

Using Line Features for 3D Face Registration

by

Fabian Brix

Bachelor Thesis

submitted in partial fulfillment
of the requirements for the degree of

BSc in Computer Science

at the **UNIVERSITÄT BASEL**

Examiner

Prof. Thomas Vetter

Supervisors

Dr. Marcel Luethi, MSc Andreas Forster

Graphics and Vision Research Group
Department of Mathematics and Computer Science
Faculty of Science

July 26, 2013



Abstract

In this bachelor thesis we present a 3D face registration pipeline by adapting Gaussian Process Regression and using labelled point and line correspondences as prior information for the registration process. Gaussian Processes are ideal for modelling smooth deformations of an arbitrary shape surface. Therefore, Gaussian Process Regression can be implemented to sample deformation fields that map a template on to a target shape using a set of corresponding facial landmarks. In our approach we extend the registration of face scans by incorporating additional information from line features. The line features define correspondence in highly characteristic areas of a set of faces, i.e. the eyes and the ears. The deformation fields are optimized - using the prior information as constraints - to obtain a plausible mapping of the template on to the target. We demonstrate that line features have a profound impact on the accuracy and quality of the registration results. In order to enhance the expressiveness of the obtained registrations, we plan to further adjust the parameters of Gaussian Process Regression to the case of 3D faces and to tweak the optimization process in the near future.

Table of Contents

1	Introduction	5
1.1	Prior Work	5
1.2	Overview	6
2	3D Model Building	7
2.1	3D Morphable Model	7
2.2	Achieving Correspondence through Registration	8
2.3	Available Data	9
3	Gaussian Processes in 3D Face Registration	11
3.1	Stochastic Processes	11
3.2	Gaussian Processes	11
3.3	1-dimensional Gaussian Processes	12
3.4	Gaussian Process Regression	13
3.5	Application to 3D Face Meshes	15
3.6	Fitting & Optimization	18
4	Registration Pipeline using Line Features	21
4.1	Line Features	21
4.2	Sampling 3D Points from 2D Line Features	22
4.3	Rigid Mesh Alignment	27
4.4	Prior Model	28
4.5	Posterior Model	28
4.6	Optimization of the Posterior	30
4.7	Robust Loss Functions	31
5	Results	34
5.1	Line Feature Sampling Inaccuracy	34

5.2	The effect of Line Features	35
6	Discussion & Future Work	42

Introduction

3D Face Models have a wide variety of applications in face image analysis. They are used in face recognition from 2D images, 3D face reconstruction from 2D images and the extraction of face attributes such as facial expressions. Given a set of scanned faces, registering these scans is the first step in the process of deriving a 3D model of morphable faces. Registration is necessary to establish the required dense correspondence between the input scans. It is the process deforming a template so that it reproduces a face scan. In order to achieve this a set of known correspondences, for example point correspondences, are used to define a transformation. In the end, replacing the scans with the respectively deformed template ensures that all faces have the same topology. In effect, the respective parts of different faces can be combined to form new faces. Defining a reproducible and accurate registration algorithm is therefore essential to obtaining a high-quality 3D model. In this thesis we present a registration algorithm which tackles the problem of 3D face registration by modeling the transformation using Gaussian Process Regression (GPR) and extending the prior information used in the transformation to parametric curves - referred to as line features - which define key regions of the face.

1.1 Prior Work

Gaussian Processes for Machine Learning were comprehensively introduced in [1] as an approach for defining a prior distribution over a space of functions. A Gaussian Process (GP) is a stochastic process, which in turn can be seen as the generalization of probability distributions to functions. Computations prove to be straightforward because a GP is modelled by a multivariate normal distribution.[2] (Using Landmarks as a Deformation Prior for Hybrid Image Registration) GPs are used to describe transformation priors in Image Registration. In this context GP allows for defining a space of possible transformations of an image and further allows for

incorporating point correspondences as additional information for inference of admissible deformations. The distinction in regard to other registration methods lies therein that the corresponding points are not hard-wired in the transformation, but instead modify the prior to include only the more feasible transformations. In [3] the process of formulating the fitting and optimization of the admissible transformations is described for image registration. GPR has further been used to infer shapes in statistical shape models [4]. In short, the reason for applying GPR is its reproducability in modelling a deformation prior for arbitrary shapes.

