}, l_T^{n_{lm}} - l_R^{n_{lm}})\} = \{(l_R^1, \hat{u}^1), \ldots, (l_R^{n_{lm}}, \hat{u}^{n_{lm}})\}, \tag{2.13}$$

with $\hat{u}$ being subject to Gaussian noise ϵ. Using GPR, we obtain the posterior model $u_p \sim GP(\mu_p, k_p)$, which models the possible surface deformations that are consistent with the given landmarks. Its mean and kernel function are given by:

$$\begin{aligned} \mu_p(x) &= \mu(x) + K_X(x)^T (K_{XX} + \epsilon)^{-1} \hat{U} \\ k_p(x, x') &= k(x, x') + K_X(x)^T (K_{XX} + \epsilon)^{-1} K_X(x'). \end{aligned} \tag{2.14}$$

Here we defined $K_X(x) = \left(k(l_R^i, x)\right)_{i=1,\ldots,n_{lm}}$, a vector of the target deformation as $\hat{u}^i = \left(l_T^i - l_R^i\right)_{i=1,\ldots,n_{lm}}$ and the kernel matrix $K_{XX} = \left(k(l_R^i, l_R^j)\right)_{i,j=1,\ldots,n_{lm}}$. Notice that the posterior model u_p is again a GP. Also notice how GPR is ideal for multi-resolution fitting as we do not need to observe the deformation field for the entire surface, but the areas in between landmark observations are instead regressed. Classical non-rigid surface registration algorithms usually define a deformation for a discretized set of points on the surface, which is similar to performing the registration between point-sets.

2.2.4 Discrete Surface Representations

Here we show how our representation of continuous surfaces can be used on point-sets by discretizing the surfaces to a set of points. This notation will be useful in Chapter 3, where we convert existing non-rigid registration methods into GiNGR which makes use of GPMMs and GPR. Many existing methods are described using matrix notation, so for convenience, we will reintroduce the GPR notation from above for point-sets. We sample the reference surface Γ_R at n points to construct the matrix $X_R \in \mathbb{R}^{n\times d}$. Likewise, the target surface Γ_T is sampled at m points $X_T \in \mathbb{R}^{m\times d}$. What we are looking for is a deformation vector for each point in X_R, stored in the matrix $U \in \mathbb{R}^{n\times d}$, such that $\tilde{U} = \arg\min_U \mathbb{d}\left[\Gamma_R + U, \Gamma_T\right] + \lambda R(U)$. We denote regularized deformations as $\tilde{U}$ while we write $\hat{U}$ for the observed deformations.

For the observed deformations $\hat{U}$, and the predicted deformations $\tilde{U}$ the joint distribution according to the prior GP can then be written as

$$\begin{bmatrix}\hat{U}\\ \tilde{U}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathrm{M}_X\\ \mathrm{M}_{X*}\end{bmatrix}, \begin{bmatrix}K(X,X)+\sigma^2 I_n & K(X,X_*))\\ K(X_*,X) & K(X_*,X_*)\end{bmatrix}\right) \tag{2.15}$$

Given n observed and n_* predicted deformations, $K(X,X)$ is the $n\times n$ covariance matrix of the observed deformations, $K(X,X_*)$ is the $n\times n_*$ matrix of covariances evaluated at all pairs of observed and predicted deformations and similarly for $K(X_*,X_*)$ and $K(X,X_*)$. For every observed deformation, we assume independent identically distributed Gaussian noise. The mean of the posterior distribution can be computed in closed form as:

$$\tilde{\mathrm{M}} = \mathrm{M}_{X*} + K(X_*,X)(K(X,X)+\sigma^2 I_n)^{-1}\hat{U}. \tag{2.16}$$

If the observed and predicted deformations are the same, which is the case when we estimate a correspondence for every point in registration algorithms, then the mean of GPR simplifies to

$$\tilde{\mathrm{M}} = \mathrm{M}_X + K(K+\sigma^2 I_n)^{-1}\hat{U}, \tag{2.17}$$

with K being a shorthand notation for the covariance matrix spanned by the observed deformations $K(X,X)$.

2.3 Model Parameters

A reference surface which is deformed with a GPMM can be completely described by the shape coefficient vector $\boldsymbol{\alpha}$. We use α_i to refer to the i'th entry in the $\boldsymbol{\alpha}$ vector. Likewise, we use $\boldsymbol{\alpha}^i$ to refer to the value of $\boldsymbol{\alpha}$ in the i'th iteration when performing iterative registration, where $\boldsymbol{\alpha}$ is updated in each iteration. In addition to the shape coefficients $\boldsymbol{\alpha}$ of GPMMs, we can also apply a global transformation. Here we just note the transformations for 3 dimensions where the translation is $\boldsymbol{t} = (t_x, t_y, t_z)^T \in \mathbb{R}^3$ and the rotation $R(\phi, \psi, \rho) \in SO(3)$. The rotation matrix R is parameterized by the Euler angles ϕ, ψ, ρ. We summarize all the parameters in the vector $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_n, \phi, \psi, \rho, t_x, t_y, t_z)$. Consequently, we write $\Gamma[\boldsymbol{\theta}]$ to refer to the resulting surface.

3

GiNGR: A Generalized Registration Framework

In this chapter we introduce GiNGR which is a generalization of existing non-rigid registration algorithms. GiNGR unifies methods from point-set (sometimes referred to as point cloud) and surface registration under the same framework. We will first start out with a motivation behind GiNGR. We will then introduce GiNGR and show how existing methods can be converted into the GiNGR framework. Finally, we will discuss related work before concluding the unifying framework.

3.1 Motivation

Registration of point-sets is used in many areas within computer vision, such as medical image analysis, Simultaneous Localization and Mapping (SLAM) to analyze an agent's environment within robotics or Geoscience. Closely related is also surface or volume registration where the points have additional connectivity information such as lines, triangles, or tetrahedrons. Within the last 20+ years, a vast majority of different registration algorithms have been published, all of which try to find an optimal transformation be-

tween a reference and a target. Many of the most popular point registration methods are summarized in [62, 103]. However, they do not include surface registration methods such as [2, 3, 11, 23, 28, 49, 61, 105]. With so many different methods to use, it can be difficult to know the difference between all of them and therefore also which one to choose for a specific problem. We do not consider many of the recent state-of-the-art methods based on neural network as they often need a large dataset to learn the deformation space from and can therefore be categorized as group-wise registration. An example is [42] which is the current best-ranked method for the FAUST dataset challenge [8]. Our framework requires no training data to learn its deformations from. However, within the same framework, we can utilize statistical deformation priors learned from data and even combine them with analytically defined deformation prior.

For any non-rigid registration approach, we can summarize their three main assumptions that have to be made:

1. Regularization - How similar should the deformations of neighboring points be?

2. Correspondence - How to estimate corresponding point pairs between the reference and the target?

3. Robustness - What is the noise assumption of the observed correspondence estimate? And how is uncertain correspondence handled?

Registration methods differ because of the different solutions they provide to the three points listed above. By clearly defining these assumptions, we are able to unify the methods into the same framework. The resulting hyper-parameters required by the framework are then easy to understand and compare. Among other explanations, we think the number of registration methods is that high because they are often tuned to specific datasets and that their code is often not available. The hyper-parameters are also often difficult to interpret and thereby difficult to choose optimally. With GiNGR we have implemented a modular setup to easily switch between different registration strate-

gies. Hence we can make algorithms better comparable and easier adaptable to new tasks.

The basics of the GiNGR framework have informally been around for years. A minimal non-rigid ICP version was introduced together with GPMMs in a Massive Open Online Course (MOOC) [1]. In this chapter we generalize the steps introduced during the MOOC and show how this generalization makes it possible to reformulate many existing algorithms into the modular GiNGR framework.

The reformulation of many existing algorithms into the modular GiNGR framework has more advantages than just making the hyper-parameters better interpretable and comparable, such as multi-resolution fitting to speed up registration of large point-sets, controllable efficiency through the GPMMs kernel approximation, a clear strategy to include expert annotations and the possibility to combine analytical and statistical learned deformations.

3.2 Method

GiNGR is a generalization of existing non-rigid registration methods. We have summarized the framework in Alg. 1. We follow the iterative optimization method from [3], where the correspondence is recomputed and fixed for every iteration. A property of iteratively re-estimating the correspondence points is that the optimization can be non-monotonically decreasing, which can help avoid getting stuck in a local optima, as highlighted in [3]. With the estimated correspondence, we perform GPR and update the model parameters with the mean of the posterior distribution. Finally, the correspondence uncertainty might be updated before iterating again. In the following sections, we will show how CPD and ICP are examples of GiNGR and how they define the GPMM kernels, correspondence estimation and noise term. As a minimal setup, one can use the Gaussian kernel from Eq. (2.8) with β set to the wanted point correlation distance. The uncertainty σ^2 can be set manually based on

[1] futurelearn.com/courses/statistical-shape-modelling

Algorithm 1: GiNGR registration algorithm

Input: Γ_T target surface,
$\hat{\Gamma}_R = \Gamma_R$,
GPMM based on the reference surface $\hat{\Gamma}_R$
Output: $\hat{\Gamma}_R$
Data: set GPMM parameter $\boldsymbol{\alpha} = \mathbf{0}$,
initialize correspondence uncertainty σ^2
while *While (not converged)* **do**
Estimate correspondence deformations L (Eq. (2.13)) or $\hat{U}$ if estimating for all points in Γ_R;
Perform GPR, i.e. compute u_p (Eq. (2.14));
Update $\boldsymbol{\alpha}$ with the mean of the posterior distribution (Eq. (2.14) or Eq. (2.16));
Update GPMM reference $\hat{\Gamma}_R$ *(Optional)*;
Update correspondence uncertainty σ^2;
end

the noise assumption for the estimated correspondences. A good initial value is given in [66]:

$$\sigma^2 = \frac{1}{dnm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left\| \hat{X}_R^i - X_T^j \right\|^2 \tag{3.1}$$

with $\hat{X}_R^i - X_T^j$ being the distance between the $i'th$ point on the reference and the $j'th$ point on the target. The correspondence deformation $\hat{U}$ can be computed using closest point estimate. The reference $\hat{\Gamma}_R$ is then updated with the mean of the GPR from Eq. (2.16). This step should be skipped if using a statistical kernel (Section 3.2.2.1). Instead, the model parameters of the GPMM should be updated by projecting the posterior mean into the model. The uncertainty value is either manually or automatically decreased in each iteration as the reference gets closer to the target. An automatic update strategy is seen in [67] where they use gradual annealing to decrease the uncertainty in each step, or [35, 66] where the updated uncertainty is computed based on the variance difference between the reference and target points and the deformation update.

Seen from a practical aspect when registering two surfaces using the

GiNGR implementation, one starts out with defining the deformation prior for the specific application using a mixture of kernels. The GPMM is then constructed for a compact description of the model. As shown in [52], we can sample and visualize the deformation prior separately from the fitting. In the fitting step, the correspondence estimation function needs to be chosen (e.g. probabilistic or heuristic closest points) and a strategy to update the uncertainty in each iteration.

3.2.1 Unifying ICP and CPD with GiNGR

In this section, we show how existing methods are special instances under the GiNGR framework. The notation from the cited papers has been modified slightly to allow for an easier connection between the different algorithms. We also provide links to the corresponding equations in the original papers for an easy comparison.

3.2.1.1 Coherent Point Drift

CPD was originally introduced as a non-rigid point-set registration method in [67]. Later on, the authors generalized the method to additionally allow for rigid and affine transformations as well as automatically compute some of the hyper-parameters in [66]. What we are searching for in CPD is the deformation vector $\tilde{U}$ from the reference point-set X_R such that a new point-set location is found in each iteration $\hat{X}_R = X_R + \tilde{U}$. Initially $\hat{X}_R = X_R$. The deformation is iteratively computed based on the new point-set location $\hat{X}$ until the distance between the reference and the target surface meets a predefined convergence threshold. The CPD algorithm uses an Expectation–maximization (EM) approach.

In the *E* step a probabilistic correspondence matrix P is computed, where each entry is defined as (eq. 6 from [66]):

$$p_{ij} = \frac{k(\hat{X}_R^i, X_T^j)}{\sum_{ii=1}^{m} k(\hat{X}_R^{ii}, X_T^j) + C(w)} \tag{3.2}$$

with k being a Gaussian kernel (Eq. (2.8)), $\hat{X}_R^i$ X_T^j being the $i'th$ and $j'th$

point on $\hat{X}_R$ and X_T respectively and $C(w)$ being an outlier distribution controlled by $0 \leq w \leq 1$ which can be seen as the percentage of outlier points that the user can set manually. Let us consider $w = 0$, then each column of the P matrix sums to 1. In other words, each entry in the matrix states the probability of the j'th point in the target point-set corresponding to the i'th point in the reference point-set. In the following, we will also be using the $P\mathbf{1}$ vector, which is the row-wise sum of the P matrix. These values give some direct insight into the confidence of the correspondence. A higher value will give more confidence to the deformation of this point from the reference.

In the *M* step, the probability matrix is kept fixed and the MAP deformation is computed. In CPD the deformation vectors are defined with $\tilde{U} = K_G W$. K_G is a Gaussian kernel matrix with entries k_{ij} according to Eq. (2.8). W is a weighted sum of the deformations between a point on the reference and all the points on the target

$$w_i = \frac{1}{\sigma^2 * \lambda} \sum_{j=1}^{m} P(x_i|y_j)(y_j - \hat{x}_i).$$

In the paper, they show how the computation of W can be restructured such that we end up with (eq. 22 from [66])

$$K_G W = K_G(K_G + \lambda\sigma^2 Q)^{-1}(QPX_T - \hat{X}_R) \tag{3.3}$$

with $Q = \mathrm{diag}(P\mathbf{1})^{-1}$ being the diagonal matrix formed using the vector $P\mathbf{1}$. By comparing this formulation to Eq. (2.17) we have that $K = K_G$, $\sigma^2 I_n = \lambda\sigma^2 Q$ and

$$\hat{U} = QPX_T - \hat{X}_R \tag{3.4}$$

To put it another way, one iteration of CPD is equivalent to one iteration of GPR where the width of the kernel of the GP is specified from the β hyper-parameter. Each deformation observation $\hat{U}^i$ is a sum of all possible correspondence pairs where each pair is scaled with their correspondence probability. We can also see that the noisy observation is scaled with the inverse of the $P\mathbf{1}$ entries. So a high $P\mathbf{1}$ value leads to a lower noise assumption on that specific observation. The λ hyper-parameter is used as a simple manual scaling of the noise observation.

3.2.1.2 Bayesian Coherent Point Drift

Hirose [35] generalized the original CPD using Variational Bayesian Inference (VBI) formulation, calling it Bayesian Coherent Point Drift (BCPD). The BCPD algorithm follows the original CPD paper in large parts. Therefore, we only mention the main differences to the original paper. The reformulation to BCPD includes an optimization scheme for non-Gaussian kernels, to which the original CPD algorithm is limited to. The reformulation also includes integration of combined optimization of the global similarity transform and the local non-rigid deformations. The work was then extended and called BCPD++, in order to speed up the BCPD algorithm using GPR to allow for down-sampling of the point-sets [34]. As we have shown in Section 3.2.1.1, CPD can be reformulated as GPR which is also the case for BCPD. Under the GiNGR framework, the two-stage BCPD++ algorithm can instead be unified under a single GP instead of having one GP for the BCPD step and one for the interpolation steps.

In the BCPD formulation, the point-sets are vectorized, so $\boldsymbol{x}_R \in \mathbb{R}^{3n}$ and $\boldsymbol{u} \in \mathbb{R}^{3n}$. The updated point-set in each iteration is:

$$\hat{\boldsymbol{x}}_R = \mathcal{T}(\boldsymbol{x}_R + \tilde{\boldsymbol{u}}) = s(I_m \otimes R)(\boldsymbol{x}_R + \tilde{\boldsymbol{u}}) + (1_m \otimes t) \tag{3.5}$$

with $\mathcal{T}$ being the similarity transformation with a scaling $\mathtt{s} \in \mathbb{R}$, rotation $R \in \mathbb{R}^{d \times d}$ and translation $\boldsymbol{t} \in \mathbb{R}^d$ and $\otimes$ being the Kronecker product. The correspondence matrix P is largely the same as in the standard CPD, but with a few improvements. The matrix is updated with an indicator function to note if a point is an outlier as well as having updated the outlier distribution for which the integral approaches zero when n becomes large in the standard CPD. We refer the reader to eq. 8 of [35].

The local deformation observations in BCPD is in comparison to Eq. (3.4):

$$\hat{U} = (\mathcal{T}^{-1}(QPX_T) - X_R) \tag{3.6}$$

which can be explained as the local deformations modeling the residual of the global similarity transformation $\mathcal{T}$. The regularized deformations are then

$$\tilde{U} = K_G W = K_G(K_G + \lambda \frac{s^2}{\sigma^2} Q)^{-1}(\mathcal{T}^{-1}(QPX_T) - X_R) \tag{3.7}$$

Finally, the updated global similarity transformation is a least-squares estimation of transformation parameters between two point-sets similar to [94]. The point-sets used to compute the transformation are the scaled target-points QPX_T and the reference points updated with the current local deformation $\boldsymbol{x}_R + \tilde{\boldsymbol{u}}$.

3.2.1.3 Iterative Closest Point - Translations

ICP methods contain a whole range of different methods, where the task is to find a transformation that aligns a reference to a target based on a preliminary set of correspondences found from closest point searching. In [6, 10] they estimate a global rigid transformation, whereas other papers focus on the non-rigid deformation [2, 3, 11, 23, 49]. In this thesis, we use the abbreviation ICP even when referring to non-rigid deformations, which sometimes is shortened as nICP. In this section we will be looking at [3] that introduces two different variants of ICP for 3-dimensional surface registration. In [3] they use a least-square optimization technique instead of a black-box energy minimization solver as in [2]. This property is more robust as the closest points are reevaluated in each step and [3] allow for outliers, which is also possible within the GiNGR framework. The ICP-T algorithm is a simple reformulation of [2], which iteratively finds the local translation deformation for each vertex in the point-set. They introduce a stiffness term to regularize how much neighboring vertices are allowed to differ in their deformations. The paper also introduces ICP-A which includes the idea from [23] to model local differences as affine transformations. According to the authors, local affine transformations is a more robust way to register partial data. In [23] they also show how local translations can end up shrinking the reference surface which in their work is avoided by using local affine transformations.

Let us start out with looking at the ICP-T algorithm which minimizes a standard quadratic cost function:

$$\arg\min_{\mathrm{x}} = \|\mathcal{A}\mathrm{x} - \mathcal{B}\|_F^2 \,, \tag{3.8}$$

with the terms

$$\arg\min_{\tilde{U}} = \left\| \begin{bmatrix} \lambda_s B \\ W I_n \end{bmatrix} \tilde{U} - \begin{bmatrix} 0 \\ W(X_c - X_R) \end{bmatrix} \right\|_F^2 \tag{3.9}$$

where λ_s is a stiffness parameter governing the regularization strength of the deformations, $W \in \mathbb{R}^{n \times n}$ is a diagonal weight matrix identifying robust correspondence points, $B \in \mathbb{R}^{r \times n}$ is the incidence matrix and r the number of edges in the reference surface. The closest point locations for each of the points in X_R are contained in $X_c \in \mathbb{R}^{n \times d}$. For simplicity in the derivation, we set $W = I_n$ and $\lambda_s = 1$. Solving equation (3.8) using least-squares leads to

$$\mathrm{x} = (\mathcal{A}^T \mathcal{A})^{-1} \mathcal{A}^T \mathcal{B} \tag{3.10}$$

Using the values from Eq. (3.9)

$$\tilde{U} = (B^T B + \lambda_s^2 I)^{-1} (X_c - X_R) \tag{3.11}$$

Here we recognize that $L = B^T B$ is the Laplacian matrix. We now make use of the Woodbury matrix identity for invertible matrices [99]:

$$(\mathcal{A} + \mathcal{U}\mathcal{C}\mathcal{V})^{-1} = \mathcal{A}^{-1} - \mathcal{A}^{-1}\mathcal{U}(\mathcal{C}^{-1} + \mathcal{V}\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}\mathcal{A}^{-1} \tag{3.12}$$

with $\mathcal{A}^{-1} \sim L^{\dagger} = K$, all $\mathcal{U}$,$\mathcal{C}$, and $\mathcal{V}$ chosen as identity matrices as well as $\hat{U} = X_c - X_R$, we can show that ICP-T chooses the deformations as

$$\tilde{U} = Z + K(K + \lambda_s^2 I)^{-1} \hat{U}. \tag{3.13}$$

The Laplacian matrix is not invertible as one of its eigenvalues is 0. Instead, the generalized inverse (pseudo-inverse) is used which is why a correction term Z is added. The general inverse of the Laplacian matrix and its applications are summarized in [31]. The correction term of the generalized inverse Laplacian is discussed in more detail in Section 3.2.1.5.

More details about the intermediate steps of the derivation can be found in Appendix A.1. In [3], a weighting matrix is also part of the optimization. This matrix identifies non-robust points and keeps them out of the cost-function.

In GPR, non-robust points are directly removed from the observation data $\hat{U}$. In Appendix A.1.1 we include the derivation showing that this is equivalent to what is done in the ICP methods. In [3], they also show how expert annotated landmarks can be integrated in the optimization and controlled via an extra hyper-parameter. Including expert annotations in GPR is however trivial as they can be added as noisy observations to the GPR formulation. In Appendix A.1.1 we also show that λ_s from Eq. (3.9) is equivalent to σ from Eq. (2.17) which is the independent Gaussian noise assumption for each observation.

3.2.1.4 Iterative Closest Point - Affine transformations

We now turn to the ICP-A algorithm, where the final transformation matrix $\mathcal{M} \in \mathbb{R}^{4n\times 3}$ no longer consists of simple deformations, but instead of an affine transformation matrix for each point in the reference.

$$\arg\min_{\mathcal{M}} = \left\| \begin{bmatrix} \lambda_s B \otimes G \\ WD \end{bmatrix} \mathcal{M} - \begin{bmatrix} 0 \\ WX_c \end{bmatrix} \right\|_F^2 \tag{3.14}$$

with G being the $\mathrm{diag}([1,1,1,\gamma]^T)$ and γ depending on the units of the data and $D \in \mathbb{R}^{4n\times 3}$ being the sparse mapping of X_R. Again, we begin with the ICP-A energy optimization terms expanded to the least square solution. For simplicity, we have set $W = I$ and $\lambda_s = 1$ as in the ICP-T example.

$$\mathcal{M} = (B^T B \otimes G + D^T D)^{-1}(X_c) \tag{3.15}$$

The full derivation can be found in Appendix A.2, which follows the same procedure as ICP-T. For ICP-A we end up with the kernel being $\mathcal{K} = DKD^T$ with K being the generalized inverse Laplacian matrix and DD^T being the dot product kernel [76], which is what makes the ICP-A algorithm invariant to rotations around the origin. To keep the observations as simple deformations, they are corrected to adjust for the modeling of affine transformations $\hat{U} = X_c - (X_R + \mathcal{K}^{-1}X_R)$.

Many papers exist which expand on [3] for their specific domains. One paper is [38] where they learn a SSM from a training set over the possible

affine transformations. As described in Section 3.2.2.1, the GiNGR framework also allows for the usage of statistically learned kernels.

3.2.1.5 Generalized Inverse Laplacian Correction Term

We inspect the correction term by isolating Z from

$$(L + I)^{-1}\hat{U} = Z + L^{\dagger}(L^{\dagger} + I)^{-1}\hat{U} \tag{3.16}$$

when $W = I$ then Z is exactly the mean observation $\bar{U}$ which is missing from $\hat{X}_R$. The same setup for ICP-A is likewise the average affine transformation that is missing. In practical terms, this means that using the generalized inverse Laplacian in the GiNGR framework cannot correct for a rigid offset like the ICP-A algorithm. Fortunately, we can compute the rigid transformation from [94] as also done in the BCPD algorithm. In Fig. 3.1 we compare the ICP algorithms and the GiNGR implementation of the ICP-A algorithm on a partial data registration problem, where a small rigid offset is found (femur data from [61]). Note that this experimental setup is similar to the synthetic shape registration experiment from [3].

As in [3], we see that the ICP-T algorithm cannot adjust for the rigid offset, whereas the ICP-A method correctly adjusts for this when we start from a very large initial noise term λ_s. The GiNGR method uses the kernel from ICP-A and it uses [94] to estimate the correction term Z, thereby approximating the initial alignment offset.

3.2.2 ICP and CPD Comparison in GiNGR

In this section, we compare the ICP and CPD algorithms as derived in Section 3.2.1. There are two main points that distinguish the methods, namely the kernels that are used and the way that the observed deformations $\hat{U}$ are found. In Fig. 3.2 the different kernels are visualized, with the color indicating the absolute kernel correlation to a landmark on the right foot of the Armadillo[2].

[2] graphics.stanford.edu/data/3Dscanrep/

(a) (b) (c) (d)

Figure 3.1: Femur registration comparison: (a) red is the partial target, yellow is the reference and the white are completions with (b) ICP-T, (c) ICP-A, (d) GiNGR with the generalized inverse Laplacian and dot product kernel.

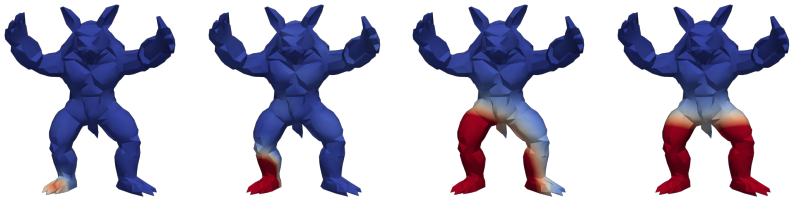
The generalized inverse Laplacian matrix as used in the ICP methods has a direct relation to the commute distance matrix [95]. In comparison to the Gaussian kernel that operates with the Euclidean distance, the commute distance is a measure between points on a graph. This makes it favorable to use if the shape that is being analyzed has spatially close but uncorrelated items as for instance the feet of the Armadillo. An interesting observation is how the dot product kernel adds correlation to points located 180° away. This explains how the femur in Fig. 3.1 is able to recover its correct position. In Fig. 3.3 we also included the simple symmetrical kernel from [27] with a Gaussian kernel being symmetrized. The symmetrical kernel is useful for registering symmetrical objects such as the face or the skull. Also notice how the Laplacian based kernel is invariant to a pose deformation of the reference surface. In Fig. 3.3 we additionally show how the variance parameter β^2 selected for the Gaussian kernels affects the point correlation.

The observed deformations in the ICP methods use the closest point heuristic and filter away points based on their normal directions, on whether

(a) (b) (c)

Figure 3.2: Visualization of the absolute point correlation to a landmark on the right foot. The top and bottom row shows the kernel correlation on the same reference surface in two different positions. (a) Gaussian with large variance, (b) Generalized inverse Laplacian, (c) Generalized inverse Laplacian and dot product. Notice how the Laplacian based methods are invariant to the pose difference as it works on the surface distance in comparison to the Gaussian kernel which uses the Euclidean distance between points. Blue: small correlation, red: large correlation.

they are located on the boundary of a surface and on whether the closest point deformation vectors intersect the surface itself. On the contrary, CPD based methods span a complete correspondence probability matrix between all point pairs in X_R and X_T. Probabilistic correspondence works well when the global structure of the target is given and can even recover large rigid transformation offsets. Another advantage of probabilistic correspondence methods are noisy point-sets where a closest point approach can have difficulties filtering away noisy points. However, the methods fail in a partial data registration scenario as shown in Fig. 3.4. We also observe that the automatic computation of σ^2 in BCPD works well to get the rigid offset and the global structure of the target correctly, but it can have difficulties to explain local intrinsic details. It is therefore advisable to use a closest point method after a global fit has been



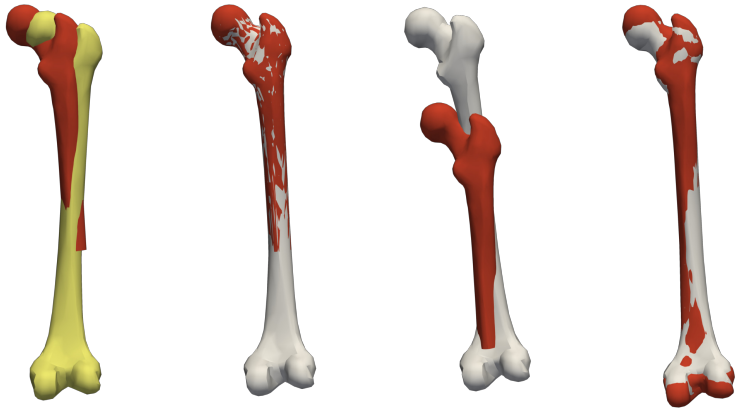
(a) (b) (c) (d)

Figure 3.3: Visualization of the absolute point correlation to a landmark on the right foot: (a) Gaussian with a small variance, (b) Gaussian with a medium variance, (c) Gaussian with a large variance, (d) Symmetrical Gaussian with a large variance. Blue: small correlation, red: large correlation.

found with a probabilistic correspondence method. If the generalized inverse Laplacian matrix is used, it is important to know that the complete matrix needs to be spanned before a low-rank approximation can be computed for the GPMM. This means that the kernel is less useful for surfaces with many points. The Gaussian kernel can be described as a kernel function and efficiently approximated based on the Nyström method as done in [35, 52].

3.2.2.1 Point Distribution Models

The simple GiNGR framework can be used with analytically defined kernels to bring point-sets in correspondence to build a statistical kernel. Given a statistical kernel (Eq. (2.10), Eq. (2.11)), we can use it the exact same way as an analytically defined kernel to analyze novel data. Statistical and analytic kernels can even be combined to create an augmented kernel as also highlighted in [52]. Note that, if the GiNGR framework is used to fit a standard PDM, then the GPMM reference update step should be skipped. The specific case of using ICP to fit a PDM has previously been shown in [11] but this without making the connection to GPMMs and other registration algorithms such as CPD.



(a) (b) (c) (d)

Figure 3.4: Femur registration comparison: (a) red is the partial target, yellow is the reference and the white are completions with (b) GiNGR with Gaussian kernels (c) GiNGR as BCPD fails to register partial data, (d) GiNGR as BCPD registration of complete target.

3.3 Related Work

In Section 3.2.1 we showed how the CPD and BCPD can be reformulated in the GiNGR framework. Furthermore, we showed how the reformulation of ICP from [3] to a least-square optimization problem allows us to reformulate it into the GiNGR framework using the Woodbury matrix identity. In this section, we highlight how other non-rigid registration methods estimate the correspondence field $\hat{U}$ and what kind of smoothness regularization they apply.

3.3.1 Correspondence Estimation

In [13] they introduce a binary linear assignment method to establish correspondence between two point-sets. A sparse correspondence matrix is used to identify corresponding points and potential outliers. The method is then extended to soft-assigning values which is similar to the P matrix in the CPD method, but without explicitly formulating the point-set registration as a max-

imum likelihood estimation problem. In [12] they introduce the Trimmed ICP (TrICP) which sorts the closest points according to their Euclidean distance to the target. The closest points with a distance below a set threshold are then filtered away to only compute the transformation on a subset of the reference points. In [63] they introduce probabilistic point matching (PPM) which is what is used in CPD. Later on, they extend their work to part-based PPM [64] which allows for local different transformations.

In [40] they generalize the rigid EM-ICP [28] to affine transformations. The EM-ICP assumes Gaussian noise on the target points and uses an EM-like algorithm to maximize the correspondence probability of the reference points. They also show that it tends toward the standard ICP when a small variance is chosen. This is again a probabilistic correspondence method for point-sets that is robust to noise and works similar to CPD.

Several feature-based alternatives exist to probabilistic correspondence estimation and the closest point heuristic. The spin image is used as a feature descriptor for 3D surfaces [44] as used in [53] or fast point feature histogram (FPFH) [82] as used in [54, 55] which captures the underlying surface model properties (local geometry, surface normals, curvatures etc.). A three-stage iterative point registration process is used in [23] where the last step involves local affine transformations for each vertex. A closest point method based on 8 dimensional (vertex position, vertex normals, and curvature) is used to estimate the correspondence between the two surfaces. A multi-level fitting is used in [9] where they first use feature points as the correspondence estimate and in a later stage they change to locally weighted ICP in order to register local details. In [49] they increase the reliability of simple closest point correspondence estimation, by introducing a 2-way closest point search. This simple filtering step can be an effective heuristic to remove bad correspondence estimates.

All of the mentioned methods have the potential to be used in the GiNGR framework as they provide a simple point deformation estimate. This list of correspondence estimation methods is by no means complete, but it provides a good overview of the different methods that exist depending on the application

at hand. Additionally, an extensive review of different ICP methods can be found in [73].

3.3.2 Smoothness Regularization

In Section 3.2.2 we showed how the regularization used in CPD and the ICP methods lead to different kernel priors and we visually compared different kernels in Fig. 3.2 and Fig. 3.3. In [86] they show how the covariance operator can generally be seen as an intermediate representation to convert between kernel functions, RKHS, regularization operators, and GPs. The majority of registration papers describe their optimization and regularization terms in a single algorithm, instead of separating them as in the GiNGR framework by using GPMMs. Besides the kernels already mentioned, in [35] they show how different positive-definite kernels can be used in the CPD algorithm, such as the inverse multiquadric and the rational quadratic kernel. Another popular regularization prior is thin-plate splines (TPS), as used in [13] to define the non-rigid deformations on a range of synthetic data and brain-structure point-sets. TPS is also used in [9] for range scanning registration. In [105] they use a spherical analog of TPS for inter-subject brain surface registration and in [27] they use a multi-scale B-spline kernel together with a symmetrical kernel to register 3D scans of faces. Unfortunately, to our knowledge, there is no single best kernel that works in all domains. The kernel function needs to be designed on a case-by-case basis. Kernels can also be added or multiplied together to complement each other. A complete list of valid kernel mixtures is given in [22].

3.3.3 Gaussian Process Regression

GPR has previously been used in different registration pipelines. In [36] they are working in a SLAM environment and use GPR to extract key-points from underwater sonar images. A simple squared exponential kernel is used to model low-frequency elevation changes on the seafloor. In a fine-tuning step, an ICP like approach is used. In [96], GPR is used in an iterative fashion

to include more and more points into the training set to estimate the road-shape based on multiple 3D lidar scans. Starting from a few inlier points, they iteratively add or reject more points to the inlier dataset based on the GP posterior. This way of iteratively adding robust points to the observations can easily be integrated as a strategy in GiNGR. A similar strategy was also used in [75] where they iteratively classified the surface into healthy and pathological regions and perform a non-rigid optimization step. In [61] they are using GPMMs and closest point correspondence to obtain a distribution of possible surface registrations using the Metropolis-Hastings method. They empirically show how the stochastic method improves a simple ICP algorithm by being able to escape local minima. The method can be seen as an extension to GiNGR for applications that are primarily interested in the distribution of different registrations instead of a single best registration.

3.4 Conclusion

In this chapter, we have shown how popular non-rigid registration methods for point-set and surface registration can be generalized with the GiNGR framework. The unifying framework separates the optimization from the modeling and allows for direct comparison of registration methods. We have shown how different ICP algorithms are fundamentally different from popular CPD based algorithms by systematically comparing their regularization, correspondence, and robustness attributes. We have identified how their correspondences are approximated, how the regularization can be reformulated in terms of the underlying kernels, and how they are using different strategies to robustly allow for finer details to be fitted in each iteration. By explaining the ICP algorithms with GiNGR, we have identified that the ICP-A is better in registering partial surfaces with a small rigid offset due to the dot product kernel in comparison to the ICP-T algorithm. In the CPD algorithm, we have identified that the underlying independent noise assumption is individually computed for each point based on the correspondence probability matrix. Beyond comparing existing registration algorithms, GiNGR can be used to easily

change between different methods during registration or to create methods that combine parts from existing algorithms. The framework gives additional benefits such as a clear concept for including expert-annotation, multi-resolution fitting for faster and more robust registration procedures, and the ability to use statistical deformation priors. In the following chapter, we will look at non-rigid registration where the individual iterations are probabilistic in comparison to the deterministic steps performed in GiNGR. The conversion to probabilistic registration has many advantages such as, soft correspondence, uncertainty quantification, and the possibility to escape local minima. We will show how GiNGR is a special instance under the probabilistic framework. This means that if we can convert existing algorithms into GiNGR, then we can also convert them from deterministic registration to probabilistic registration.

4

P-GiNGR: Probabilistic Registration

In this chapter, we introduce P-GiNGR, which is a generalization of GiNGR to a probabilistic registration framework. We will show how the standard deterministic GiNGR is just a special instance under the P-GiNGR framework. The main parts of this chapter have previously been published in [61]. We will first start out with the motivation for a probabilistic registration framework before we formally introduce P-GiNGR. We also show how P-GiNGR and GiNGR compare in registering femur bones, where the biggest challenge is establishing correspondence along the long smooth surface of the femur shaft. Finally, we refer to related work on probabilistic registration and uncertainty quantification in registration methods.

4.1 Motivation

In this chapter, we propose an efficient, fully probabilistic method for surface registration. With P-GiNGR we formulate non-rigid surface registration as an approximation of the posterior distribution over all possible instances of point-to-point correspondences, given a target surface. With this approach, we can also quantify the registration uncertainty, which is the remaining variance in the posterior distribution. The ability to quantify the uncertainty of a surface

registration is important in many areas of shape analysis. It is especially useful for the reconstruction of partial data, or for the analysis of data where the exact correspondence is unclear, such as smooth surfaces. Within the medical field, the uncertainty of partial data reconstruction is needed to make informed surgical decisions [77]. Uncertainty estimates can also be used to build better generalizing Point Distribution Models (PDMs) by assigning an uncertainty measure to each landmark [40, 56].

P-GiNGR makes use of the MH algorithm to sample surface registrations from the posterior distribution. In comparison to many deterministic methods, our method can escape local optima and aims to capture the full posterior distribution of registrations. The efficiency of the method is gained by introducing a specialized way of constructing proposals based on estimated correspondence pairs. Our proposal can make informed updates while maintaining the theoretical convergence properties of the MH algorithm.

As in the previous chapters, we focus on non-rigid registration. However, MH allows for easy integration of proposal distributions of parameters other than the shape. The proposal distribution can therefore easily be extended to include translation, rotation, scaling as shown in [65], or texture, illumination, and camera position as in [83].

4.2 Method

Instead of formulating the registration as an optimization problem, we formulate it as Bayesian inference. We obtain the posterior distribution of parameters $\boldsymbol{\alpha}$ given the target surface Γ_T as:

$$P(\boldsymbol{\alpha}|\Gamma_T) = \frac{P(\Gamma_T|\boldsymbol{\alpha})P(\boldsymbol{\alpha})}{\int P(\Gamma_T|\boldsymbol{\alpha})P(\boldsymbol{\alpha})d\boldsymbol{\alpha}}. \tag{4.1}$$

The prior probability, computed with Eq. (2.7), pushes the solution towards a more likely shape given the GPMM space by penalizing unlikely shape deformations. The likelihood term can easily be customized with different distance measures and probability functions depending on the application goal at hand. As in the majority of deterministic registration methods, we are usually inter-

ested in modeling the average Euclidean distance between two surfaces ($\mathbb{d}_{l^2}$), for which we can use the independent point evaluator likelihood:

$$P(\Gamma_T|\boldsymbol{\alpha}) = \prod_{i=1}^{n} \mathcal{N}(\mathbb{d}_{l^2}(\Gamma_T^i, \Gamma[\boldsymbol{\alpha}]^i); 0, \sigma_{l^2}^2), \tag{4.2}$$

as also used in [65]. The Euclidean distance between the i-th point $\Gamma[\boldsymbol{\alpha}]^i \in \mathbb{R}^d$ and its closest point on the surface Γ_T^i is rated using a zero-mean normal distribution with the expected standard deviation for a good registration. The variance $\sigma_{l^2}^2$ is the observation noise of the points of our target surface. We can register for a better Hausdorff distance [4] by changing the likelihood to:

$$P(\Gamma_T|\boldsymbol{\alpha}) = Exp(\mathbb{d}_H(\Gamma_T, \Gamma[\boldsymbol{\alpha}]); \lambda_H) \tag{4.3}$$

with $\mathbb{d}_H$ being the Hausdorff distance between the two surfaces and Exp being the exponential distribution controlled by the rate parameter λ_H, with probability density function (pdf) $p(d) = \lambda_H e^{-\lambda_H d}$.

4.2.1 Approximating the Posterior Distribution

The posterior distribution defined in Eq. (4.1) can unfortunately not be obtained analytically. Yet, we can compute the unnormalized density value for any shape described by $\boldsymbol{\alpha}$. This allows us to use the Metropolis-Hastings algorithm [79] to generate samples from the posterior distribution in the form of a Markov chain. The MH algorithm is summarized in Alg. 2. A general way to explore the parameter space is to use a random-walk proposal, i.e. a Gaussian update distribution in the parameter space

$$Q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}) \sim \mathcal{N}(\boldsymbol{\alpha}, \sigma_l). \tag{4.4}$$

We usually combine differently scaled distributions, each with a specified σ_l, to allow for both local and global exploration of the posterior distribution. For each proposal, one distribution is chosen at random.

Algorithm 2: Metropolis-Hastings sampling

$\boldsymbol{\alpha}^0 \leftarrow$ arbitrary initialization
for $i = 0$ to S **do**
  $\boldsymbol{\alpha}' \leftarrow$ sample from $Q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}^i)$
  $t \leftarrow \dfrac{q(\boldsymbol{\alpha}^i|\boldsymbol{\alpha}')p(\Gamma_T|\boldsymbol{\alpha}')p(\boldsymbol{\alpha}')}{q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}^i)p(\Gamma_T|\boldsymbol{\alpha}^i)p(\boldsymbol{\alpha}^i)}$. {acceptance threshold}
  $r \leftarrow$ sample from $\mathcal{U}(0,1)$
  if $t > r$ **then**
    $\boldsymbol{\alpha}^{i+1} \leftarrow \boldsymbol{\alpha}'$
  else
    $\boldsymbol{\alpha}^{i+1} \leftarrow \boldsymbol{\alpha}^i$

4.2.2 Correspondence-proposal

A random walk in the parameter space of the prior GPMM is time-consuming as it usually is high-dimensional. Instead, we propose to accelerate convergence by using an informed proposal. For the proposal to reach a unique stationary distribution, we have to be able to compute the transition probability which requires the proposal to be stochastic. We propose the Correspondence-Proposal (CP) which takes estimated correspondence pairs between the model and the target into account to guide the proposed change. Internally, we use an analytical posterior model u_p based on estimated correspondences to propose randomized informed samples. From a current state $\boldsymbol{\alpha}$, we propose an update $\boldsymbol{\alpha}'$ by executing the following steps (visualized in Fig. 4.1):

1. Construct the set of observations L based on corresponding landmark pairs as in Eq. (2.13) and define the noise $\epsilon_i \sim \mathcal{N}(0, \Sigma_i)$ using Eq. (4.6).

2. Compute the analytic posterior u_p (Eq. (2.14)) with L and $\{\Sigma_i\}$.

3. Get $\boldsymbol{\alpha}^o$ by first drawing a random surface[3] from the posterior model u_p and then projecting it into the prior model u.

[3] Sampling all $\boldsymbol{\alpha}_i$ independently from $\mathcal{N}(0,1)$ and constructing the shape with Eq. (2.6).

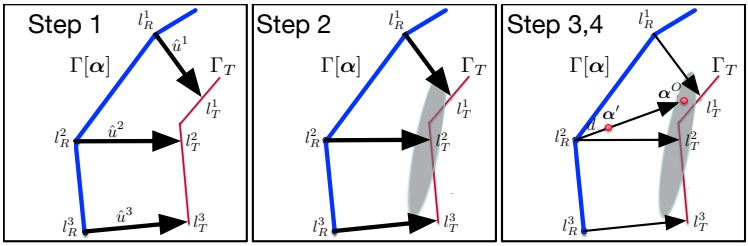


Figure 4.1: Visualization of the CP with the current instance (blue) from the model, and the target surface (red). The correspondence has been estimated using closest-points in the given example. The grey ellipse in the center window shows the landmark noise for l_R^2. The right window shows how an update for l_R^2 is generated based on the posterior sample $\boldsymbol{\alpha}^o$ and the step size d.

4. We generate

$$\boldsymbol{\alpha}' = \boldsymbol{\alpha} + d(\boldsymbol{\alpha}^o - \boldsymbol{\alpha}) \tag{4.5}$$

with $d \in [0.0 \ldots 1.0]$ being a step-length.

A way to model the noise ϵ in 3-dimensions and account for correlated uncertainty is with a low variance along the normal direction and high variance along the surface. The variance at each point in $\Gamma[\boldsymbol{\alpha}]$ can be set to:

$$\Sigma_i = [\boldsymbol{v}_n, \boldsymbol{v}_1, \boldsymbol{v}_2] \begin{bmatrix} \sigma_n^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix} [\boldsymbol{v}_n, \boldsymbol{v}_1, \boldsymbol{v}_2]^T, \tag{4.6}$$

where $\boldsymbol{v_n}$ is the surface normal at a landmark l_R^i on the surface and $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are perpendicular vectors to the normal. The variances along the vectors are set such that $\sigma_v^2 \gg \sigma_n^2$. This noise term ensures that the posterior model from step 2 takes the uncertain correspondence along the surface into account, which is not well defined in flat regions.

If a small step length is chosen in step 4, the current proposal is only adjusted slightly in the direction of the target surface, resulting in a locally restricted step. With a step size of 1.0, the proposed sample is an independent sample from the posterior in Eq. (2.14). This parameter is mainly used to adjust the proposal to have a good acceptance rate.

To construct L as described in step 1, we can make use of simple closest points as used in ICP or take all the possible combinations of correspondence pairs into account to compute the probabilistic correspondence pairs as in CPD.

4.2.2.1 Computing the Transition Probability

Unlike the random-walk proposal, the CP is not symmetric. Therefore, we need to be able to compute the transition probability to ensure convergence of the MH algorithm. For each new proposal $\boldsymbol{\alpha}'$ from the CP distribution, we need to compute the transition probability as part of the acceptance threshold (see Alg. 2 step 4). The transition probability $q(\boldsymbol{\alpha}'|\boldsymbol{\alpha})$ is equal to the probability of sampling the shape corresponding to $\boldsymbol{\alpha}^o$ from the posterior model u_p, computed in step 2.14 of the CP. For $q(\boldsymbol{\alpha}|\boldsymbol{\alpha}')$, the transition probability is computed in the same way. We solve Eq. (4.5) for $\boldsymbol{\alpha}'_o$ after swapping $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ and evaluate the corresponding shape likelihood under the posterior distribution c'_p, which is the posterior distribution computed based on $\boldsymbol{\alpha}'$

4.3 Probabilistic GiNGR

P-GiNGR is a simple extension to the GiNGR framework in Alg. 1. In Alg. 3 we show a rewriting of the MH algorithm introduced in Alg. 2 by using the informed CP. The additional steps needed for P-GiNGR is highlighted with bold text. By comparing Alg. 1 and Alg. 3 we see that GiNGR is a special version of its probabilistic version. In P-GiNGR a random sample is taken from the posterior distribution, whereas GiNGR makes use of the Maximum a posteriori probability (MAP) estimate (the posterior mean). Furthermore, P-GiNGR has a validation/rejection step to check if the newly proposed $\boldsymbol{\alpha}$ should be updated or kept. In GiNGR the newly proposed parameter is in contrast always accepted. The chain of proposals obtained after a burn-in phase of P-GiNGR contains samples from the unnormalized posterior distribution. In the following experiments, we will first look at the convergence properties of P-GiNGR before we compare GiNGR with P-GiNGR where the same

Algorithm 3: Probabilistic GiNGR registration algorithm

Input: Γ_T target surface,
$\hat{\Gamma}_R = \Gamma_R$,
GPMM based on the reference surface $\hat{\Gamma}_R$
Output: $\hat{\Gamma}_R$
Data: set GPMM parameter $\boldsymbol{\alpha} = 0$,
initialize correspondence uncertainty σ^2
while *While (not converted)* **do**
  Estimate correspondence deformations L (Eq. (2.13));
  Perform GP-regression, i.e. compute u_p (Eq. (2.14));
  Draw random sample $\boldsymbol{\alpha}^o$ from u_p;
  Accept of reject, i.e. keep $\boldsymbol{\alpha}$ or update to $\boldsymbol{\alpha}^o$;
  Update GPMM reference $\hat{\Gamma}_R$ (Optional);
  Update correspondence uncertainty σ^2;
end

correspondence estimate is used.

4.4 Experiments

In the following, we perform registration experiments on surfaces of femur bones. We use 50 healthy femur surfaces extracted from computed tomography (CT) images[4]. Each surface is complete, i.e. no holes or artifacts. This setting is optimal for the standard ICP and CPD algorithms. We compare the methods using the same GPMM model using either GiNGR or P-GiNGR. As the surfaces are complete, without artifact and with no rigid offset, this dataset serves as a fair comparison between the probabilistic and deterministic methods.

4.4.1 Convergence Comparison

We compare the convergence properties of our CP and random-walk. For the CP, we use $d = 0.5$ and set the noise along the normal to $\boldsymbol{v}_n = 5.0$ and

[4] Available via the SICAS Medical Image Repository [46].

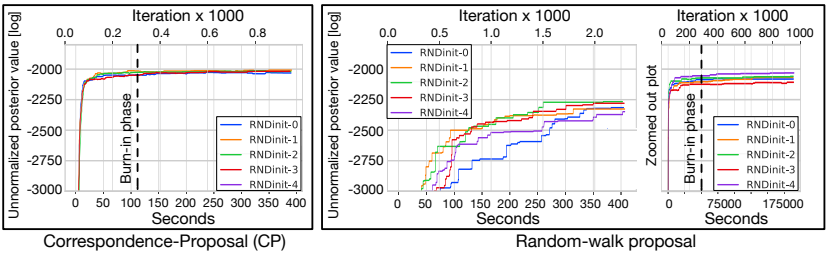


Figure 4.2: Convergence plots for the femur GPMM registrations with 50 components. Our CP is shown to the left and the random-walk (including a zoomed-out plot) to the right. The CP needs 300 iterations, while the random-walk needs more than 200k samples for the burn-in phase without reaching the same registration quality even after 1M iterations. Run-time in seconds is shown on the lower x-axis and the number of MH iterations on the upper.

along the surface to $\boldsymbol{v} = 100.0$. For the random-walk, we use a mixture of the proposals defined in Eq. (4.4), with σ_l being set to six different levels, from $1.0\,\mathrm{mm}$ to $0.01\,\mu\mathrm{m}$, and all six proposal distributions equally likely to be sampled from. For the CP we use a closest point correspondence estimate to compute the correspondence set. In Fig. 4.2, the convergence time of the standard random-walk and the CP is shown. The experiment is performed with a GPMM with a low-rank approximation of 50 components. We randomly initialize the model parameters and start 5 registrations in parallel. As expected, our proposal leads to much faster convergence. In Fig. 4.3 we see a posterior plot comparison of the two proposals. Note how the less likely samples are often accepted which makes it different from deterministic methods such as standard ICP and CPD.

4.4.2 Registration Accuracy - GiNGR vs P-GiNGR

In this experiment we make a direct comparison of GiNGR and P-GiNGR using either a closest-point correspondence estimate or CPD correspondence. As in the previous experiment, we use a femur-GPMM using a mixture of Gaussian kernels and approximated with 50 basis functions. From the model,

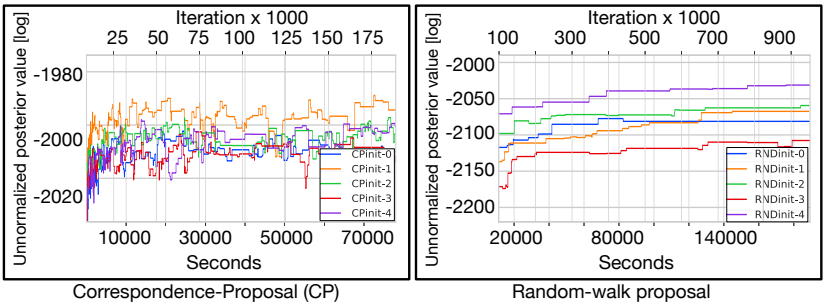
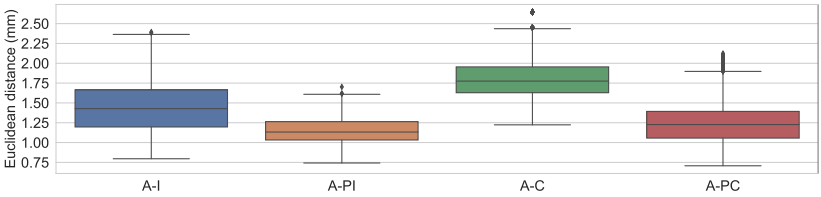


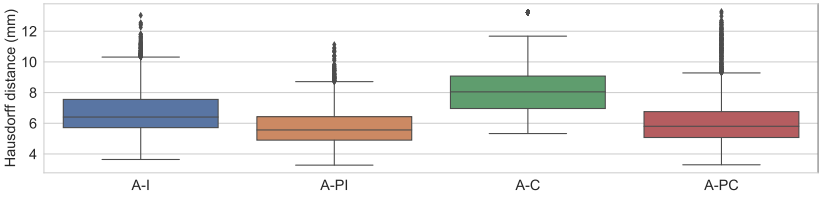
Figure 4.3: Posterior plot comparison of the CP and random-walk. Even with a very small update step to the random-walk, it has difficulties to explore the posterior in the high-dimensional setting. The CP, on the other hand, can more efficiently explore the high-dimensional space. This can be seen in the fluctuation of the posterior value where samples are often accepted even when being less likely. For the random walk, the majority of samples are in contrast rejected due to the difficulty of exploring the distribution in higher dimensional space.

we sample 100 random surfaces as starting points for the registration. We, therefore, end up with 100 registrations for each target. In Fig. 4.4 we show the summary of all 5000 registrations. We see that the probabilistic method performs better both in terms of average Euclidean distance and Hausdorff distance between the final registration and the target surfaces. The GiNGR configurations converge at maximum 100 iterations, whereas we let P-GiNGR explore the posterior distribution for 1.000 samples. In Section 5.1 we show more examples of GiNGR compared to P-GiNGR in a partial data challenge.

In Fig. 4.5a we show the box plot results of 5 target surfaces. The plots for the remaining target surfaces can be found in the Appendix B. We see that GiNGR with CPD is very stable no matter its initialization as it always ends up with the same distance to the target. The method however has problems with not getting finer details of the target. Its probabilistic version with P-GiNGR nearly always converges to a better solution but has some variation in the MAP that is found. In general, we see the same trend for ICP that P-GiNGR almost always converges to better solutions. In Fig. 4.5b and Fig. 4.5d we show the



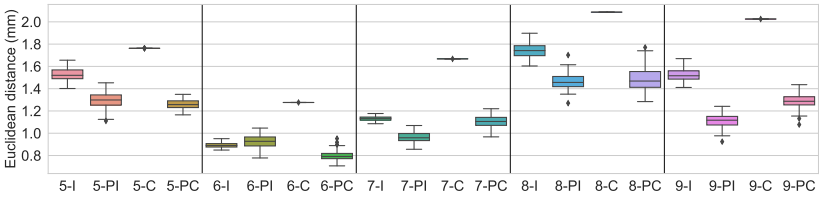
(a) Average Euclidean distance.



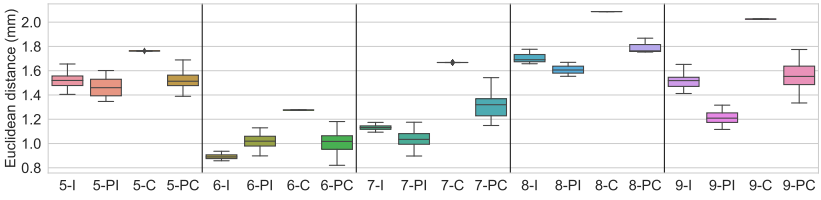
(b) Hausdorff distance.

Figure 4.4: Distances (Euclidean (a) and Hausdorff (b)) between the final registrations and their target surfaces. For each target surface, we randomly initialize 100 registrations. The MAP sample from P-GiNGR is superior to GiNGR using the same correspondence method. *A* refers to the plot being the combined result of all 5000 registrations. The letters refer to the registration method, *I*: GiNGR as ICP, *PI*: P-GiNGR as ICP, *C*: GiNGR as CPD, *CP*: P-GiNGR as CPD.

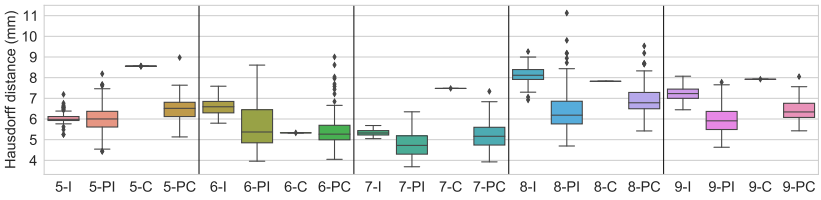
same experiment as in Fig. 4.5a and Fig. 4.5c but we limit the iterations of P-GiNGR to 100. We see that even with as few iterations as its deterministic counterparts, P-GiNGR is better than GiNGR on the majority of the targets and on average over all the targets. The few outliers for our method are therefore probably cases where the chain has not had time to explore the posterior distribution. In Fig. 4.2 we see that the chain usually converges within 100-300 samples, so this should be the absolute minimum number of samples to draw. By drawing more samples, we can additionally explore the distribution of different point-correspondences and therefore estimate the correspondence uncertainty. In Fig. 4.6, we show the uncertainty of the established correspondence of a registration from 1.000 samples (300 samples for the burn-in phase).



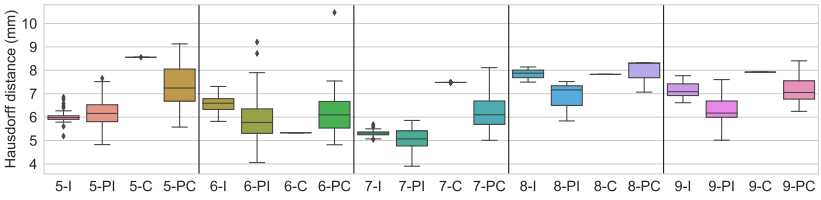
(a) Euclidean distance. 1000 samples with P-GiNGR.



(b) Euclidean distance. 100 samples with P-GiNGR.



(c) Hausdorff distance. 1000 samples with P-GiNGR.



(d) Hausdorff distance. 100 samples with P-GiNGR.

Figure 4.5: The box plot shows the variation of the average Euclidean distances (a, b) and Hausdorff distance (c, d) from all 100 registrations of each target. In (a and c) 1.000 samples are drawn, whereas (b and d) are limited to 100 samples. The MAP sample from our CP is superior to their deterministic equivalents. The number refers to the target id (0 to 49), the letters refer to the registration method, *I*: GiNGR as ICP, *PI*: P-GiNGR as ICP, *C*: GiNGR as CPD, *CP*: P-GiNGR as CPD.

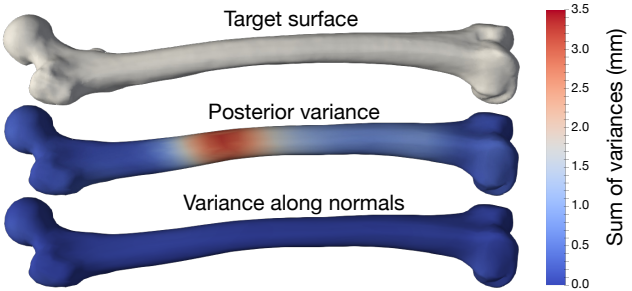
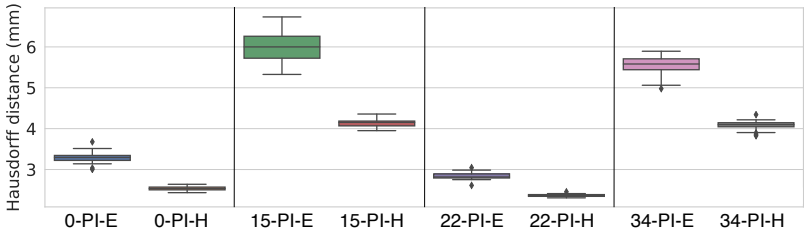


Figure 4.6: Femur registration with uncertainty. The registration uncertainty is visualized with the point-wise sum of variances. As expected, the higher uncertainty of the established correspondence coincides with the shaft region with the least characteristic shape. No variance is observed along the normals, so the uncertainty is only in the correspondence along the surface.

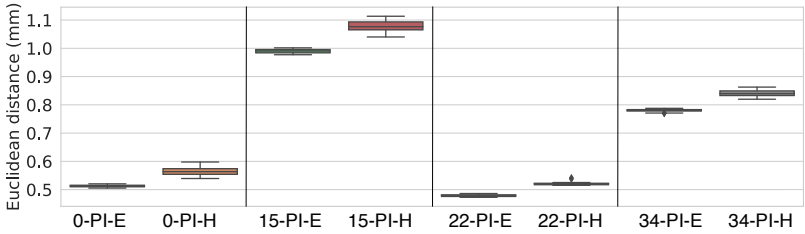
Being able to estimate the correspondence uncertainty is an additional benefit of using P-GiNGR over GiNGR. The uncertainty of individual points from the posterior distribution from a single registration is depicted. Note the high uncertainty values along the shaft, which indicate that the established correspondence is less reliable in that region. No variance along the normals indicates that the uncertainty is purely a correspondence shift within the surface. In Section 5.1 we show additional uncertainty maps when registering partial femurs.

4.4.2.1 Alternative Hausdorff Distance Likelihood

In this section we compare a P-GiNGR setup using two different likelihood functions and see how this influences the MAP of femur registrations. For the likelihood function, we use either the independent point evaluator (E), Eq. (4.2), which models the average Euclidean distance to the target or the Hausdorff likelihood (H), Eq. (4.3), which models the Hausdorff distance to the target. In Fig. 4.7a, we compare the registration results based on their Hausdorff distance. As expected, we see that the P-GiNGR with a Hausdorff likelihood function also obtains the best MAP solution when comparing their



(a) Hausdorff distance.



(b) Average Euclidean distance.

Figure 4.7: Registration result comparison using either the Hausdorff or the Euclidean distance likelihood. (a) Euclidean distance and (b) Hausdorff distance between the MAP samples and the target surfaces. The plots show that the average Euclidean surface distance is only slightly worse when the Hausdorff (H) likelihood is used. The naming on the x-axis first mentions the target id, then the registration method which is P-GiNGR as ICP and finally, (E) refers to a Euclidean metric being used in the likelihood function and (H) refers to a Hausdorff metric being used.

Hausdorff distances. In Fig. 4.7b, we show that while optimizing for the Hausdorff distance, the average Euclidean distance is only slightly increased. This demonstrates the capability to change the likelihood in our framework based on the application at hand.

4.4.2.2 Drawbacks of Deterministic Methods

The main problem with ICP and CPD is that they cannot recover from local optima. If the algorithm finds a corresponding point on the wrong part of the target, we end up with a bad registration. In Fig. 4.8, we show a registration

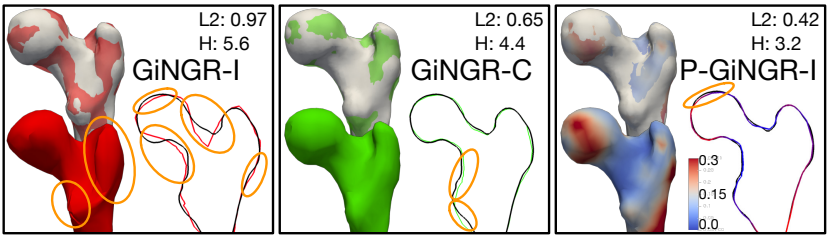


Figure 4.8: The 3 registration methods (GiNGR as ICP or CPD and P-GiNGR as ICP) are shown in separate windows. The registration accuracy for the same target is noted in the form of Euclidean- (E) and Hausdorff- (H) distances (mm). The orange ellipses highlight problematic areas of the registration for each method. For each method, we show the target with the registration overlaid, the 3D-registration, and a 2D slice of the registration (colored) and the target (black). Notice how P-GiNGR (in comparison to GiNGR) shows the correspondence uncertainty (summed point variances (mm^2) for each landmark).

result of the 3 registration methods. The ICP with GiNGR can get the overall length of the bone correct but might end up with a registration, where the structure is folding around itself. The CPD approach with GiNGR is more robust than ICP as it preserves the topological structure of the point-sets but often fails to get local details. We compare the two deterministic methods to P-GiNGR as ICP. As shown in Fig. 4.8, this method has the additional benefit of providing an uncertainty map of the final registration result.

4.4.2.3 Run-time Comparison

The number of components in the low-rank approximation can be seen as regularization of the deformations. More complex local deformations can be obtained using more components. The algorithm run-time scales near linearly in the number of components, with the run-time being 1.5 times slower for P-GiNGR. For models with rank 50, 100 and 200, the CP takes: 46 s, 110 s and 275 s. In comparison, the same correspondence method with GiNGR takes 30 s, 69 s and 155 s. While P-GiNGR method is slower than GiNGR, we still get reasonable run-times while providing more accurate results and estimate

the full posterior instead of a single estimate. The timings have been measured with the closest point estimate as the correspondence estimation method.

4.5 Related Work

In Section 4.3 we showed how GiNGR is a special case of P-GiNGR. This means that the related work suggestions from Section 3.3 mentioning on correspondence estimation and deformation modeling are applicable to P-GiNGR as well. In this section, we highlight related work on informed proposals, uncertainty estimation in non-rigid surface registration and the usage of the MH framework together with GPMMs.

Several alternatives already exist to the random-walk proposal, such as MALA [29], Hamiltonian [69] or NUTS [37]. While the mentioned proposals work well in lower-dimensional spaces and for smoother posteriors, we experienced that they get computationally demanding in high dimensional spaces and have problems when the posterior is far from smooth. In [50], the ICP method is improved using Simulated Annealing and Markov Chain Monte Carlo (MCMC). Their method is a robust version of the ICP algorithm, which can find the global optimal rigid registration of point-sets. However, they do not measure the registration uncertainty, nor are they able to perform non-rigid registrations.

P-GiNGR improves on previous works in the literature in different ways. In [25, 78], the MH algorithm is used to estimate the uncertainty in non-rigid registration. These papers are working on the image domain, and are not transferable to the surface registration setting. MH has also been used in [65] to fit an Active Shape Model to images and in [83] to fit a Morphable Face Model to an image, both of which only make use of the framework to avoid local optima and to easily integrate different likelihood terms. The main problem with the MH algorithm is the commonly used random-walk approach, which suffers from very long convergence times when working in high-dimensional parameter spaces. To overcome this problem, informed proposal distributions can be designed to perform directed sample updates. In [47], a Bayesian Neu-

ral Network is learned to produce informed samples for the MH framework in the case of 3D face reconstruction from a 2D image. This, however, requires training a neural network for each class of shapes to be registered. In [41], local random-walk is combined with an image-dependent global proposal distribution. This image-dependent distribution is, however, not directly transferable to the problem of surface registration. Our method targets settings with no, or limited, available training data and adds to existing methods by providing an uncertainty measure, which is especially important within the medical domain.

4.6 Conclusion

In this chapter, we presented P-GiNGR, which is a probabilistic registration framework. Our main contribution is the formulation of the CP for the MH algorithm, which makes it possible to work in a high-dimensional model space that would be difficult to explore with pure random-walk. Our informed proposal integrates knowledge about the correspondence pairs, which results in faster convergence. Using our framework, different likelihood terms can be combined and used, while the choice is restricted to the Euclidean norm in standard ICP and CPD. By converting existing non-rigid registration algorithms into GiNGR, we have shown how P-GiNGR can directly convert these methods into probabilistic versions. Besides generally obtaining a better global MAP estimate, the probabilistic version can be used to estimate the correspondence uncertainty of a registration. In the following chapter, we will look at different applications for the P-GiNGR framework. We will look at cases of missing data, where our method provides an estimate of the posterior over possible reconstructions. Thus our framework can provide uncertainty measures for critical tasks such as surface reconstruction in the medical domain, as required for surgical decision making. Furthermore, we will show how to build statistical priors that generalize better from partial data observations.

5
Applications

In this chapter, we utilize the P-GiNGR framework for two different applications that involve partial data processing, namely:

- How to analyze partial data with a PDM.
- How to create a PDM from partial data.

In Section 5.1 we make use of P-GiNGR on a partial data reconstruction challenge and show how it performs better than deterministic methods. For analyzing partial data, we also show applications where we go beyond using P-GiNGR to obtain the MAP solution as done in Section 4.4.2. In Section 5.2 we compare posterior estimation methods to analyze partial femur data. We show how the analytical posterior introduced in Section 2.2.3 vastly underestimates the remaining uncertainty and highlight the difficulty when using hard correspondence. On a face dataset with missing noses, we discuss in Section 5.3 a limitation of P-GiNGR when using the independent point evaluator as the likelihood function. In Section 5.4 we show how a femur-PDM can be constructed from only partial data observations. We also show how a hand-PDM can be constructed from increasingly less data and how this influences the generalization and specificity properties of PDMs. In Section 5.5 we show an application of creating a PDM constructed from multiple data sources for the

creation of a molar-tooth-PDM. Finally, we show how P-GiNGR can be used to create a PDM from disjoint data sources such as Computerized Tomography (CT) images and surface scans of faces to create a joint face-skull-PDM.

5.1 Partial Femur Reconstruction - Challenge

In this section, we show the results of using P-GiNGR on a femur reconstruction challenge[5]. For the challenge, one reference surface is given, 50 femur surfaces for learning a PDM and 10 partial surfaces as targets which we show in Fig. 5.1. The goal of the challenge is to reconstruct the missing part of the target surface and get as close as possible to the ground-truth solution in terms of Euclidean and Hausdorff distances. To construct the PDM we use a GPMM with a mixture of Gaussian kernels. We register the model to each of the 50 femur surfaces using GiNGR and obtain an average Euclidean distance between the reference and all the targets of $0.32\,\mathrm{mm}$ and the average Hausdorff distance being $2.65\,\mathrm{mm}$. To put these numbers in context, the reference surface is discretized to 18.879 vertices with an average distance between vertices of $2.00\,\mathrm{mm}$, maximum distance of $3.86\,\mathrm{mm}$ and a minimum distance of $0.08\,\mu\mathrm{m}$. With the femur-PDM constructed from the 50 training surfaces, we use P-GiNGR as ICP, as well as adding random walk proposals for the rotation and translation parameters, such that we are able to recover the pose even if the initialization is slightly off. The description of the complete model parameter vector $\boldsymbol{\theta}$ is described in Section 2.3. As a likelihood model, we use the independent point evaluator as mentioned in Eq. (4.2). This likelihood normally measures the Euclidean distance between the reference and the target. Here we will instead use it to measure the distance between the target and the reference. The reason for this is the target being a subset of the reference surface. If measuring from the reference to the target, some areas will have a large distance due to the missing area on the target surface. After taking 10.000 steps, we choose the MAP solution from the unnormalized pos-

[5] shapemodelling.cs.unibas.ch/femur-reconstruction-challenge

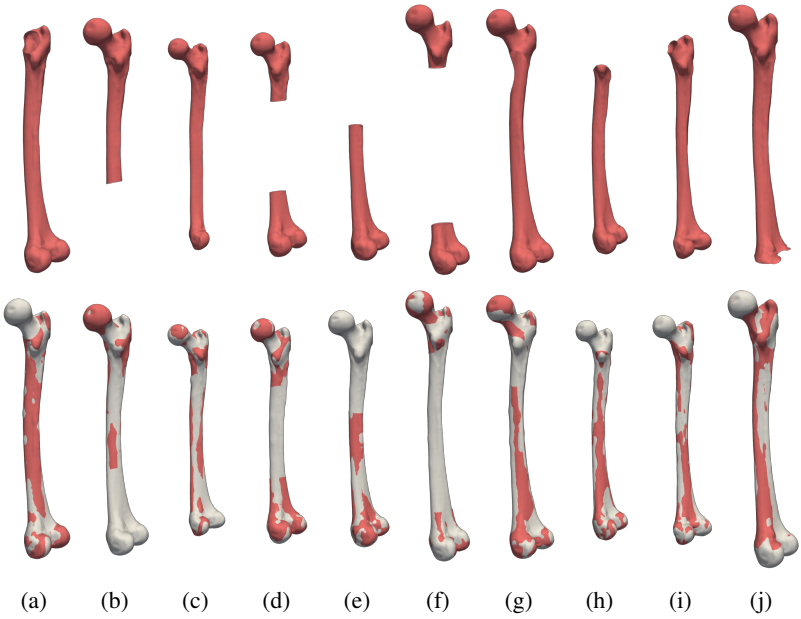


Figure 5.1: Top row shows the 10 target femurs. On the bottom row we overlay the reconstructions with the partial targets.

terior distribution. We obtains an average Euclidean distance of $0.60\,\mathrm{mm}$ and an average Hausdorff distance of $3.02\,\mathrm{mm}$ to the partial surface. In Table 5.1 and Table 5.2 we show the results from the challenge which only measures the distance on the missing part. We end up with the best reconstruction results on average, where the majority of participants are using a standard ICP method. We only use enough landmarks (3-5) to make a decent initial alignment of the PDM. With the same initial alignment we also perform GiNGR as ICP as a comparison and obtain a much worse result. Especially one target (j) has a Euclidean distance of $17\,\mathrm{mm}$. This is mainly a result of a bad pose initialization which we automatically recover with P-GiNGR. In Fig. 5.1 we also show the MAP reconstructions that we obtain. While we obtain the best average Euclidean distance to most targets, we only have a single reconstruction with the best Hausdorff measure. As shown in Section 4.4.2.1 we can

Table 5.1: Average Euclidean distance score on the femur reconstruction challenge. The best results are highlighted in bold. If the best reconstruction was not in top 3, it is shown in the *other* row. Notice how P-GiNGR obtains the best result across many of the targets and it even gets lower than $1.00\,\mathrm{mm}$ on 3 targets as the first method.

Rank Total EEuclidean Distance (mm)

Rank		**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
1 (P-GiNGR)	**1.44**	1.92	1.76	**0.92**	**0.80**	**1.91**	**1.18**	**1.16**	2.58	**1.18**	**0.98**
2	1.67	1.56	1.67	1.34	1.25	2.75	1.33	1.29	2.35	1.46	1.68
3	1.85	2.06	1.58	1.66	1.65	2.29	1.72	1.33	2.73	1.37	2.11
Other	-	**1.11**	**1.47**	-	-	-	-	-	**2.15**	-	-
GiNGR	5.60	4.45	5.93	1.52	3.47	3.32	3.38	1.71	8.31	6.38	17.47

Table 5.2: Average Hausdorff distance score on the femur reconstruction challenge. The best results are highlighted in bold. If the best reconstruction was not in top 3, it is shown in the *other* row.

Rank Total HHausdorff Distance (mm)

Rank		**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
1 (P-GiNGR)	**5.3**	6.35	7.24	4.74	3.65	8.25	**4.00**	5.69	8.19	3.65	**3.53**
2	5.64	3.77	**5.07**	4.85	4.60	11.68	5.02	5.53	6.66	3.50	5.75
3	6.47	5.81	7.30	5.35	8.83	8.28	5.41	6.65	8.25	3.00	5.82
Other	-	**3.31**	5.54	**3.92**	**3.52**	**6.95**	-	**2.28**	-	**5.88**	-
ICP	15.25	12.06	19.80	5.94	8.16	16.12	12.25	5.70	26.28	13.78	32.43

however easily change the likelihood function to the distance measure which is most describing to the task at hand. This is in contrast to traditional methods in GiNGR which mainly use the Euclidean distance. As also shown in Section 4.4.2.1, changing the likelihood to the Hausdorff likelihood negatively affects the average Euclidean distance measure. What is ultimately gained in a better Hausdorff distance then produces a worse average Euclidean distance.

As an additional benefit of P-GiNGR we do not only obtain a single best reconstruction, but a whole distribution of reconstructions. In Fig. 5.2 we visualize the individual point uncertainty from the posterior distributions for each target. We both show the sum of variances and the variance along the

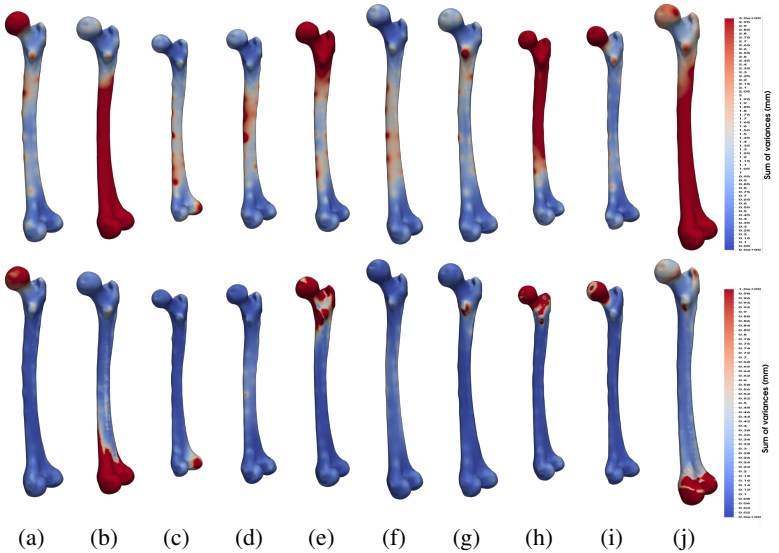


Figure 5.2: Top row shows the sum of variances from the femur posterior distributions. On the bottom row we compute the uncertainty in the normal direction. We see that especially if an end of the femur bone is missing, a large uncertainty is found along the femur shaft as many different correspondence possibilities are explored. The normal direction uncertainty can in contrast be used to automatically detect the missing part of the femur.

normals within the reconstructions. We see how the uncertainty in the normal directions can be used to automatically detect the missing area of the target in areas of high curvature.

5.1.1 Discussion

In this section we showed how P-GiNGR can be used to obtain good results on a simple reconstruction challenge. We also showed how the probabilistic method can provide us with a correspondence uncertainty map and how the uncertainty in the normal direction can be used to automatically detect the missing part of the femur. In the following section we will show how the underlying task of the reconstruction challenge is in fact ill-posed as there

does not only exist a single possible reconstruction, but a whole distribution of possible reconstructions. We will especially focus on how analytical posterior methods vastly underestimate the variance in reconstruction cases.

5.2 Partial Femur Surfaces - Posterior Comparison

In this section we compare P-GiNGR to GPR on the task of computing the posterior distribution when a partial target is given. In medical applications, a certainty estimate for a reconstruction is often required. This estimate can be computed as the likelihood of the chosen reconstruction in the distribution of all other possible reconstructions. These experiments can also be seen as comparing hard and soft correspondence properties in registration. In Fig. 5.3 we show how the hard correspondence property in GPR has close to zero variance left even in areas far away from the partial observation. In contrast, with P-GiNGR we are able to have a more reliable estimate in the uncertainty of the reconstruction results. When using analytical posterior methods as introduced in Eq. (2.14) both a fixed pose and fixed point-to-point correspondence is assumed [1, 7]. Furthermore, the analytical-posterior requires an initial dataset alignment before it can be computed as highlighted in [1]. In absence of exact point-to-point correspondence, those conditions are impossible to fulfill. With P-GiNGR we are able to estimate the posterior distribution from partial surface knowledge without having to perform a dataset realignment.

Initially, the reference is aligned to the partial target from a few manually clicked landmarks. This alignment might not be the global ideal alignment as we do not have the global frame of the target. In the specific case of the femur it is therefore advisable to mix an informed CP with a random-walk proposal Eq. (4.4) for the rotation and translation parameters. In the following, we will even show how the CP and random-walk proposal can be combined into a projection-proposal. The experiments in this section have previously been published in [59].

5.2.1 Projection-Proposal: A Correspondence-Proposal

Once the system has converged, we would ideally like to keep the part of the PDM fixed where correspondence can be established and is stable, as we are interested in the posterior distribution given partial surface information. To do so, we can make a small add-on to the CP which steps are given in

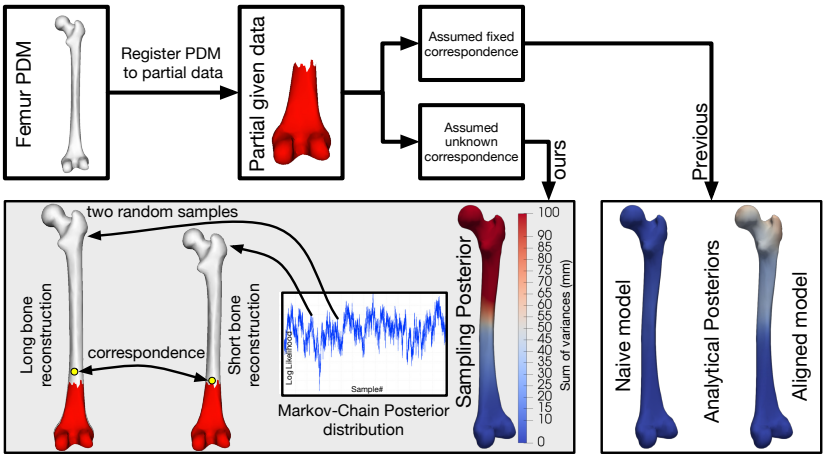


Figure 5.3: A femur PDM is first registered to the partial surface to establish point-to-point correspondence. If we assume fixed correspondence, the posterior is computed analytically. With our method, we assume unknown correspondence and compute the posterior as a Markov-Chain. Note the correspondence difference to the available data visualized with the yellow markers on the long and short bone reconstructions. The colored bones show the uncertainty, computed as the sum of the variances at each landmark with the different methods. Our method clearly shows more variability (red) far away from the partial surface, but at the same time has low variability (blue) at the known area.

Section 4.2.2:

1. Construct corresponding landmark pair
2. **Propose a random pose update from $Q(\boldsymbol{t}'|\boldsymbol{t})+Q(\boldsymbol{R}'|\boldsymbol{R})$, while keeping the current shape parameters $\boldsymbol{\alpha}$. fixed**.
3. Compute the analytic posterior.
4. Draw a random surface sample.
5. Scale predicted update with step length.

The highlighted step 2 is the extension to the existing steps in P-GiNGR. For the estimated correspondence landmark pairs, we define a zero deformation

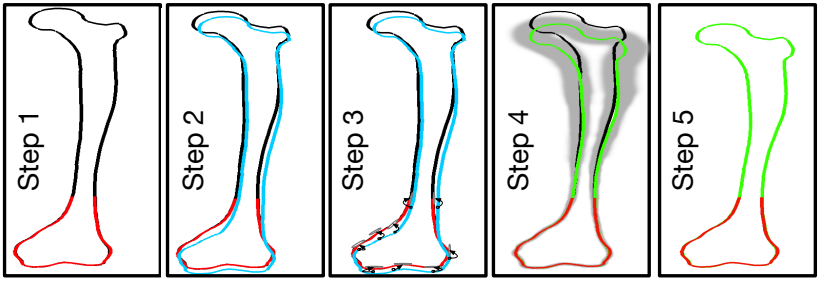


Figure 5.4: Visualization of the projection-proposal steps. Step 1: partial target (red), current reference (black), compute correspondence L based on closest points. Step 2: Reference after applying a random rigid transformation (cyan). Step 3: Compute the analytical posterior with L based on Eq. (2.14). Step 4: Draw a random sample from the posterior distribution (green) and project back into the model space. Step 5: Adjust step length.

vector for all points on the model which has a correspondence point on the target (i.e. not on the edge of the surface). Before computing the analytical-posterior, we then make a random rotation or translation proposal. Just as in P-GiNGR we use an anisotropic noise term ϵ, to simulate correspondence shift along the surface, where a low variance along the normal and a higher variance along the surface is chosen. The steps of the projection-proposal are visualized in Fig. 5.4. The transition probability can again be computed using the posterior distribution from step 3 as also shown in Section 4.2.2.1.

5.2.1.1 Projection-Proposal Importance

We need i.i.d. samples to compute the variance, which means that we need to find the number of samples to be discarded from the Markov-Chain before an independent sample is found. We compute the auto-correlation of the individual shape parameters and look for the number of samples needed to reach near zero correlation. We report 50 samples for the projection-proposal and 30×10^3 samples for a random-walk. While the random-walk requires 600 times more samples, the projection-proposal with the closest point correspondence estimate is only 6 times slower, making it overall 100 times faster.

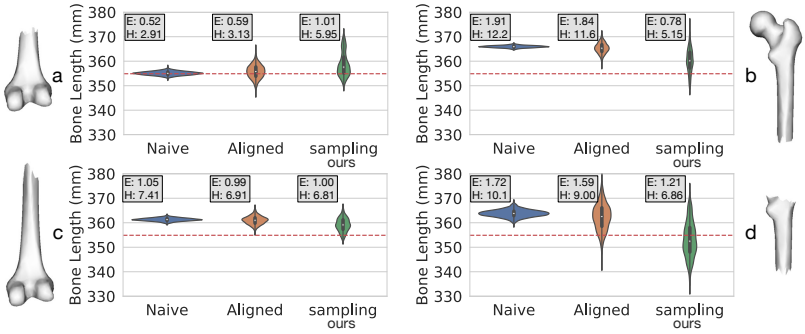


Figure 5.5: Violin and box plots of bone length prediction in mm using the analytical-posterior and our method P-GiNGR. All plots (a,b,c,d) concern the same ground-truth bone (length visualised with the red-dashed line), but differ in how much of the bone is given. The average Euclidean (E) and Hausdorff (H) distances from the ground-truth surface to the mean surface from the distributions are in mm.

We compute the bone length variance based on the distance between two landmarks. The length variation converges at 10^3 samples with the projection-proposal. For the random-walk, we need 500×10^3 samples to achieve the same length variance. With the projection-proposal, we can, therefore, explore similar variance numbers with less than $1\,\%$ of the samples needed by the random-walk.

5.2.2 Experiments

For the experiments, we use 61 femur meshes extracted from CT images. We use 50 femurs for the femur PDM (femur lengths, mean: $372\,\mathrm{mm}$, min: $322\,\mathrm{mm}$, max: $437\,\mathrm{mm}$) and 11 for the test-set (femur lengths, mean: $372\,\mathrm{mm}$, min: $322\,\mathrm{mm}$, max: $441\,\mathrm{mm}$). The PDM contains a total of 1622 landmarks. Each test femur is divided into several partial meshes from where the posteriors are estimated. In Fig. 5.5 a subset of the cuts are shown.

5.2.2.1 Experimental Setup

We compare the posterior from P-GiNGR with two different PDMs created from the same data. We will refer to the analytical-posterior computed without aligning according to the dataset as the naive-posterior and the analytical-posterior with the dataset alignment as the aligned-posterior. The naive-posterior is created as a normal PDM and the aligned-posterior is computed based on the estimated corresponding points in the target. This realignment step is mentioned as a necessary step in [1] when using a PDM to register partial targets. For the aligned-posterior, we need to estimate the observed points in the PDM. For this we use a standard ICP method with GiNGR.

In the overview image of our method (Fig. 5.3), the posterior variability of the different methods is visualized with colors on the full femur bone. Very little variance is maintained in the naive-posterior, which highlights the importance of dataset alignment when computing the analytical-posterior. P-GiNGR with the projection proposal contains 2 to 3 times more variability than the aligned-posterior, suggesting that the full variability cannot be obtained using hard correspondence.

5.2.2.2 Length Estimation of Partial Femurs

We compare the mean and the variance of bone lengths of the different posterior estimation methods. A landmark is placed at each end of the femurs and the variability of the distance between the two landmarks is reported. For the analytical-posteriors, we randomly sample 10^3 shapes from the posterior models to be used for the estimate. For the projection-proposal we take 10^3 samples with 50 sample spacing in between. The bone length results for test femur 1 are shown in Fig. 5.5. Notice the difference between the results for partial bone *a* and *c*. More data is available in *c*, which results in a more narrow distribution, whereas the correspondence used in *c* is worse, making the ground-truth surface very unlikely under its distribution. The P-GiNGR results of the remaining test femurs are shown in Fig. 5.6.

We observe that both of the analytical-posterior methods sometimes fail to estimate the ground-truth length within their variability for most of the cuts.

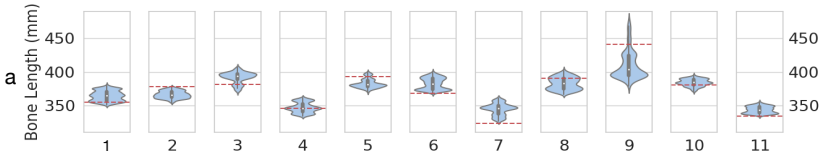


Figure 5.6: Violin and box plots of bone length prediction in mm. All plots of different ground-truth bones. The posteriors are computed with P-GiNGR and the projection proposal.

In contrast, P-GiNGR can explain the shape length accurately.

5.2.2.3 Importance of Correct Correspondence

The quality of a surface reconstruction can be measured with the average Euclidean or Hausdorff distance to the ground-truth. These measures are, however, not a good indicator for the registration quality when large uncertainty exists in the correspondence. In Fig. 5.7 we show the same bone length experiment as in Fig. 5.5, but only for the aligned-posterior computed using different correspondences. The different correspondences have been found by initializing the PDM either as a very short, medium or long bone. We see that a close to perfect reconstruction can be found if the ground-truth correspondence is known, but at the same time can extremely over or underestimate the bone length if wrong correspondences are used. The average Euclidean distance from the partial surface to the reconstruction is, in all the cases, less than 0.25 mm, which suggests that the model represents the surface well in the available part.

5.2.3 Discussion

It is difficult to infer the full surface from a bone fragment. This is due to missing exact point-to-point correspondence. Existing methods deterministically find a set of correspondences to estimate the posterior using GPR. We showed how this can result in overconfident posterior estimates if incorrect correspondences are used. We have shown how previous methods even fail to

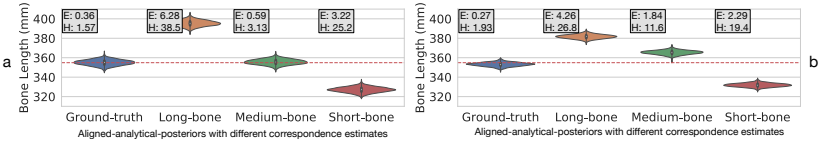


Figure 5.7: Violin and box plots of bone length prediction in mm. All plots concern the same ground-truth bone (length visualized with the red-dashed line). The posteriors are computed with the aligned-analytical-posterior method and differ only in the correspondence which has been used. The average Euclidean (E) and Hausdorff (H) distances from the ground-truth surface to the mean surface from the distributions are in mm. (a) and (b) refers to the same partial shapes as in Fig. 5.5.

explain the ground truth solution under their distributions in an experimental setup with synthetic data. By using P-GiNGR to perform probabilistic registration, we can estimate the posterior distribution of reconstructions without relying on hard correspondences. With P-GiNGR, both correspondence and reconstruction estimates are more accurate than with a traditional analytical approach. We are also able to more reliably estimate the uncertainty of the reconstruction results.

5.3 A Missing Nose - Surface Reconstruction

In this section, we want to highlight one of the difficulties with using P-GiNGR, namely how the likelihood function should be designed. We perform a reconstruction experiment of missing data on face scans. The main difficulty with the face dataset is the incompatibility with the independent point evaluator as the likelihood function. In Section 5.1 we changed the direction of the distance function as the target was a subset of the reference surface. In Fig. 5.8, we show that there is no perfect overlap between the reference and the target surfaces. The likelihood function, therefore, needs to take into account that a different number of points might be evaluated in each iteration. In this section, we show a possible simple extension to overcome this limitation. We compute the posterior distribution of registrations for faces where the nose has been removed. As the face reference, we use the PDM part of the publicly available Basel Face Model (BFM) from [27] together with 10 available face scans as the targets. For this experiment, we use a mixture of CP using ICP and a random-walk proposal for the pose parameters. For each target, we take 10.000 samples and use every 200th sample to estimate the variance of the distribution.

If the number of points that are used to compute the independent point likelihood differ from iteration to iteration, the likelihood function is not well defined, so we might see large jumps in the likelihood value. We therefore need to take the number of evaluated points into account. One possibility from the literature is the collective average likelihood as introduced in [83]:

$$P(\Gamma_T|\boldsymbol{\alpha}) = \mathcal{N}(\mathbb{d}_{CL}(\Gamma_T, \Gamma[\boldsymbol{\alpha}]); 0, \sigma^2_{CL}), \tag{5.1}$$

where

$$\mathbb{d}_{CL} = \frac{1}{N}\sum_{i=1}^{n_{lm}} \|\Gamma^i_R - \Gamma^i_c\|^2. \tag{5.2}$$

Here the closest point of Γ^i_R on the target is Γ^i_c and n_{lm} is the number of landmarks in the reference mesh Γ_R. The usage of the collective average likelihood gives a good global fit, but many of the intrinsic features are missing.

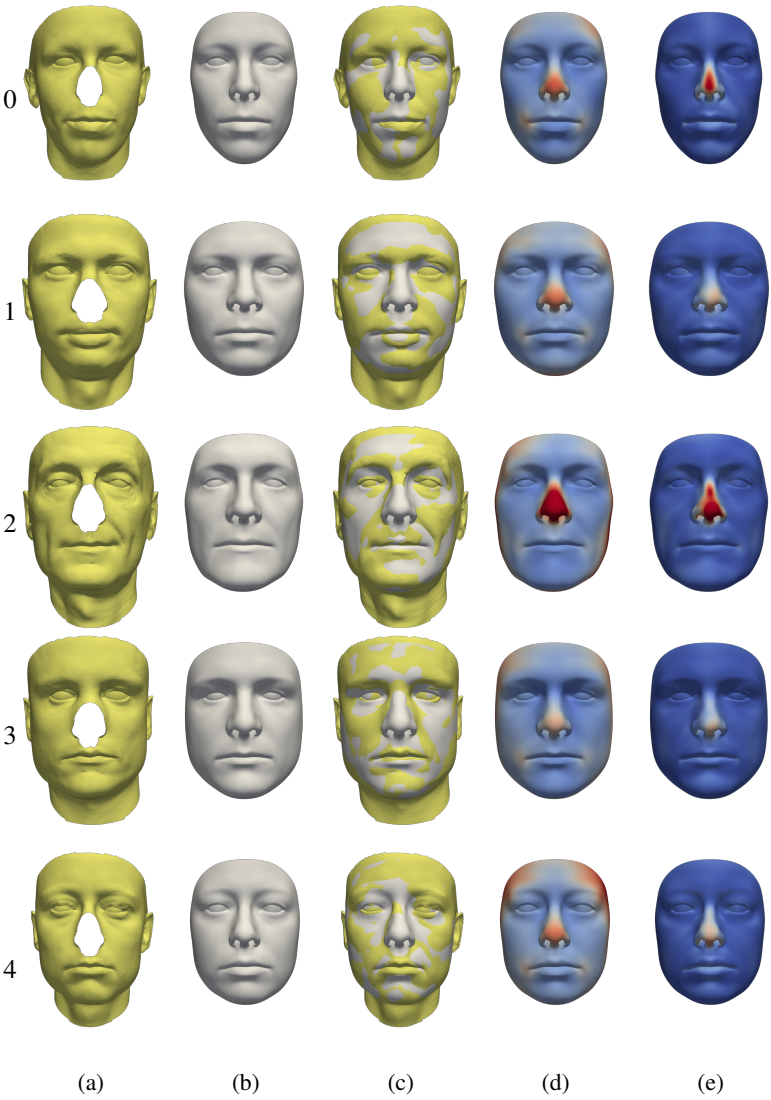


Figure 5.8: Face registration experiment. (a) target surface, (b) MAP estimate, (c) overlay of target and MAP, (d) visualization of the correspondence uncertainty (sum of point variance), (e) visualization of the posterior uncertainty in the normal direction (color-map, red: large variance, blue: no variance).

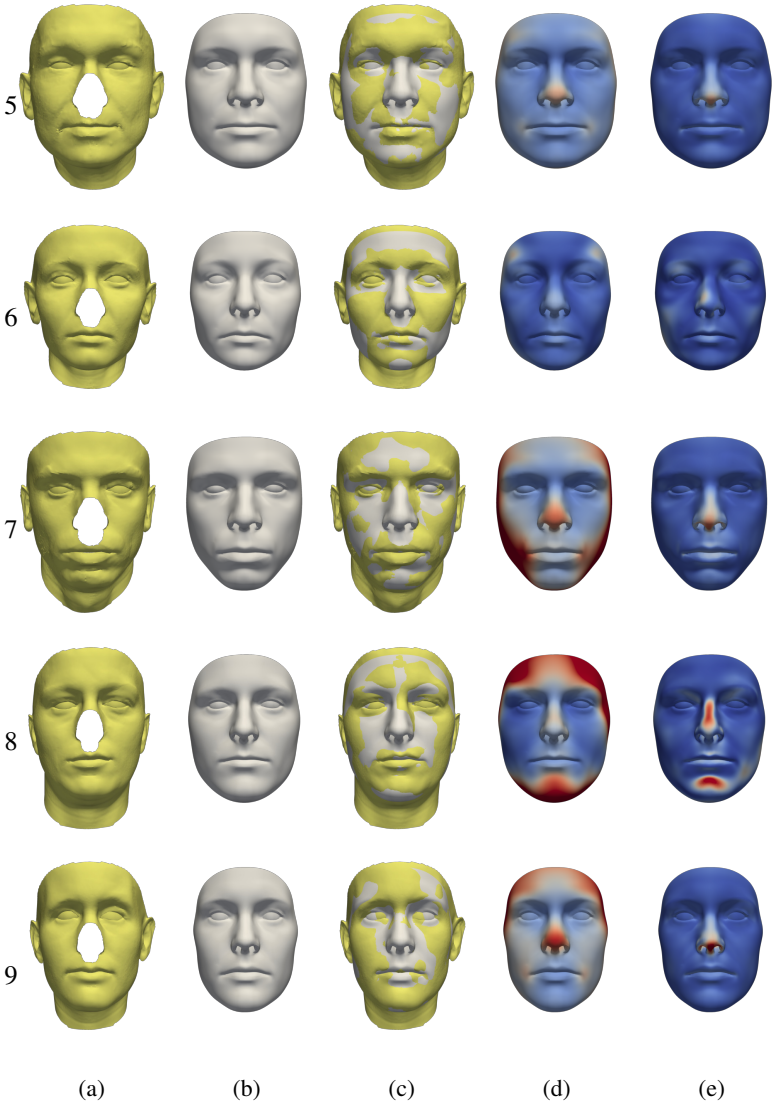


Figure 5.9: Face registration experiment. (a) target surface, (b) MAP estimate, (c) overlay of target and MAP, (d) visualization of the correspondence uncertainty (sum of point variance), (e) visualization of the posterior uncertainty in the normal direction (color-map, red: large variance, blue: no variance).

The collective average likelihood is a valid likelihood function as a single likelihood is used in each iteration which takes the average distance between the surfaces into account.

In [61] we experimented with combining the collective average distance likelihood with the Hausdorff distance likelihood to obtain a good average registration while also avoiding individual points being far off:

$$P(\Gamma_T|\boldsymbol{\alpha}) \propto \mathcal{N}(\mathbb{d}_{CL}(\Gamma_T, \Gamma[\boldsymbol{\alpha}]); 0, \sigma_{CL}^2) \cdot Exp(\mathbb{d}_H(\Gamma_T, \Gamma[\boldsymbol{\alpha}]); \lambda_H). \quad (5.3)$$

However, from a practical side, it is very difficult to adjust the hyper-parameters of the two distributions to get the desired result. Additionally, we do not end up with a valid pdf function when multiplying the two pdfs. Therefore, as an alternative likelihood function, we choose to simply scale the independent point evaluator based on the number of corresponding points n_c in each iteration:

$$P(\Gamma_T|\boldsymbol{\alpha}) \propto \frac{1}{n_c} \prod_{i=1}^{n_c} \mathcal{N}(\mathbb{d}_{l^2}(\Gamma_T^i, \Gamma[\boldsymbol{\alpha}]^i); 0, \sigma_{l^2}^2). \quad (5.4)$$

We find that the hyper-parameter is easy to set and can be kept constant over the 10 faces we have as targets. Most importantly is that the acceptance rate is kept at 20-30%, whereas it drops to near 0% when using the collective average distance for some targets. In Fig. 5.8 and Fig. 5.9, we show the correspondence uncertainty from the posterior registration distribution. Note however that this likelihood function is also not a valid pdf due to the change in evaluated points in each iteration.

With P-GiNGR we infer a larger correspondence uncertainty in the outer region. However, as the surface is observed, the uncertainty is low in the direction of the face surface normals but high within the surface. This is because there is no anatomical distinctive feature to predict the correspondence more precisely. Higher uncertainty is also inferred on the nose, where data is missing. In contrast to the outer region, the uncertainty in the direction of the normal of the reconstructed nose surface is large for most of the targets. This shows that uncertainty visualization could possibly be used to detect missing areas. As quantitative results, we obtain an average Euclidean distance

of $0.55\,\mathrm{mm}$ and an average Hausdorff distance of $5.79\,\mathrm{mm}$ over all the targets. In contrast, with the collective average likelihood, we get $0.73\,\mathrm{mm}$ and $6.10\,\mathrm{mm}$.

5.3.1 Discussion

In this section, we highlighted one of the difficulties when using P-GiNGR, namely the design of likelihood functions. We showed an alternative likelihood function that can be adapted to work for target surfaces that are not a strict subset of the PDM or the target. Unfortunately, this likelihood function is merely a hacky solution that does not provide us with a theoretically sound likelihood function. The design of a more general likelihood function that works across a large range of different domains is therefore up for future research. In Section 5.6 we will be looking more into designing application-specific likelihood functions. But in the following section, we will first explore how we can directly use P-GiNGR to construct PDMs from partial data.

5.4 Shape Priors (PDMs) from Pieces

In this section, we show that P-GiNGR can be used as a principled way to go from *only* partial data observations to a PDM, without the need for multiple intermediate steps. The experiments in this section have previously been published in [60, 61]. From Bayesian statistics, we know that there are three ways to handle missing data [32]:

- Discard observations with any missing values.
- Rely on the learning algorithm to handle missing values in the training phase.
- Impute all missing values (complete the data-matrix before using any learning algorithm).

From [26] we know that if we choose a reasonable missing-data model, the imputed dataset is likely to provide a more accurate estimate than a strategy which discards the data with missing values. In our setting where we are only working with partial data, it is not even a possibility to discard the data. For data imputation, a range of different strategies exist. The simplest strategy is to impute the missing values with the mean or the median of the non-missing values for that feature. This approach assumes that the missing values were known in the complete-data and will bias the variance of the dataset towards zero. One would, in the worst case, end up only learning the mean shape of the dataset and nothing about the shape variation. A better option for imputation is to perform regression based on the remaining dataset. This usually overestimates correlations which are then reflected by the model, as no uncertainty is given to the missing part. Instead of a single imputation, we can make use of multiple imputations. Multiple imputation was initially introduced in [81] to fill out non-responses in surveys. If the data is arbitrarily missing, a MCMC method can be used to create multiple imputations by simulating draws from a Bayesian predictive distribution given the partial data [101]. An overview of the multiple imputation method is visualized in Fig. 5.10. We use P-GiNGR to compute the posterior distribution of registrations for each partially observed

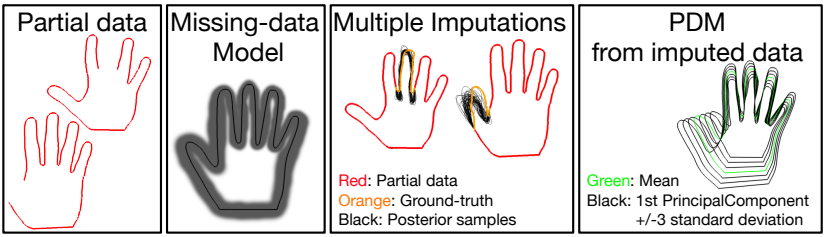


Figure 5.10: Overview of our method. Given the partial data (red), the missing-data model (GPMM) is used to draw samples from the posterior distribution over registrations to impute the data. Finally, multiple imputations from each partial data item are used to compute the PDM.

target surface. The posterior distribution contains the distribution of possible completions under our missing data model. These completions are in the Bayesian setting referred to as multiple imputations. From the posterior distribution $P(\boldsymbol{\alpha}|\Gamma_T)$, we randomly sample I imputations, such that the inferred data-matrix will be $X \in \mathbb{R}^{dn \times NI}$, with N being the number of target surfaces. By including multiple imputations into the PDM, we are able to take the uncertainty of the imputation under our missing-data model into account. The PDM can then be computed using standard PCA on the complete data-matrix.

5.4.1 Experiments

In the following experiments, we first demonstrate our method on a synthetic dataset of 2D hands and then a 3D clinical dataset of femurs with artificially removed parts. As we have complete hands and femurs, we can compare our method with the ground-truth PDMs. For a model to be useful, it has to be specific - i.e. only able to represent valid shapes. Alternatively, a model might be very general, making it able to generalize to any type of new data, but this at the cost of explainability. To evaluate PDMs, we use three common measurements: specificity S(PDM) (evaluate if the model only generates instances that are similar to those in the training set), generalization G(PDM) (the ability to describe instances outside of the training set) and compactness

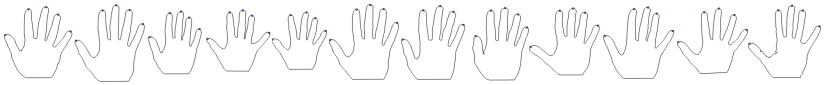


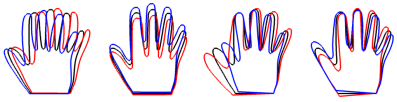
Figure 5.11: The dataset of hands with marked fingertips.

C(PDM) (a model's ability to use a minimal set of parameters to represent the data variability) as described in [89]. From these measures, specificity and generalization are the most important measures when using the model as prior information in other learning algorithms. We want the model to stay within the shape space, but also to be able to explain new data from the same shape space. For most of the experiments we do not plot the compactness as this is less important and very similar for all of the models.

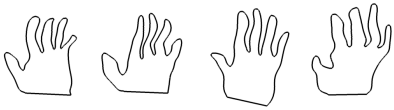
The target surfaces in all the experiments have been initially landmark aligned to a reference surface. For the synthetic experiments where parts are cut from the target surfaces, we still keep the global alignment to avoid factoring pose difference into the model comparison.

5.4.1.1 2D Synthetic Hand Experiment

For the 2D hand experiment, we make use of 12 synthetic 2D hand surfaces as the targets, visualized in Fig. 5.11. An additional hand surface is used as the reference surface to construct a hand-GPMM with a mixture of Gaussian kernel as shown in Fig. 5.12. The computations of the generalization measures for the hand models are computed by a "leave-one-out" approach. For the specificity measure, we use 1000 random samples.



(a) Mean, +/-3 (red/blue) standard deviation of the 4 first GPMM components.



(b) Random samples from the GPMM.

Figure 5.12: Visualization of the analytically defined 2D hand-GPMM.

(a) 5 (b) 10 (c) 15 (d) 20 (e) 30 (f) 40 (g) 50 (h) 60 (i) 70 (j) 80

Figure 5.13: Hand example of partial data. The caption lists how much of each shape has been removed (5-80%) starting from the fingertip of the thumb.

Missing Finger Experiments. We clip the hands dataset starting from a landmark on the top of each finger and cut away an increasing amount (5-90%), as visualized in Fig. 5.13. We perform 2 different experiments with this setup. In the first experiment, a random finger is increasingly cut away from each of the hands. In the second experiment, all 12 hands are increasingly missing the thumb. In Fig. 5.14 we show the model measures of the random missing finger experiment and in Fig. 5.15 we show the model measures of the missing thumb experiment. From the figures we see that both experiments follow the same trend. We see that the models computed from multiple imputations are able to generalize much better than when using the MAP or mean solutions. Note, that the curve flattens for the MAP and the mean experiments as only 12 surfaces are used to compute these models, whereas we use 100 imputations for each of the 12 target surfaces in the multiple imputation experiment. If less than 30% of the surfaces are missing, we also obtain a better specificity measure compared to the MAP and mean solutions. Finally, we see that the multiple imputations are able to maintain much more of the total variance within the model when large parts of the targets are missing.

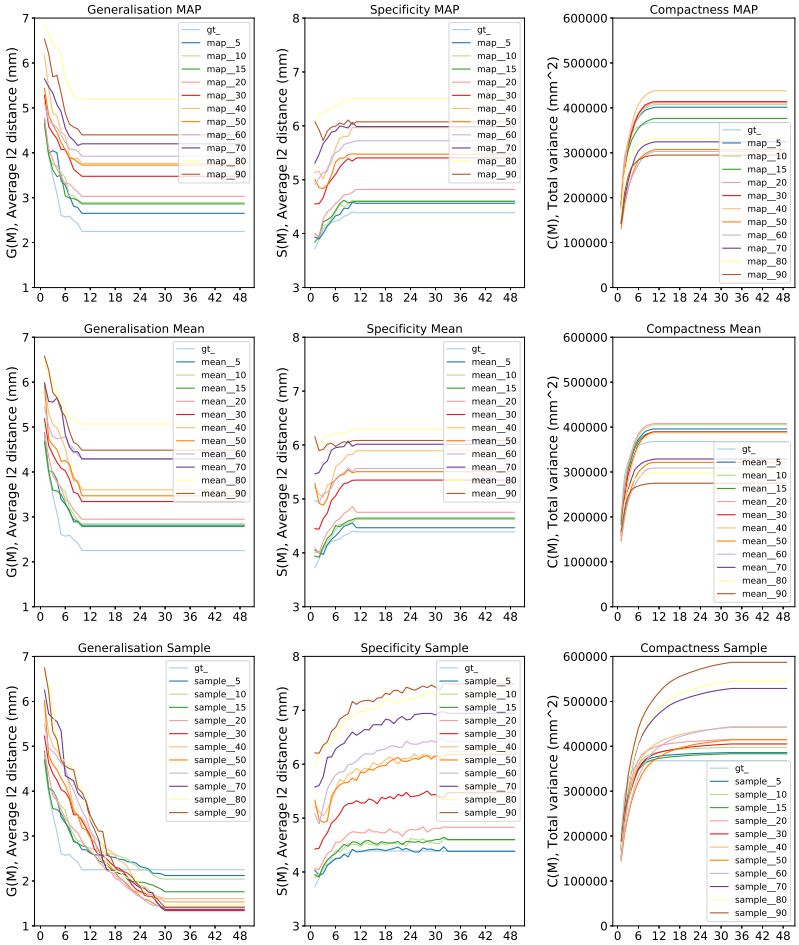


Figure 5.14: Hand model measures from the random missing finger experiment. The left column shows the results for models computed with the MAP from each posterior distribution. The middle column is computed using the mean surface from each posterior distribution. The right column shows the results using 100 imputations from each of the 12 posterior distribution. The last-mentioned model is clearly superior.

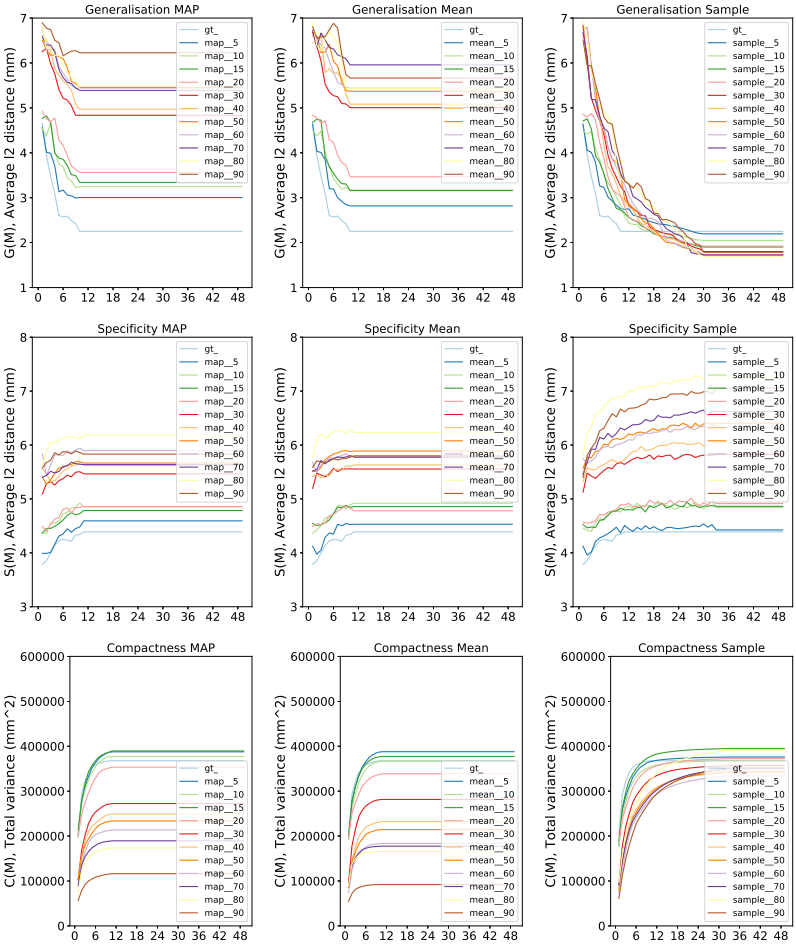


Figure 5.15: Hand model measures from the missing thumb experiment. The left column shows the results for models computed with the MAP from each posterior distribution. The middle column is computed using the mean surface from each posterior distribution. The right column shows the results using 100 imputations from each of the 12 posterior distribution. The last-mentioned model is clearly superior.

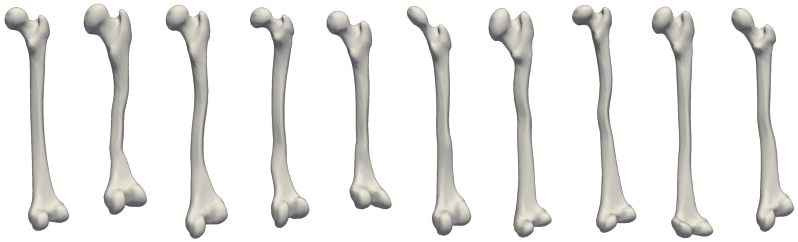


Figure 5.16: Reference (left) and random samples from the femur-GPMM.

5.4.1.2 3D Femur Experiment with Ground-Truth

We use a Gaussian kernel for the femur-GPMM. The reference surface and random samples from the femur-GPMM are shown in Fig. 5.16. We use the publicly available dataset of 50 complete femurs as also used in Section 5.1 and Section 5.2, where the surfaces are extracted from computed tomography images[6]. We choose ten complete femur surfaces as our training data, from where we clip a varying amount (5-25%) at different landmark locations. For eight of the bones, this clipping was done in a single location. For the remaining two, we spread the removal over four locations. The partial dataset is visualized in Fig. 5.17. Similar to before we build PDMs from multiple imputations, the MAP solution and the mean and compare them additionally to the ground-truth PDM. For the specificity computation, the ground-truth registrations are used (these are also used to build the ground-truth PDM). For the generalization, we use 40 registered surfaces which are not included in the dataset of the PDMs. In Fig. 5.18 we see an example of a target shape (red) with multiple different imputations. We see how both the mean and the MAP solutions fail to perfectly predict the true shape. From the random samples we see that there is a broad distribution of possible imputations for the missing part. In Fig. 5.19 we compare PDMs created from the MAP, mean and multiple imputations. In the case of multiple imputations, we compare models created with a different number of samples. From the given example,

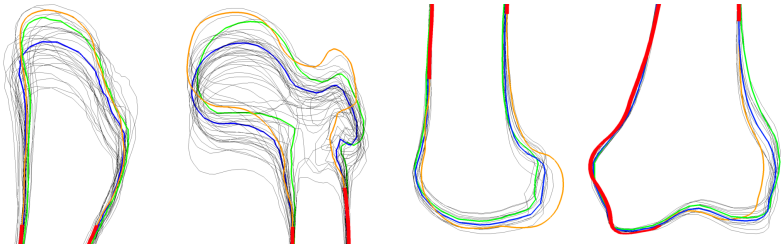
[6] Available at the SICAS Medical Image Repository [46]

Figure 5.17: The partial femur data that is used to construct the PDM. Numbered 1 to 10 from left to right.

we see that by using 5 or more imputations, we are able to create PDMs that generalize better, than by using the MAP or the mean solutions, while also maintaining better specificity.

5.4.2 Related Work

PDMs are usually built from complete and healthy shapes in point-to-point correspondence. However, in the medical domain, either the data is captured because there is a pathology, or it is scanned only partially to capture the essential part of the structure while minimizing the radiation danger. Hence, usually



(a) Target femur 1. (b) Target femur 5.

Figure 5.18: Posterior distribution visualizations of the two first partial femurs from Fig. 5.17. Red: Partial surface, Orange: Ground-truth surface, Green: MAP, Blue: Mean, Black: Posterior samples. It can be seen that the posterior samples better cover the ground-truth than both, the MAP and mean shape.

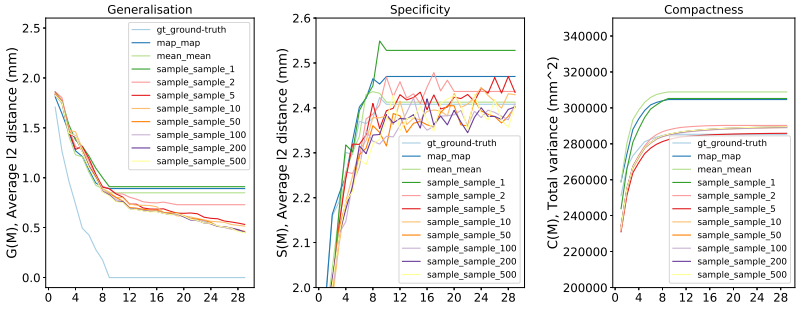


Figure 5.19: Femur model measures. Multiple imputations lead to a better generalization, while also keeping a better specificity than models built from the MAP or mean imputations.

only a part of the healthy anatomy is observed. Several methods address the construction of PDMs where some of the training shapes are partial. In [43], each landmark is assigned a probability of being an outlier. This is used to compute a mean of the dataset where landmarks with lower probability of being an outlier have more influence on the mean shape. In [51], the training surfaces are divided into patches and each patch is assigned a probability of being an outlier. The outlier detection is performed with PCOut [24] which identifies samples that do not fit well into the distribution. Probabilistic PCA [90] is then used to iteratively build the PDM and replace the outlier parts with healthy parts. In [30], the shape model is computed using robust PCA (RPCA) to be able to marginally improve the model through partial data. In [56], they extend this idea to have a probability of being an outlier assigned to each landmark in a shape. The authors then extend their method to a robust kernel SSM, to have a non-linear model for better compactness [57]. All of the aforementioned methods use off-the-shelf registration methods and focus on building the models robustly from the noisy registration results. Often the registrations shrink substantially where parts are missing, to which the learning algorithm then has to be robust. The majority of the methods decompose the data matrix into a low-rank matrix containing correct data and a sparse matrix with the corrupted data. These matrices are mainly found via convex optimization and

usually require non-corrupted data to be present in the dataset.

In comparison, with P-GiNGR, we do not need to detect if parts are outliers or to assign weights for each landmark in a shape.

5.4.3 Discussion

In this section, we showed a principled method to create PDMs from partial data observations. Our method uses the idea of multiple imputation from Bayesian statistics for point-to-point registration of partial data. With this, we can create multiple imputations of a single partial data observation with the use of P-GiNGR, to span the data-matrix. We show how this improves model specificity and generalization. Using GPMMs, we can not only design the model in an analytical way to encode smoothness or symmetry but also iterate the model building process. Once data has been registered, we can create multiple imputations to compute a statistical kernel. The statistical kernel can then be augmented and the partial data can be registered again. In the second iteration, the imputations will be partly based on a kernel learned from data, and therefore be more reliable than the imputations based purely on an analytical defined kernel. In the following section, we will show how the general principles introduced in this section can be used to create a PDM from hybrid imaging with complementary overlapping regions.

5.5 PDMs from Overlapping Data

In this section, we show an application of how a molar-tooth-PDM can be created from partial data. This setting is different from the previous section, as we are working with a hybrid of imaging with complementary and overlapping areas. In Fig. 5.20 we show the raw surfaces that we use to build the PDM from. Ideally, the tooth-PDM would be created from scans of full teeth in high quality. However, this requires the teeth to be pulled out and is therefore not a viable solution in creating a model with a broad population distribution. We base the reference surface on a scan of an extracted tooth. We manually edit the reference surface to remove patient-specific deformations and discretize it with ~ 25.000 vertices, such that a higher detail level is found in the crown level of the tooth. To capture the tooth root shape, we use surfaces extracted from CBCT images with a U-Net-based segmentation network [80]. The teeth extracted from the CBCT data are very noisy and with very little visible details in the crown area. The idea is therefore to append the high frequency crown details from the intraoral scans to the CBCT data. It should be noted that the CBCT data and the intraoral scans are from different patients. Because of that we cannot manually fuse the two data sources to create the dataset that

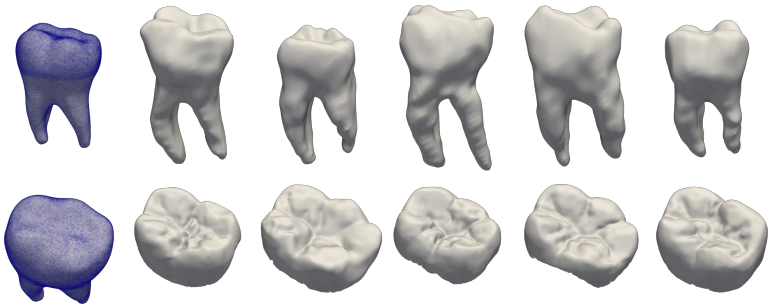


Figure 5.20: Visualization of the molar tooth training data. The reference surface is depicted on the left side with the surface wireframe. The remaining figures are the (top) segmented CBCT surfaces (bottom) intra-oral scans (bottom). Note that the intra-oral scans and the Cone-beam computed tomography (CBCT) data are obtained from different patients.

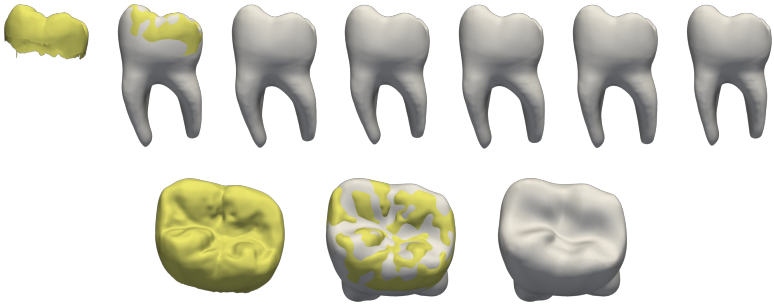


Figure 5.21: Completions of intraoral scans. The intraoral scan is shown in yellow. In gray we show different samples from the P-GiNGR posterior. In the bottom row, we zoom in on the crown area and see that many of the larger structures are captured in the registration, but many high-frequency details are still smoothed out.

we need. In total, we make use of 30 complete teeth extracted from CBCT and 200 intra-oral scans of lower-molar teeth. The procedure for building the tooth-PDM from the hybrid dataset is:

1. Register the CBCT data using P-GiNGR.
2. Create a statistical kernel from the registered CBCT data.
3. Augment the PDM in the crown region.
4. Register the intraoral scans using P-GiNGR.
5. Create the tooth-PDM from multiple imputations.

For step 1, we use the dot-product kernel together with a Gaussian kernel (Eq. (2.8)) with a large β value to avoid fitting all the noisy bumps as seen in the extracted CBCT data in Fig. 5.20. In step 3 we only augment the PDM in the crown region by the use of a change-point kernel, which was also used in [27]. The augmented kernel is a multi-scale of different Gaussian kernels and the statistical kernel learned from the CBCT data. In Fig. 5.21 we show an example of the registration performed in step 4, where a single intraoral

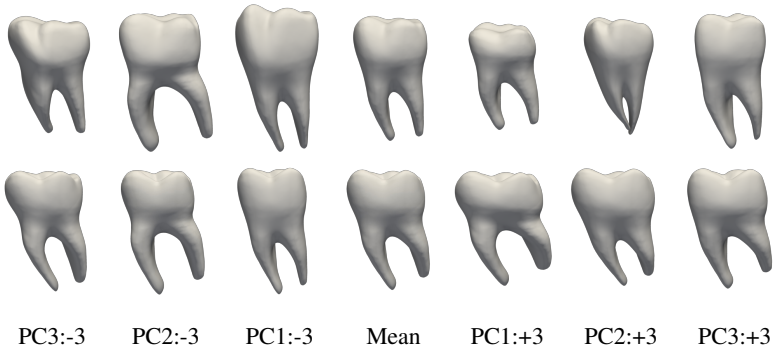


Figure 5.22: Comparison of the first 3 principal components (PC) from the PDMs created from (top) CBCT data only, (bottom) mixture of CBCT data and intraoral scans. -/+3 denotes 3 standard deviation.

target has many different completions. In Fig. 5.22 we compare the major directions of variation between the final tooth-PDM and the PDM created in step 2 from only using the CBCT data. The expectation is that the first principle components will stay almost the same, while the intraoral data mainly add deformation information in the crown area. We see that the distribution has changed slightly based on the larger dataset, but the root deformations from the CBCT model are largely kept the same. As we have no complete combined data available, it is difficult to quantitatively assess the model using the model metrics (specificity, generalization, compactness). We are therefore only qualitatively evaluating samples from the tooth-PDM. In Fig. 5.23 we see the main difference between the model created in step 2 and step 5, namely the detailed crown deformations.

The majority of literature focuses on creating PDMs from complete data. The most similar approach to what we have shown in this section is [72]. They create a face model from coarse full head scans and highly detailed face scans. They show how this can either be done using a regression-based solution to complete the partial data (in their case the face scans) or by directly smoothing two kernels using the GPMM framework. The main problem with this method is the need for smoothing. If using a regression-based solution, one needs to

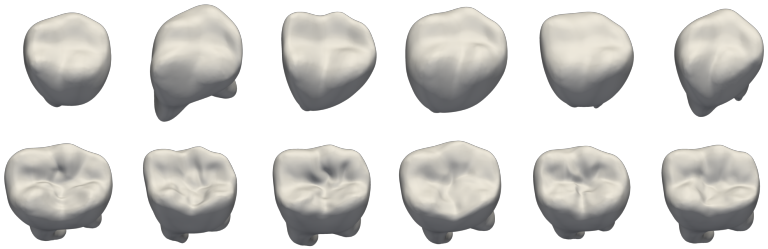


Figure 5.23: Comparison of random samples from the PDMs created from (top) CBCT data only, (bottom) mixture of CBCT data and intraoral scans. Notice how the samples from the CBCT are completely smoothed out, while the addition of the intraoral scans adds details to the crown region.

smoothen the surfaces in the overlapping region between the high detailed partial data and the coarse data. The same goes for smoothing the kernels, which we found can be problematic, especially if the mean deformations are different from one another as can be observed in Fig. 5.22 when comparing the two mean surfaces.

5.5.1 Discussion

In this section, we showed an additional use case of the P-GiNGR framework for creating PDMs from a hybrid of overlapping imaging with different properties. While further evaluation still needs to be carried out to fully validate the statistical properties of the combined model, the qualitative results look promising and the method does not require any hand-tuning to smoothly blend between different kernels. However, as shown in Fig. 5.21, there is still room for improvements in the hyper-parameter settings of P-GiNGR in order to get all the high-frequency intricate details from the intraoral scans. One possible solution to overcome this, would be a few deterministic ICP iterations with a very low noise assumption, or alternatively, directly projecting the model onto the target. In the following section, we will go a step further and show how PDMs can be created from independent data with no overlapping regions.

5.6 PDMs from Disjoint Data

As a final application of the probabilistic registration framework, we show how to create a combined face-skull-PDM from two disjoint data sources, namely face surface scans and skull surfaces from Magnetic Resonance Imagings (MRIs). For this application, we are working with data that is already in point-correspondence and from which an independent face-PDM and an independent skull-PDM has been created. The main idea behind this experiment is to show alternative usages of the probabilistic framework other than establishing point-correspondence in a dataset for the creation of PDMs. The mean of our final combined face-skull model is illustrated in Fig. 5.24. In this application, we will not be using CP but instead a standard random-walk proposal. The emphasis is instead on the careful design of application specific likelihood models.

The most obvious use for a combined face-skull-PDM is facial reconstruction, which is the process of recreating the face of an unknown individual solely from their skeletal remains [97]. Reconstructing the face shape from a skull is an ill-posed problem. It is commonly known that the facial reconstruction process is highly artistic and different forensic artists might produce different face predictions [98]. One reason for the problem to be ill-posed is that many features of a shape are not due to the skull itself, but due to genetics

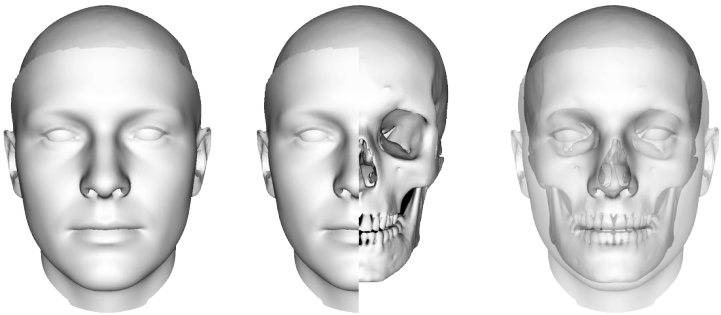


Figure 5.24: Illustration of the face-skull model with the face fully visible, one side visible and fully visible with an opacity of 70%.

and environmental factors such as diet and ageing. As a result, for a given skull there exists a whole distribution of possible faces. A combined model would likewise not only provide us with a single best prediction, but a full distribution of possible faces given a skull. In Section 5.6.3 we will discuss alternative methods that have been used for facial reconstruction. In [58] we showed how the combined model can be used for craniofacial superimposition within forensics. This is the task of comparing a photo of a skull to a photo of a face to evaluate their similarity [93]. We did so by ranking a database of faces given a skull. However, in this thesis we mainly focus on the model construction and on applications directly applicable to medical image analysis.

5.6.1 Method

We model the joint face-skull distribution and build a joint face-skull-PDM. Because it is difficult or even impossible to observe the joint face-skull distribution directly, we construct it from the conditional distribution over face shapes given skull shapes. Fig. 5.25 shows the general idea of combining the skull and the face model. The two models are independent and contain scans from different individuals. We compute the posterior distribution of faces given a skull using MH. A sparse set of distance constraints between the surface of the skull and the surface of the sampled faces is everything the method needs. The distance constraints are known as tissue-depth markers within facial reconstruction and represent the depth of the tissue in between the skull and the face surface such as fat, muscle fibers and cartilage. We generate face-skull pairs by sampling from the posterior face distribution found for each skull. Then, we model the joint distribution according to the generated data.

5.6.1.1 Training Data

We use three sources of data; the publicly available BFM, a skull-PDM and tissue marker data as defined in the publicly available T-Table [87, 88] and shown on a skull shape in Fig. 5.26. All three sources have to our knowledge

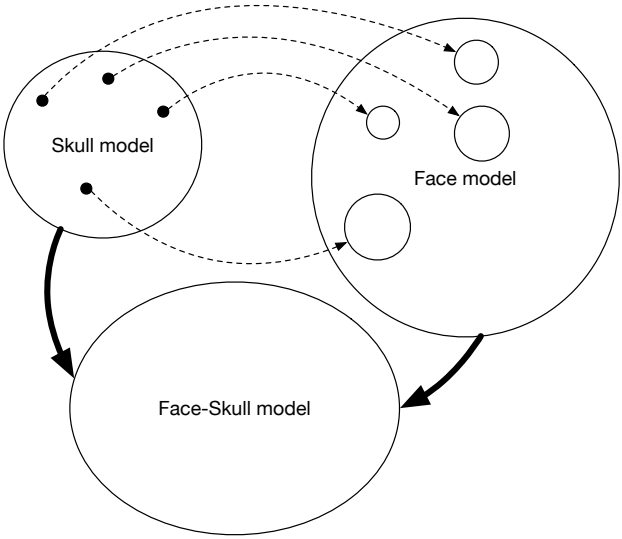
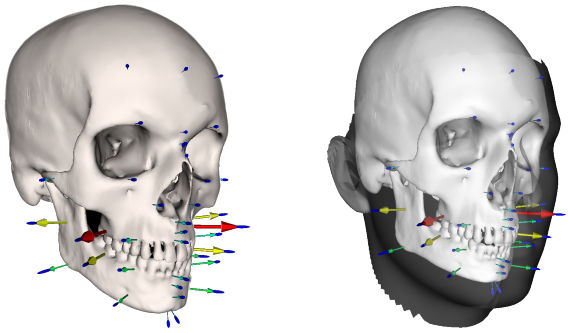


Figure 5.25: Combining the models. We map each skull in the skull-PDM to a distribution of face shapes in the face model. The face-skull model is a joint distribution of the individual skulls from the skull-PDM and their individual face distributions.

no overlap in the individuals from which the data was obtained.

5.6.1.2 Tissue-depth Gaussian assumption

The Gaussian distribution is assumed for each of the individual tissue-depth markers. While the distribution of tissue-depth over age, weight and gender differences might follow a more complex distribution, tissue-depth studies do only provide mean and standard deviation and not the individual measurements. However, the nature of our method would allow easy use of more intricate distributions such as Gaussian mixture models if one has raw data measures available. As our raw skull data does not contain any metadata, we use the T-Table which uses the law of large numbers to combine measures from a variety of age, Body Mass Index (BMI) and gender groups. If skull



(a) (b)

Figure 5.26: Tissue-depth markers placed on a skull. The landmark distributions are shown at the end of the vectors where one standard deviation is visualized. (b) shows a transparent face shape fit to the tissue-depth measures in (a).

metadata would be available, alternative data tables could be used to limit the variance and update the mean, as found in [18].

5.6.1.3 Simulating the Joint Face-Skull Distribution

We model the joint face-skull distribution as a multivariate normal distribution over face and skull shape.

$$\Gamma_F, \Gamma_S \sim \mathcal{N}(\mu_F, \Sigma_F), \mathcal{N}(\mu_S, \Sigma_S) \tag{5.5}$$

As we cannot observe the full joint face-skull distribution directly, we construct it from the skull shapes and tissue depth marker information. Probability theory allows us to write the joint distribution as a product of a conditional distribution and a prior. We write the joint probability of observing a specific face-skull pair as the distribution of face shapes Γ_F given the skull shape Γ_S and the prior of the skull shape.

$$P(\Gamma_F, \Gamma_S) = P(\Gamma_F|\Gamma_S)P(\Gamma_S) \tag{5.6}$$

The skull-PDM provides the prior $P(\Gamma_S)$. The posterior distribution over plausible face shapes, given a skull shape, consists of all the face shapes that

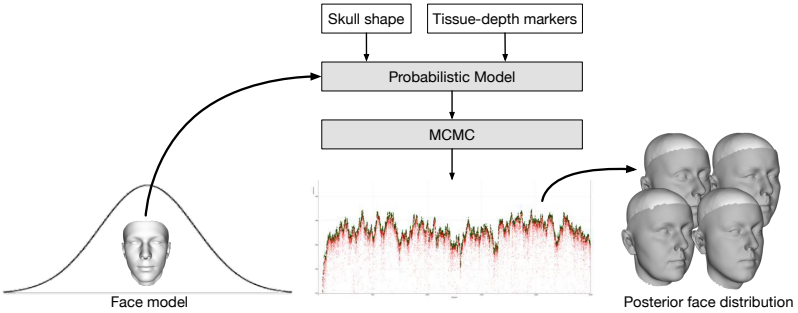


Figure 5.27: Illustration of the MCMC fitting procedure.

fulfill the given tissue depth constraints. Just as in P-GiNGR, we approximate the posterior distribution with MH. The posterior distribution we are interested in is the face shapes given an observed skull shape. Having a posterior distribution over faces given a skull allows us to construct the joint distribution. According to Eq. (5.6), we first sample a random shape from the skull-PDM. Then we approximate the posterior face distribution given the skull. To create a face-skull-PDM, we fit a Gaussian distribution to the skull-PDM samples and their corresponding posterior face shape samples. Fig. 5.27 gives an overview of how we approximate the posterior. Fig. 5.28 shows how we build the joint face-skull-PDM from skull samples and posterior face shape samples. Just as in Section 5.4 we use multiple imputations to complete the face-skull pairs given a skull instance.

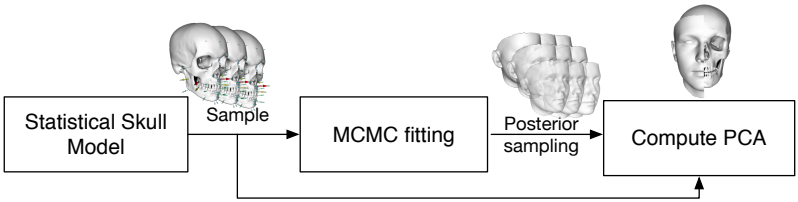


Figure 5.28: Pipeline describing how to compute the face-skull-PDM by combining skull samples with the posterior face distributions found with MCMC.

5.6.1.4 Likelihood Model

In this section, we describe how the probability distribution over possible face shapes is modeled (determined by the model parameter vector $\boldsymbol{\theta}$ as introduced in Section 2.3) to fit a given skull. We model the likelihood as:

$$P(\boldsymbol{\theta}|D^{tvi}, D^{sym}, c, D^{cs}), \tag{5.7}$$

where D^{tvi}, D^{sym}, c, D^{cs} are four different measurements that are derived from the given skull and the tissue depth markers. We will explain them in detail below. Using Bayes theorem we can write the conditional distribution as a product of its prior probability and a likelihood term. As with P-GiNGR we are only interested in comparing samples, so the evidence term can be ignored. By assuming independence between the individual likelihoods, we write it as the product:

$$\begin{aligned} P(\boldsymbol{\theta}|D^{tvi}, D^{sym}, c, D^{cs}) \propto \\ P(\boldsymbol{\theta})P_{tvi}(D^{tvi}|\boldsymbol{\theta})P_{tvs}(D^{sym}|\boldsymbol{\theta})P_{fs}(c|\boldsymbol{\theta})P_{cs}(D^{cs}|\boldsymbol{\theta}). \end{aligned} \tag{5.8}$$

With the face shape prior $P(\boldsymbol{\theta})$ we penalize unlikely shapes. We enforce the tissue-depth distribution with the tissue-vector intersection likelihood P_{tvi}. Tissue-vector symmetry is encouraged by P_{tvs} and face-skull intersection is discouraged with P_{fs}. Furthermore, we enforce correspondence in a single point with the likelihood P_{cs}.

For a given skull shape Γ_S we evaluate the probability of a face shape from its face model parameters $\boldsymbol{\theta}$. We place n_{lm} landmarks $lm_i^S \in \mathbb{R}^3, i = 1...n_{lm}$ on Γ_S. For each landmark we have a tissue-depth vector $\boldsymbol{v}_i^S \in \mathbb{R}^3$ with direction towards the face surface (see Fig. 5.26), which is obtained from [88]. We define the landmark $lm_i^F[\boldsymbol{\theta}] \in \mathbb{R}^3$ as the intersection point of $\boldsymbol{v}_i^S$ with the face model $\Gamma_F[\boldsymbol{\theta}]$.

Tissue-vector intersection We evaluate where the n_{lm} vectors $\boldsymbol{v}_i^S$ intersect with the face model at a given $\boldsymbol{\theta}$ by the distance $d_i \in \mathbb{R}$ for each vector. d_i is found from the skull landmark lm_i^S to the intersection landmark $lm_i^F(\boldsymbol{\theta})$ on the face shape as $d_i(\boldsymbol{\theta}, \Gamma_S) = \|lm_i^F[\boldsymbol{\theta}] - lm_i^S\|$, with D^{tvi} denoting the

d_i distances for all the tissue-depth vectors. The probability over all the n_{lm} points is the product of each individual likelihood term:

$$P_{tvi}(D^{tvi}|\boldsymbol{\theta}) = \prod_{i=1}^{n_{lm}} \mathcal{N}(d_i(\boldsymbol{\theta}, \Gamma_S), \sigma_i^2) \tag{5.9}$$

with σ_i^2 being the variance of the individual tissue-depth vector from the T-Table [87, 88].

Tissue-vector symmetry We extend the evaluation of tissue-vector intersection by evaluating how similarly mirroring vectors intersect the face surface. Each landmark lm_i^S not placed in the center of the head has a mirror landmark. There is a total of n_{center} points placed in the center of the head. We define $lm_{-i}^S \in \mathbb{R}^3$ to be the mirror point of lm_i^S and $\boldsymbol{v}_{-i}^S \in \mathbb{R}^3$ to be the mirror vector of $\boldsymbol{v}_i^S$ around the sagittal plane. Let $d_{-i}(\boldsymbol{\theta}, \Gamma_S)$ be the tissue-depth vector intersection mirroring $d_i(\boldsymbol{\theta}, \Gamma_S)$, then $d_i^{sym}(\boldsymbol{\theta}, \Gamma_S) = \|d_i(\boldsymbol{\theta}, \Gamma_S) - d_{-i}(\boldsymbol{\theta}, \Gamma_S)\|$ is the difference between a tissue-vector intersection depth and its mirror vector intersection depth. D^{sym} is the tissue-depth intersection difference. The combined likelihood for all the mirroring points is:

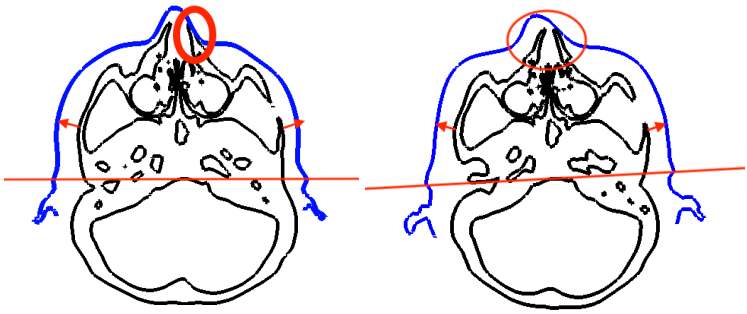
$$P_{tvs}(D^{sym}|\boldsymbol{\theta}) = \prod_{i=1}^{(n_{lm}-n_{center})/2} \mathcal{N}(d_i^{sym}(\boldsymbol{\theta}, \Gamma_S), 1) \tag{5.10}$$

The likelihood will return a high probability if the symmetry vectors intersect the face surface at the same distance. Fig. 5.29a shows a sample where the tissue-vector symmetry likelihood was not used.

Face in skull detection The intersection likelihood in Eq. (5.11) penalizes the number of points $c(\boldsymbol{\theta})$ of the face model which intersect the skull shape Γ_S. By choosing λ large, faces intersecting with the skull region become very unlikely.

$$P_{fs}(c|\boldsymbol{\theta}) \sim Exp(\lambda) \tag{5.11}$$

Point-correspondence In our implementation, we have made use of a single point-correspondence instance. The point is placed in the center of the



(a) (b)

Figure 5.29: Face-skull alignment problem from missing likelihood terms. (a) Missing tissue-vector symmetry likelihood. The center of the face and the skull do not align. This is especially visible at the nose which is shifted to the side. The illustrated tissue-vectors have the same length and are not intersecting the tissue at the same distance on both sides of the face. (b) Missing point-correspondence likelihood. The face is rotated slightly which can be seen from the red line which should be horizontal as in (a). The nose of the skull is also not aligned at the center of the nose in the face, similar to the problem without the point-correspondence.

face just below the nose. The correspondence point on the skull is placed just below the anterior nasal spine. The likelihood is necessary in order to avoid misalignment at the nose region as shown in Fig. 5.29b. The general point-correspondence likelihood consists of n_{center} points where the exact face intersection position on the face model surface has been defined as $lm_j^{F_{ideal}} \in \mathbb{R}^3$, with $j = 1...m$. The distance between the actual intersection point $p_i^F(\boldsymbol{\theta}, \Gamma_S)$ and the ideal intersection point $lm_j^{F_{ideal}}$ is found from $d_i^{cs}(\boldsymbol{\theta}, \Gamma_S) = \|lm_i^F(\boldsymbol{\theta}, \Gamma_S) - lm_j^{F_{ideal}}\|$, with D^{cs} denoting all the d_i^{cs} distances. The combined likelihood is then:

$$P_{cs}(D^{cs}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \mathcal{N}(d_i^{cs}(\boldsymbol{\theta}, \Gamma_S), 1) \tag{5.12}$$

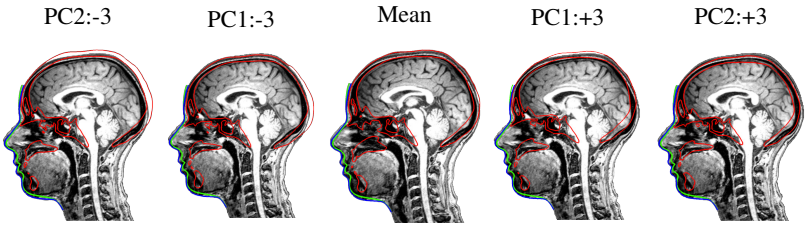


Figure 5.30: Mean and the first two principal components of the face-skull model conditioned on a face in an MRI. Blue: ground-truth face, green: face-skull model conditioned on the face, red: skull from the face-skull model. Notice the skull from the face-skull model (red) in the back and on top of the head. The model is very good in describing the correct placement of the skull over the 2 components.

5.6.1.5 Approximating the Probabilistic Model

As in Chapter 4 we approximate the distribution Eq. (5.7) by sampling directly from the unnormalized posterior distribution with MH. For the proposal we choose multiscale independent random walk proposals for the shape, rotation and translation parameters. These individual proposals are combined into a mixture proposal for the full parameter vector

$$Q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = z_\alpha Q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}) + z_t Q(\boldsymbol{t}'|\boldsymbol{t}) + z_R Q(\boldsymbol{R}'|\boldsymbol{R}) \tag{5.13}$$

where z_α, z_t, z_R are weighting factors of the different proposal distributions with $z_\alpha + z_t + z_R = 1$. The standard deviations for each of the proposal distributions have been scaled to have a 20% acceptance rate.

5.6.2 Applications of a Face-Skull PDM

We obtain a distribution over plausible skull shapes by conditioning on a face shape. An immediate application for this is the segmentation of the skull from an MRI. Segmenting the skull in an MRI is a difficult task as the bone has the same intensity as air and consequently they cannot be distinguished. Therefore, any successful segmentation method necessarily needs to include a strong shape prior. In contrast, segmentation of the face from an MRI can

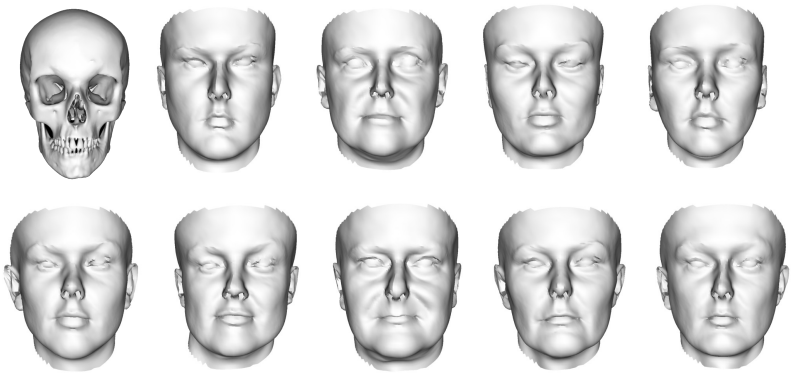


Figure 5.31: Random face samples from the posterior model conditioned on skull.

be done using simple threshold segmentation. Hence, by conditioning on the known face shape, we can obtain a shape model that only assigns a high probability to skulls which fit the given face. We have illustrated this approach in Fig. 5.30, where we show the mean and the first two principal components ($\pm$ 3 standard deviations). We observe that while the model still has the flexibility to adapt to the image, it already constrains the solution space well, and a model based segmentation algorithm for this task would be likely to benefit from such additional prior knowledge.

In Fig. 5.31 we qualitatively look at random samples of the posterior model of faces, when conditioning the combined model on a skull. We observe that the faces from the posterior face distribution look different from one another but are similar in shape.

5.6.3 Related Work

Forensic artists reconstruct the face surface by adding layers of muscle fibers and placing tissue depth markers on the skull whose identity is unknown. As already mentioned, and as highlighted in [98], face reconstruction is an ill-posed problem which is highly artistic. The majority of methods within auto-

matic forensic facial reconstruction share a general framework which consists of finding the mapping between skulls and faces [15]. Independent models of the face and the skull are constructed and the parameter mapping between the two models is found from a small test set. In more recent work [85], two individual PDMs were created from a large dataset of CT head scans and a linear mapping is found between the two. Similar methods have previously been attempted with smaller datasets from various sources [17, 20, 21, 48, 70, 100]. The model mapping approach maps a skull to a single face and therefore does not account for shape variance due to genetics and environmental factors such as diet and health. Several approaches, however, suggest changing the facial attributes in a final step to adjust for soft tissue variation due to age and BMI differences [20, 71, 84].

Alternative simple methods place tissue-depth pins on a dry skull and deform a face-mesh according to the mean tissue-depth measures [14].

More recent approaches fit several local shape models to a skull and glue them together into one joint face by using partial least squares regression to find a smooth surface transition [19, 39]. In [68] they are working with images of skulls and try to identify the matching face from a larger set of unnormalized facial photographs. The evaluation method here does also mimic craniofacial superimposition by only using pure classification techniques and without taking the soft-tissue variability into account.

As an alternative approach to traditional facial reconstruction, we have shown how a combined face-skull-PDM can be constructed from independent existing PDMs. By conditioning the combined model on the skull surface, we obtain a full distribution of possible faces instead of a single face as in traditional methods.

5.6.4 Discussion

In this section we presented an alternative usage of the probabilistic registration framework in order to combine two independent PDMs. The main novelty of this approach is that it does not require a paired dataset of head images, but can construct possible face-skull pairs using readily available statis-

tics about the soft tissue depth at individual landmarks. From a computational point of view, our method works by sampling possible faces from a posterior distribution using the MH algorithm. The generated faces for each skull are then used to build the joint face-skull model using multiple imputation. We showed how plausible face shapes can be generated which fit a given skull. This is a task of great interest in forensics. The distribution of possible faces that we obtain implies that current approaches of facial reconstruction, which only predict a single face for a given skull, exclude a lot of important information. As an application, we have shown the possibility to use our model for model-based segmentation of the skull in MRIs. We believe, that our approach of combining individual PDMs is very general and that the same principle can be used to join other shape models, in order to obtain flexible shape priors of different combined structures.

5.7 Conclusion

In this chapter, we have shown different applications for the probabilistic registration framework P-GiNGR. The common characteristic about all the experiments is the focus on partial data which is often the setting in medical image analysis. We have shown how P-GiNGR can be used to obtain the MAP solution in a femur reconstruction challenge. We then showed the importance of using soft correspondence to analyze partial data, especially when the global frame of the structure is not known as was the case for the femur. We have shown an example of how the likelihood function in P-GiNGR can be modified to account for the target surface not being a strict subset of the reference which is needed for the independent point likelihood. We also showed different applications of using P-GiNGR to create PDMs from partial data using multiple-imputations. We showed how the PDMs can be created purely from partial data, from data that has overlapping areas, or how a combined model can be created from completely disjoint data. In the following chapter, we discuss future research directions and what parts still need to be implemented from a practical aspect.

6
Future Directions

In this chapter, we highlight future research directions and practical aspects of the P-GiNGR framework.

For GiNGR there is still more work to be conducted in converting existing non-rigid registration algorithms into GiNGR. In Section 3.3 we discussed related work which might be possible to use within GiNGR, such as correspondence approximation methods and different kernel functions. Further analysis would need to be conducted to find an optimal configuration that works on a wide variety of applications. This would mean that the method has to be efficient to compute if used in robotic applications such as SLAM but at the same time precise enough to be used within medical image analysis. Another aspect of GiNGR is kernel design, where it is necessary with an efficient kernel method for GPMMs which works on the surface instead of Gaussian, b-spline, or similar kernels that compute the covariance based on the Euclidean distance. When moving from GiNGR to P-GiNGR we introduced the CP in Section 4.2.2, where the observation noise was manually set to a simple anisotropic Gaussian noise, with a large uncertainty along the surface and a small uncertainty in the normal direction of the surface. While this method works well for the femur and the face, it might not work for structures without large smooth regions. A more dynamic noise term would therefore

be a good idea which not only looks at the correspondence distance between the reference and the target, but also takes the surface curvature into account. This could possibly be done by making the noise isotropic when the surface is having large curvatures and anisotropic in large smooth areas.

In general, the above-mentioned research directions revolve around automatic hyper-parameter setting such that the end-user can use it as a black-box method. Another question is the practical aspect of using P-GiNGR in a commercial application to compute registration uncertainty. In comparison to GiNGR, with P-GiNGR we are approximating the posterior distribution by sampling. We can analyze individual domains and see how many samples might be sufficient. But it is difficult to define a limit that works for a broad range of applications with the current configuration. This has the implication that we theoretically will have to take infinitely many samples. Future research directions could be on VBI methods to approximate the posterior distribution instead of sampling methods.

The uncertainty estimation that we are able to compute with P-GiNGR can be contributed to multiple factors. It would therefore be useful to divide the uncertainty into how much is due to the unknown correspondence and how much is due to the model approximation which might not be flexible enough to explain parts of the target.

With P-GiNGR we are using the independent point evaluator likelihood (Section 4.2). On a closer look, the independence assumption is obviously flawed. As an example, we can discretize the surface such that the point distance is close to zero. However, the likelihood would still assume that two points in such close proximity are independent. In Section 5.3 we showed different practical solutions to the likelihood function. But a good likelihood function for multiple domains, which also needs to be theoretically sounds, needs to be found. Future research directions would therefore be on searching for a better likelihood function, which ideally would take the point dependence into account.

6.1 Open Source Software

All of the experiments throughout this thesis have been conducted with the open-source scalable image analysis and shape modeling library, Scalismo[7]. Until now we have published code for an ICP version of the CP proposal[8] together with [61]. We also published code to construct PDMs from partial data[9] together with [60]. What still needs to be done is combining it all into the P-GiNGR software package to make it easily approachable for a wide range of people. As an entry-level method, it could be possible to create P-GiNGR-UI by utilizing Scalismo-UI[10] and therefore make non-rigid registration even easier to perform for non-technical people without coding knowledge.

[7] scalismo.org/
[8] github.com/unibas-gravis/icp-proposal
[9] github.com/unibas-gravis/shape-priors-from-pieces
[10] github.com/unibas-gravis/scalismo-ui

7
Conclusion

In this thesis, we introduced P-GiNGR which is a generalized method for probabilistic non-rigid surface and point-set registration. With P-GiNGR we showed how to perform non-rigid registration without using hard correspondence. Instead, the correspondence is described through a likelihood function. We showed how different likelihood functions can be used such as independent point evaluation, collective average, or a likelihood which takes the Hausdorff distance between surfaces into account. With GiNGR we showed how non-rigid registration can clearly separate the prior deformation information from the general registration algorithm using GPMMs and GPR.

We converted existing algorithms, such as ICP and CPD into GiNGR and showed that the regularization hyper-parameters are converted to an easily understandable property, namely a Gaussian noise assumption of the estimated correspondence pairs. Beyond comparing existing registration algorithms, GiNGR can be used to easily switch between different methods during registration or to create methods that combine parts from existing algorithms. The framework gives additional benefits such as a clear concept for including expert-annotation, multi-resolution fitting for faster and more robust registration procedures, and the ability to use statistical deformation priors.

With P-GiNGR we can even make the iterations in existing registration

algorithms stochastic in comparison to their deterministic nature. The probabilistic interpretation is obtained with the MH algorithm and a specially designed correspondence-proposal from which we are able to take informed steps while keeping the convergence properties of the MH algorithm. In addition to making existing algorithms more robust with P-GiNGR, we also show how the posterior distribution of registrations can be used to assess correspondence uncertainty. Finally, we showed a wide variety of applications of using P-GiNGR on analyzing partial data and how P-GiNGR can be used to create PDMs from partial data.

With the modular design and open-source availability of P-GiNGR, we hope to bring different non-rigid point-set and surface registration communities together to develop deformation priors and correspondence estimation methods that work on a large variety of different domains.

A

Derivation of ICP-T and ICP-A

A.1 ICP-T Full Derivation

The ICP-T minimization problem

$$\arg\min_{\tilde{U}} = \left\| \begin{bmatrix} \lambda_s B \\ WI \end{bmatrix} \tilde{U} - \begin{bmatrix} 0 \\ W\hat{U} \end{bmatrix} \right\|_F^2 \tag{A.1}$$

To simplify the derivation, at first we set $W = I$ and $\lambda_s = 1$, so the minimization instead becomes

$$\arg\min_{\tilde{U}} = \left\| \begin{bmatrix} B \\ I \end{bmatrix} \tilde{U} - \begin{bmatrix} 0 \\ \hat{U} \end{bmatrix} \right\|_F^2 \tag{A.2}$$

Then we expand it into the least square solution:

$$\tilde{U} = (B^T B + I)^{-1} \hat{U} \tag{A.3}$$

Here we recognize that $L = B^T B$ is the Laplacian matrix and $\hat{U} = X_c - X_R$. We now make use of the Woodbury matrix identity

$$(\mathcal{A} + \mathcal{U}\mathcal{C}\mathcal{V}^T)^{-1} = \mathcal{A}^{-1} - \mathcal{A}^{-1}\mathcal{U}(\mathcal{C}^{-1} + \mathcal{V}^T\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}^T\mathcal{A}^{-1}. \tag{A.4}$$

We set $\mathcal{A}^{-1} = L^\dagger = K$, and $\mathcal{U}$, $\mathcal{C}$, and $\mathcal{V}$ to identity matrices. Note, the Laplacian matrix L does not have full rank and therefore we use the pseudo

inverse. We can rewrite

$$\begin{aligned}\tilde{U} &= (L+I)^{-1}\hat{U} \\ &= (K - K(K+I)^{-1}K)\hat{U}\end{aligned}$$

The above equation is then set equal to the mean of GPR (eq. 8) when choosing the inverse Laplacian as the kernel. Then we can show that under these assumptions $u = \hat{U}$ holds, and hence both methods lead to the same prediction.

$$K(K+I)^{-1}u = (K - K(K+I)^{-1}K)\hat{U} \tag{A.5}$$

Multiply with $(K+I)K^{-1}$ on both sides

$$u = (K+I)K^{-1}(K - K(K+I)^{-1}K)\hat{U} \tag{A.6}$$

$$\begin{aligned}u &= ((K+I) - K)\hat{U} \\ &= \hat{U}\end{aligned}$$

which means that

$$\tilde{U} = K(K+I)^{-1}\hat{U}. \tag{A.7}$$

A.1.1 Inclusion of Weighting Matrix W, λ_s^2 and σ^2

We now focus on what the W matrix, stiffness λ_s and the noise assumption σ^2 variables change in the derivation. So we start out with the minimization problem in Eq. (A.1)

$$\begin{aligned}\tilde{U} &= (L + \lambda_s^2 I W^T I W)^{-1} W^T I W \hat{U} \\ &= (K - K W^T (\sigma^2 I + W K W^T)^{-1} W K) W \hat{U}\end{aligned}$$

Continuing the derivation similar to Eq. (A.5), this time equating to the GPR: $KW^T(\sigma^2 + WKW^T)^{-1}Wu$ extended with W, leads to

$$\begin{aligned}KW^T(\sigma^2 I + WKW^T)^{-1}Wu = \\ \big(\frac{1}{\lambda_s^2}K - \frac{1}{\lambda_s^2}KW(I + \frac{1}{\lambda_s^2}WKW)^{-1}\frac{1}{\lambda_s^2}WKW\big)W\hat{U}\end{aligned} \tag{A.8}$$

Using $(I + \frac{1}{\lambda_s^2}WKW)^{-1} = \lambda_s^2(\lambda_s^2 I + WKW)^{-1}$ and also $W = WW = W^T$, as well as assuming $\lambda_s = \sigma$, the same approach as in Eq. (A.6) results in

$$\begin{aligned} Wu &= (\sigma^2 I + WKW)WK\big(\frac{1}{\lambda_s^2}K - \\ &\qquad \frac{1}{\lambda_s^2}KW(\lambda_s^2 I + WKW)^{-1}WKW\big)W\hat{U} \\ &= \big(\frac{1}{\lambda_s^2}(\sigma^2 I + WKW)W - \frac{1}{\lambda_s^2}WKW\big)W\hat{U} \\ &= W\hat{U} \end{aligned} \tag{A.9}$$

Therefore, the stiffness term λ_s and the uncertainty assumption σ are describing the same even when including W. When comparing this to (eq. 7), we see that WKW^T is the kernel function, but with all the unknown items removed. The same goes for $\hat{U}$ which now only contains the points that we have observed values for, the rest will be 0. Likewise, KW is the kernel matrix between all pairs of predicted points and the observed points, where the correlation between predicted points and unobserved points has been set to 0.

A.2 ICP-A Full Derivation

The ICP-A minimization problem

$$\arg\min_{\mathcal{M}} = \left\| \begin{bmatrix} \lambda_s B \otimes G \\ WD \end{bmatrix} \mathcal{M} - \begin{bmatrix} 0 \\ WX_c \end{bmatrix} \right\|_F^2 \tag{A.10}$$

with G being the $\mathrm{diag}([1, 1, 1, \gamma]^T)$ and γ depending on the units of the data. Again, we begin with the ICP-A energy optimization terms expanded to the least square solution. For simplicity we have set $W = I$ and $\lambda_s = 1$. The inclusion of W and λ_s leads to the same outcome as shown in Appendix A.1.1.

$$\mathcal{M} = (B^T B \otimes G^T G + D^T D)^{-1} D^T X_c \tag{A.11}$$

$$= (L + D^T D)^{-1} D^T X_c \tag{A.12}$$

We now make use of the Woodbury matrix identity Eq. (A.4) and set $\mathcal{U} = D^T$, $\mathcal{V} = D, \mathcal{C} = I$ and $\mathcal{A}^{-1} = L^\dagger = K$.

$$\mathcal{M} = (KD^T - KD^T(I + DKD^T)^{-1}DKD^T)X_c \tag{A.13}$$

As $\mathcal{M}$ is the stacked affine transformation matrices per point, we then multiply the solution with D to get the updated point locations:

$$\hat{X}_R = (DKD^T - DKD^T(I + DKD^T)^{-1}DKD^T)X_c \tag{A.14}$$

Setting $\mathcal{K} = DKD^T$ leads to

$$\hat{X}_R = (\mathcal{K} - \mathcal{K}(I + \mathcal{K})^{-1}\mathcal{K})X_c \tag{A.15}$$

Similar to in the ICP-T derivation, we set the predicted point locations $\hat{X}_R$ equal to the GPR mean (see eq. 8) using $\mathcal{K}$ as the kernel. Then we solve for the observation field $\hat{U}$. Notice, in the ICP-A formulation, the updated locations of the points are predicted directly and not a deformation field $\tilde{U}$ that warps the reference. We can get the updated point locations in the standard GPMM formulation by adding the predicted mean deformation field to the reference point locations X_R,

$$(\mathcal{K} - \mathcal{K}(I + \mathcal{K})^{-1}\mathcal{K})X_c = X_R + \mathcal{K}(I + \mathcal{K})^{-1}\hat{U}. \tag{A.16}$$

We now isolate $\hat{U}$ by multiplying both sides with $(\mathcal{K} + I)\mathcal{K}^{-1}$

$$(((\mathcal{K} + I)\mathcal{K}^{-1})(\mathcal{K} - \mathcal{K}(I + \mathcal{K})^{-1}\mathcal{K}))X_c = (\mathcal{K} + I)\mathcal{K}^{-1}X_R + \hat{U} \tag{A.17}$$

$$((\mathcal{K} + I) - \mathcal{K})X_c = (X_R + \mathcal{K}^{-1}X_R) + \hat{U} \tag{A.18}$$

$$X_c - (X_R + \mathcal{K}^{-1}X_R) = \hat{U} \tag{A.19}$$

The additional term $\mathcal{K}^{-1}X_R$ (in comparison to ICP-T) is to adjust for the ICP-A algorithm defining its closest point deformations as affine transformations.

B

GiNGR vs P-GiNGR Plots

Complete individual results for comparison of GiNGR and P-GiNGR using either ICP or CPD to estimate correspondence in each iteration. Each boxplot in Fig. B.1, Fig. B.2, Fig. B.3 and Fig. B.4 is estimated based on 100 random initialization of the GPMM model parameters $\boldsymbol{\alpha}$. Notice how the figures both show the Euclidean distance and the Hausdorff distance which are both in general better when using P-GiNGR instead of the same configuration (ICP or CPD and noise model) with GiNGR.

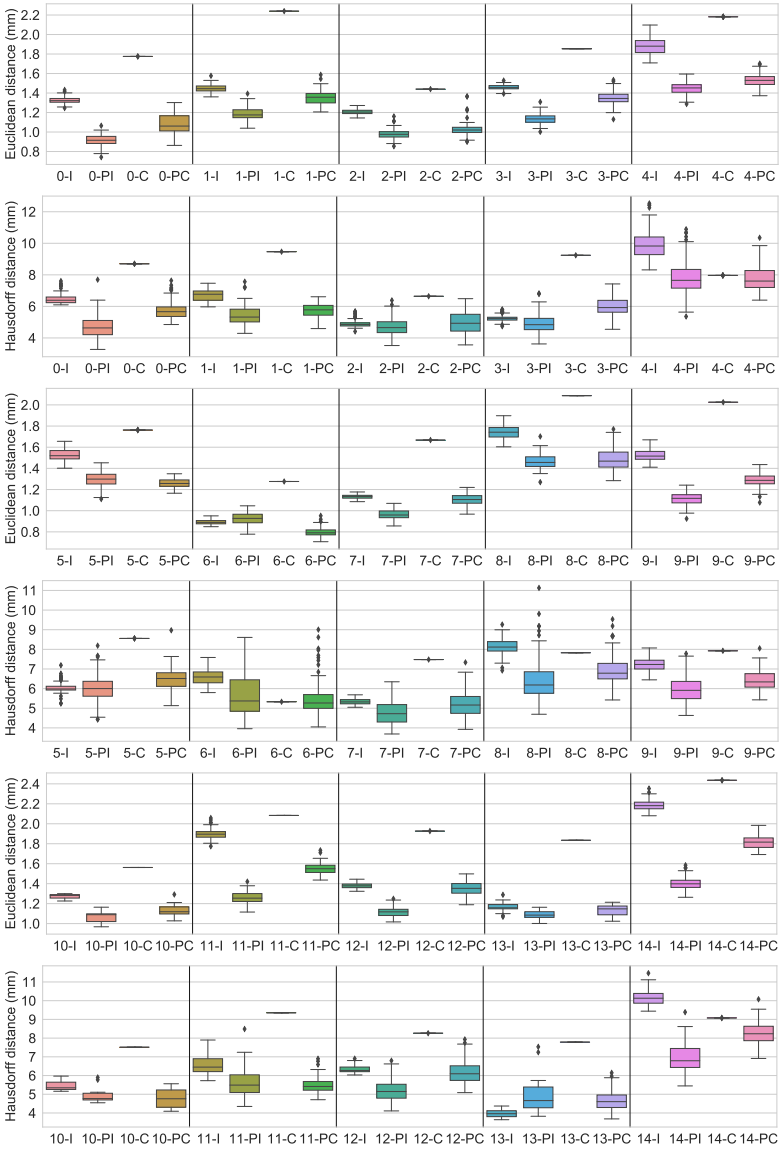


Figure B.1: GiNGR vs P-GiNGR across 50 target femurs where the GPMM is initialized with 100 random shape coefficients.

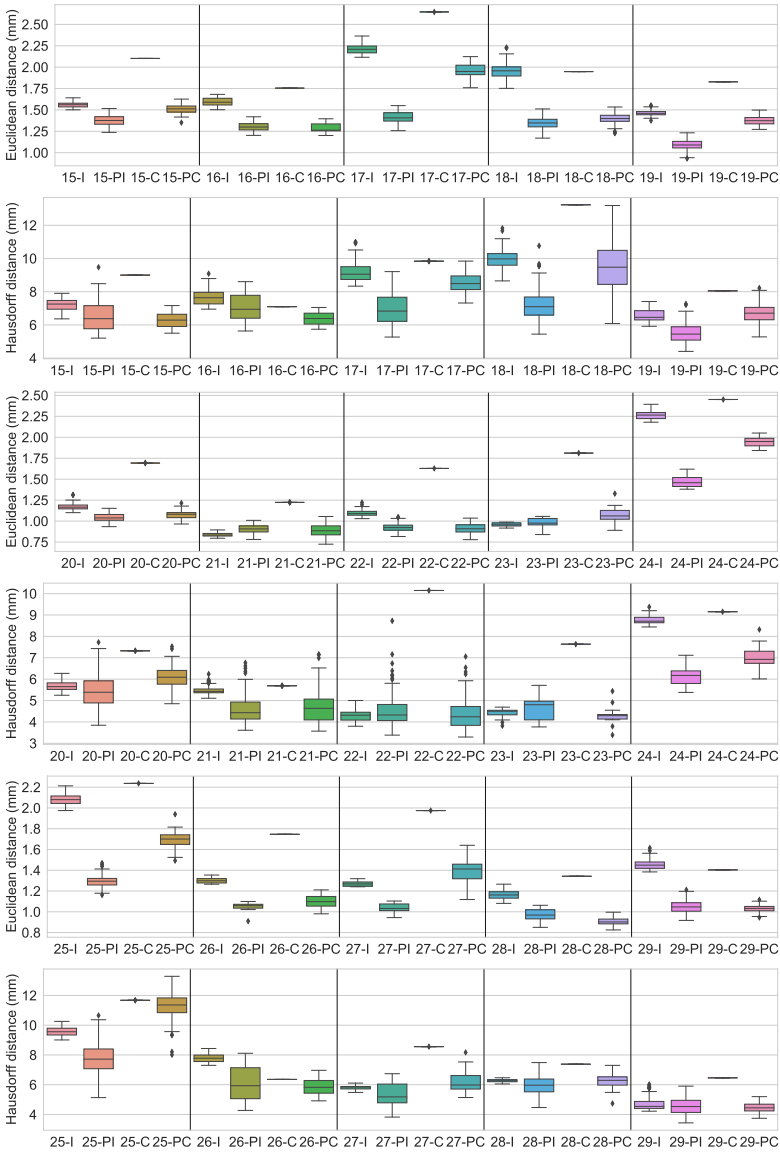


Figure B.2: GiNGR vs P-GiNGR across 50 target femurs where the GPMM is initialized with 100 random shape coefficients.

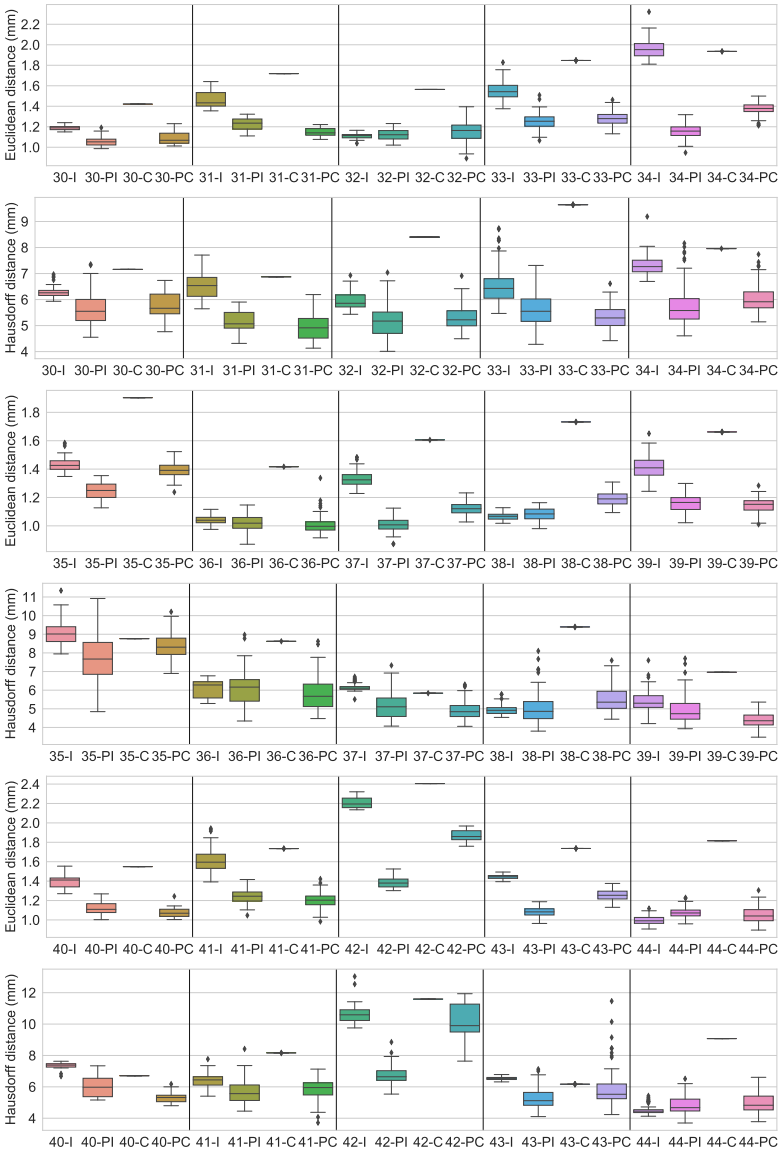


Figure B.3: GiNGR vs P-GiNGR across 50 target femurs where the GPMM is initialized with 100 random shape coefficients.

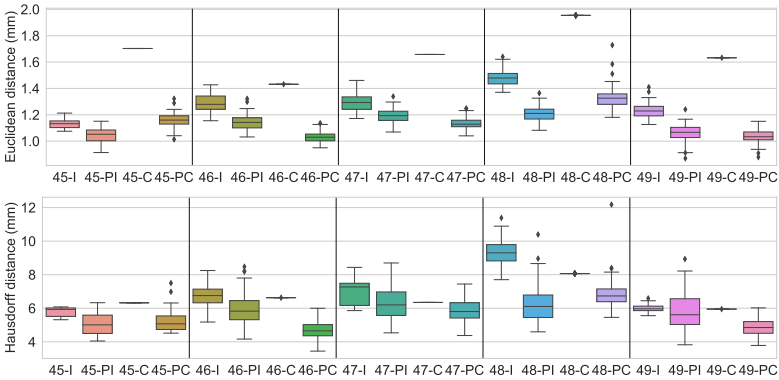


Figure B.4: GiNGR vs P-GiNGR across 50 target femurs where the GPMM is initialized with 100 random shape coefficients.

Bibliography

[1] Thomas Albrecht, Marcel Lüthi, Thomas Gerig, and Thomas Vetter. Posterior shape models. *Medical image analysis*, 17(8):959–973, 2013.

[2] Brett Allen, Brian Curless, and Zoran Popovic. The space of human body shapes: reconstruction and parameterization from range scans. *TOG*, 22(3):587–594, 2003.

[3] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[4] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. MESH: measuring errors between surfaces using the Hausdorff distance. In *Proceedings. IEEE International Conference on Multimedia and Expo*, volume 1, pages 705–708 vol.1, August 2002.

[5] Florian Bernard, Luis Salamanca, Johan Thunberg, Alexander Tack, Dennis Jentsch, Hans Lamecker, Stefan Zachow, Frank Hertel, Jorge Goncalves, and Peter Gemmar. Shape-aware surface reconstruction from sparse 3D point-clouds. *Medical Image Analysis*, 38:77–89, May 2017.

[6] Paul J. Besl and Neil D. McKay. Method for registration of 3-D shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

[7] Rémi Blanc, Mauricio Reyes, Christof Seiler, and Gábor Székely. Conditional variability of statistical shape models based on surrogate variables. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 84–91. Springer, 2009.

[8] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3794–3801, 2014.

[9] Benedict J. Brown and Szymon Rusinkiewicz. Global non-rigid alignment of 3-D scans. In *ACM SIGGRAPH 2007 papers*, pages 21–es. 2007.

[10] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[11] Shiyang Cheng, Ioannis Marras, Stefanos Zafeiriou, and Maja Pantic. Statistical non-rigid ICP algorithm and its application to 3D face alignment. *Image and Vision Computing*, 58:3–12, February 2017.

[12] Dmitry Chetverikov, Dmitry Stepanov, and Pavel Krsek. Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing*, 23(3):299–309, 2005.

[13] Haili Chui and Anand Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89 (2):114–141, February 2003.

[14] Peter Claes, Dirk Vandermeulen, Sven De Greef, Guy Willems, and Paul Suetens. Craniofacial reconstruction using a combined statistical model of face shape and soft tissue depths: methodology and validation. *Forensic science international*, 159:S147–S158, 2006.

[15] Peter Claes, Dirk Vandermeulen, Sven De Greef, Guy Willems, John Gerald Clement, and Paul Suetens. Computerized craniofacial

reconstruction: conceptual framework and review. *Forensic science international*, 201(1-3):138–145, 2010.

[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training Models of Shape from Sets of Examples. In *In Proc. British Machine Vision Conference*, pages 9–18. Springer-Verlag, 1992.

[17] Maya de Buhan and Chiara Nardoni. A mesh deformation based approach for digital facial reconstruction. 2016.

[18] Sven De Greef, Peter Claes, Dirk Vandermeulen, Wouter Mollemans, Paul Suetens, and Guy Willems. Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction. *Forensic science international*, 159:S126–S146, 2006.

[19] Qingqiong Deng, Mingquan Zhou, Zhongke Wu, Wuyang Shui, Yuan Ji, Xingce Wang, Ching Yiu Jessica Liu, Youliang Huang, and Haiyan Jiang. A regional method for craniofacial reconstruction based on coordinate adjustments and a new fusion strategy. *Forensic science international*, 259:19–31, 2016.

[20] Fuqing Duan, Sen Yang, Donghua Huang, Yongli Hu, Zhongke Wu, and Mingquan Zhou. Craniofacial reconstruction based on multi-linear subspace analysis. *Multimedia Tools and Applications*, 73(2):809–823, 2014.

[21] Fuqing Duan, Donghua Huang, Yun Tian, Ke Lu, Zhongke Wu, and Mingquan Zhou. 3d face reconstruction from skull by regression modeling in shape parameter spaces. *Neurocomputing*, 151:674–682, 2015.

[22] David Duvenaud. *Automatic model construction with Gaussian processes*. PhD Thesis, University of Cambridge, 2014.

[23] Jacques Feldmar and Nicholas Ayache. Rigid, affine and locally affine registration of free-form surfaces. *Int J Comput Vision*, 18(2):99–119, May 1996.

[24] Peter Filzmoser, Ricardo Maronna, and Mark Werner. Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, 52(3):1694–1711, January 2008.

[25] L. Le Folgoc, H. Delingette, A. Criminisi, and N. Ayache. Quantifying Registration Uncertainty With Sparse Bayesian Modelling. *IEEE Transactions on Medical Imaging*, 36(2):607–617, February 2017.

[26] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. CRC press, 2013.

[27] Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Luthi, Sandro Schönborn, and Thomas Vetter. Morphable face models-an open framework. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 75–82. IEEE, 2018.

[28] Sébastien Granger and Xavier Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *European Conference on Computer Vision*, pages 418–432. Springer, 2002.

[29] Ulf Grenander and Michael I. Miller. Representations of Knowledge in Complex Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(4):549–581, 1994.

[30] B. Gutierrez, D. Mateus, E. Shiban, B. Meyer, J. Lehmberg, and N. Navab. A sparse approach to build shape models with routine clinical data. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 258–261, April 2014.

[31] Ivan Gutman and W. Xiao. Generalized inverse of the Laplacian matrix and some applications. *Bull Class Sci Math*, 129(29):15–23, 2004.

[32] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[33] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, August 2009.

[34] O. Hirose. Acceleration of non-rigid point set registration with downsampling and Gaussian process regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[35] Osamu Hirose. A Bayesian Formulation of Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[36] Thomas Hitchcox and James Richard Forbes. A Point Cloud Registration Pipeline Using Gaussian Process Regression for Bathymetric SLAM. page 8, 2020.

[37] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

[38] Hidekata Hontani, Takamiti Matsuno, and Yoshihide Sawada. Robust nonrigid ICP using outlier-sparsity regularization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 174–181. IEEE, 2012.

[39] Yongli Hu, Fuqing Duan, Mingquan Zhou, Yanfeng Sun, and Baocai Yin. Craniofacial reconstruction based on a hierarchical dense deformable model. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–14, 2012.

[40] Heike Hufnagel, Xavier Pennec, Jan Ehrhardt, Nicholas Ayache, and Heinz Handels. Generation of a statistical shape model with probabilistic point correspondences and the expectation maximization- iterative closest point algorithm. *Int J CARS*, 2(5):265–273, 2008.

[41] Varun Jampani, Sebastian Nowozin, Matthew Loper, and Peter V. Gehler. The informed sampler: A discriminative approach to bayesian

inference in generative computer vision models. *Computer Vision and Image Understanding*, 136:32–44, 2015.

[42] B. Jiang, J. Zhang, J. Cai, and J. Zheng. Disentangled Human Body Embedding Based on Deep Hierarchical Neural Network. *IEEE Transactions on Visualization and Computer Graphics*, 26(8):2560–2575, August 2020.

[43] Ma Jingting, Katharina Lentzen, Jonas Honsdorf, Lin Feng, and Marius Erdt. Statistical shape modeling from gaussian distributed incomplete data for image segmentation. In *Workshop on Clinical Image-Based Procedures*, pages 113–121. Springer, 2015.

[44] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, May 1999.

[45] Junmo Kim, Müjdat Çetin, and Alan S. Willsky. Nonparametric shape priors for active contour-based image segmentation. *Signal Processing*, 87(12):3021–3044, December 2007.

[46] Michael Kistler, Serena Bonaretti, Marcel Pfahrer, Roman Niklaus, and Philippe Büchler. The Virtual Skeleton Database: An Open Access Repository for Biomedical Research and Collaboration. *Journal of Medical Internet Research*, 15(11):e2930, November 2013.

[47] Adam Kortylewski, Mario Wieser, Andreas Morel-Forster, Aleksander Wieczorek, Sonali Parbhoo, Volker Roth, and Thomas Vetter. Informed MCMC with Bayesian Neural Networks for Facial Image Analysis. *arXiv:1811.07969 [cs]*, November 2018.

[48] Yan Li, Liang Chang, Xuejun Qiao, Rong Liu, and Fuqing Duan. Craniofacial reconstruction based on least square support vector regression. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1147–1151. IEEE, 2014.

[49] Luming Liang, Mingqiang Wei, Andrzej Szymczak, Anthony Petrella, Haoran Xie, Jing Qin, Jun Wang, and Fu Lee Wang. Nonrigid iterative closest points for registration of 3D biomedical surfaces. *Optics and Lasers in Engineering*, 100:141–154, January 2018.

[50] H. Liu, T. Liu, Y. Li, M. Xi, T. Li, and Y. Wang. Point Cloud Registration Based on MCMC-SA ICP Algorithm. *IEEE Access*, 7:73637–73648, 2019.

[51] Marcel Lüthi, Thomas Albrecht, and Thomas Vetter. Building shape models from lousy data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1–8. Springer, 2009.

[52] Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter. Gaussian process morphable models. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1860–1873, 2017.

[53] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu. Robust L_2e Estimation of Transformation for Non-Rigid Registration. *IEEE Transactions on Signal Processing*, 63(5):1115–1129, March 2015.

[54] J. Ma, J. Zhao, and A. L. Yuille. Non-Rigid Point Set Registration by Preserving Global and Local Structures. *IEEE Transactions on Image Processing*, 25(1):53–64, January 2016.

[55] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng. Nonrigid Point Set Registration With Robust Transformation Learning Under Manifold Regularization. *IEEE Transactions on Neural Networks and Learning Systems*, 30(12):3584–3597, December 2019.

[56] Jingting Ma, Feng Lin, Jonas Honsdorf, Katharina Lentzen, Stefan Wesarg, and Marius Erdt. Weighted robust PCA for statistical shape modeling. In *International Conference on Medical Imaging and Augmented Reality*, pages 343–353. Springer, 2016.

[57] Jingting Ma, Anqi Wang, Feng Lin, Stefan Wesarg, and Marius Erdt. A novel robust kernel principal component analysis for nonlinear statistical shape modeling from erroneous data. *Computerized Medical Imaging and Graphics*, 77:101638, October 2019.

[58] Dennis Madsen, Marcel Lüthi, Andreas Schneider, and Thomas Vetter. Probabilistic joint face-skull modelling for facial reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5295–5303, 2018.

[59] Dennis Madsen, Thomas Vetter, and Marcel Lüthi. Probabilistic Surface Reconstruction with Unknown Correspondence. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures*, pages 3–11. Springer, 2019.

[60] Dennis Madsen, Jonathan Aellen, Andreas Morel-Forster, Thomas Vetter, and Marcel Lüthi. Learning Shape Priors from Pieces. In *International Workshop on Shape in Medical Imaging*, pages 30–43. Springer, 2020.

[61] Dennis Madsen, Andreas Morel-Forster, Patrick Kahr, Dana Rahbani, Thomas Vetter, and Marcel Lüthi. A closest point proposal for MCMC-based probabilistic surface registration. In *European Conference on Computer Vision*, pages 281–296. Springer, 2020.

[62] Baraka Maiseli, Yanfeng Gu, and Huijun Gao. Recent developments and trends in point set registration methods. *Journal of Visual Communication and Image Representation*, 46:95–106, July 2017.

[63] G. McNeill and S. Vijayakumar. A Probabilistic Approach to Robust Shape Matching. pages 937–940, October 2006.

[64] Graham McNeill and Sethu Vijayakumar. Part-based probabilistic point matching. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 382–386. IEEE, 2006.

[65] Andreas Morel-Forster, Thomas Gerig, Marcel Lüthi, and Thomas Vetter. Probabilistic fitting of active shape models. In *International Workshop on Shape in Medical Imaging*, pages 137–146. Springer, 2018.

[66] Andriy Myronenko and Xubo Song. Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, December 2010.

[67] Andriy Myronenko, Xubo Song, and Miguel Á Carreira-Perpiñán. Non-rigid point set registration: Coherent Point Drift. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1009–1016. MIT Press, 2007.

[68] Shruti Nagpal, Maneet Singh, Arushi Jain, Richa Singh, Mayank Vatsa, and Afzel Noore. On matching skulls to digital face images: A preliminary approach. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 813–819, October 2017.

[69] Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

[70] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3D face model for pose and illumination invariant face recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 296–301. Ieee, 2009.

[71] Pascal Paysan, Marcel Lüthi, Thomas Albrecht, Anita Lerch, Brian Amberg, Francesco Santini, and Thomas Vetter. Face reconstruction from skull shapes and physical attributes. In *Joint Pattern Recognition Symposium*, pages 232–241. Springer, 2009.

[72] Stylianos Ploumpis, Haoyang Wang, Nick Pears, William A. P. Smith, and Stefanos Zafeiriou. Combining 3D Morphable Models: A Large

Scale Face-And-Head Model. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10926–10935. IEEE, June 2019.

[73] François Pomerleau, Francis Colas, and Roland Siegwart. A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends in Robotics*, 4(1):1–104, 2015.

[74] Ruma Purkait. Triangle identified at the proximal end of femur: a new sex determinant. *Forensic science international*, 147(2-3):135–139, 2005.

[75] Dana Rahbani, Andreas Morel-Forster, Dennis Madsen, Marcel Lüthi, and Thomas Vetter. Robust Registration of Statistical Shape Models for Unsupervised Pathology Annotation. In *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*, pages 13–21. Springer, 2019.

[76] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.

[77] P. Risholm, A. Fedorov, J. Pursley, K. Tuncali, R. Cormack, and W. M. Wells. Probabilistic non-rigid registration of prostate images: Modeling and quantifying uncertainty. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 553–556, March 2011.

[78] Petter Risholm, Steve Pieper, Eigil Samset, and William M Wells. Summarizing and visualizing uncertainty in non-rigid registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 554–561. Springer, 2010.

[79] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[80] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[81] Donald B. Rubin. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In *Proceedings of the survey research methods section of the American Statistical Association*, volume 1, pages 20–34. American Statistical Association, 1978.

[82] R. B. Rusu, N. Blodow, and M. Beetz. Fast Point Feature Histograms (FPFH) for 3D registration. In *2009 IEEE International Conference on Robotics and Automation*, pages 3212–3217, May 2009.

[83] Sandro Schönborn, Bernhard Egger, Andreas Morel-Forster, and Thomas Vetter. Markov Chain Monte Carlo for Automated Face Image Analysis. *Int J Comput Vis*, 123(2):160–183, June 2017.

[84] Sarah Shrimpton, Katleen Daniels, Sven De Greef, Francoise Tilotta, Guy Willems, Dirk Vandermeulen, Paul Suetens, and Peter Claes. A spatially-dense regression study of facial form and tissue depth: towards an interactive tool for craniofacial reconstruction. *Forensic science international*, 234:103–110, 2014.

[85] Wuyang Shui, Mingquan Zhou, Steve Maddock, Taiping He, Xingce Wang, and Qingqiong Deng. A PCA-Based method for determining craniofacial relationship and sexual dimorphism of facial shapes. *Computers in Biology and Medicine*, 90:33–49, November 2017.

[86] Florian Steinke and Bernhard Schölkopf. Kernels, regularization and differential equations. *Pattern Recognition*, 41(11):3271–3286, November 2008.

[87] Carl N. Stephan. The application of the central limit theorem and the law of large numbers to facial soft tissue depths: T-Table robustness

and trends since 2008. *Journal of forensic sciences*, 59(2):454–462, 2014.

[88] Carl N. Stephan and Ellie K. Simpson. Facial soft tissue depths in craniofacial identification (part I): an analytical review of the published adult data. *Journal of Forensic Sciences*, 53(6):1257–1272, 2008.

[89] 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 *Biennial International Conference on Information Processing in Medical Imaging*, pages 63–75. Springer, 2003.

[90] 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.

[91] Mildred Trotter and Goldine C. Gleser. Estimation of stature from long bones of American Whites and Negroes. *American journal of physical anthropology*, 10(4):463–514, 1952.

[92] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154, February 2003.

[93] Douglas H. Ubelaker. Craniofacial superimposition: historical review and current issues. *Journal of forensic sciences*, 60(6):1412–1419, 2015.

[94] S. Umeyama. Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(04):376–380, April 1991.

[95] Ulrike von Luxburg. A Tutorial on Spectral Clustering. *arXiv:0711.0189 [cs]*, November 2007.

[96] D. Wang, J. Xue, D. Cui, and Y. Zhong. A robust submap-based road shape estimation via iterative Gaussian process regression. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1776–1781, June 2017.

[97] Caroline Wilkinson. *Forensic facial reconstruction*. Cambridge University Press, 2004.

[98] Caroline Wilkinson. Facial reconstruction–anatomical art or artistic anatomy? *Journal of anatomy*, 216(2):235–250, 2010.

[99] Max A. Woodbury. *Inverting modified matrices*. Statistical Research Group, 1950.

[100] Zedong Xiao, Junli Zhao, Xuejun Qiao, and Fuqing Duan. Craniofacial reconstruction using gaussian process latent variable models. In *International Conference on Computer Analysis of Images and Patterns*, pages 456–464. Springer, 2015.

[101] Yang C. Yuan. Multiple imputation for missing data: Concepts and new development (Version 9.0). *SAS Institute Inc, Rockville, MD*, 49(1-11): 12, 2010.

[102] Guoyan Zheng, Xiao Dong, Kumar T. Rajamani, Xuan Zhang, Martin Styner, Ramesh U. Thoranaghatte, Lutz-Peter Nolte, and Miguel A. Gonzalez Ballester. Accurate and robust reconstruction of a surface model of the proximal femur from sparse-point data and a dense-point distribution model for surgical navigation. *IEEE transactions on biomedical engineering*, 54(12):2109–2122, 2007.

[103] Hao Zhu, Bin Guo, Ke Zou, Yongfu Li, Ka-Veng Yuen, Lyudmila Mihaylova, and Henry Leung. A review of point set registration: From pairwise registration to groupwise registration. *Sensors*, 19(5):1191, 2019.

[104] Zhonglin Zhu and Guoan Li. Construction of 3D human distal femoral surface models using a 3D statistical deformable model. *Journal of biomechanics*, 44(13):2362–2368, 2011.

[105] Guangyu Zou, Jing Hua, and Otto Muzik. Non-rigid surface registration using spherical thin-plate splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 367–374. Springer, 2007.