The registration methods used so far [5] in 3D face registration are specifically designed for this context are therefore not very well reproducible. Furthermore, these registration methods have to be directly constrained on point correspondences in the optimization process, whereas with GPR only admissible deformations have to be optimized. The advantage of our approach of applying GPR is that we have derived a general framework for 3D face registration.

1.2 Overview

In face registration a common parametrization has to be established for all face scans. Using a GP in 3D face registration models a space of possible prior deformations of a facial scan for the purpose of registering it with another scan. The given a priori point correspondences - called landmarks - are used as additional information constraining the space of possible deformations. We additionally use line features to match semantically prominent regions - which have to be brought into dense correspondence - for all scans. With the help of line features we want to infer deformations that result in a more accurate mapping of these regions and therefore better registration results. Using GPR in the registration process results in a highly customizable algorithm, that is dependent on the covariance of the surface points on the a priori given information. The following chapters cover the basics of building 3D morphable models, GPR theory and the implementation of a face registration pipeline with the above-mentioned algorithm. The registration results shown in the last chapter give a first expression of what this method is capable of and how the incorporation of line features affects the registration outcome. The question of how to incorporate the line features in the registration process will be answered in the course of this thesis.

3D Model Building

This chapter describes how to build a generative textured 3D face model from an example set of 3D face scans. A morphable model is derived from the set of scans by transforming their shape and texture into a vector space representation. The term generative implies that - as the face space is spanned by the set of examples - new faces can be generated by calculating linear combinations of these examples.

2.1 3D Morphable Model

The 3D Morphable Model (3DMM) published by Blanz and Vetter in 1999 (bib) is a multidimensional function for modelling textured faces derived from a large set of 3D face scans. A vector space of face shapes can be constructed from the available data set where each face is represented by a shape-vector $S \in \mathbb{R}^{3n}$ that consists of a component-wise listing of its n vertices. The texture-vector then $T \in \mathbb{T}^{3n}$ contains the corresponding RGB values in similar fashion. New shapes and textures can now be constructed with a linear combination of the set of m available faces governed by shape and texture coefficients $\vec{\alpha}, \vec{\beta} \in \mathbb{R}^m$.

$$\mathcal{S}_{new} = \sum_{i=1}^m \alpha_i S_i, \quad \mathcal{T}_{new} = \sum_{i=1}^m \beta_i T_i \quad (2.1)$$

However, the goal of such a 3D face model is not just to construct arbitrary faces with linear combinations of the set of examples, but to derive plausible faces from them. This is achieved by estimating two multivariate normal distributions for the coefficients in $\vec{\alpha}$ and $\vec{\beta}$. These multivariate normal distributions are constructed from the average face shapes $\bar{S} \in \mathbb{R}^{3N}$ and textures $\bar{T} \in \mathbb{R}^{3N}$ of the datasets and the covariance matrices K_S and K_T , which are defined over the residuals of each example and with both the shape $R_S = S - \bar{S}$ and texture $R_T = T - \bar{T}$ averages. The covariance matrices are then used to perform Principal Component Analysis

(PCA) which defines a basis transformation to an orthogonal coordinate systems. The axes are defined by the eigenvectors of the respective covariance matrix.

$$\mathcal{S}(\vec{\alpha}) = \bar{S} + M_S \vec{\alpha}, \quad \mathcal{T}(\vec{\beta}) = \bar{T} + M_T \vec{\beta} \quad (2.2)$$

In (2.2) the $N = m$ principal eigenvectors of K_S and K_T respectively are assembled column-wise in M_S and M_T and scaled in a way such that the prior distribution over the face shape and texture parameters is given by a multivariate normal distribution with unit covariance (Amberg).

$$p(\vec{\alpha}, \vec{\beta}) = \mathcal{N}(\vec{\alpha} || \mathbf{0}, \mathbb{I}) \mathcal{N}(\vec{\beta} || \mathbf{0}, \mathbb{I}) \quad (2.3)$$

Calculating the likelihood for a set of coefficients is straightforward and therefore it is now possible to find out how likely a face is given the model.

2.2 Achieving Correspondence through Registration

In order for a 3D Morphable Model to generate plausible faces we have to make sure that all faces share the same representation, so that only the same parts of the respective faces are combined with one another. For example a nose is combined with another nose and not with a mouth. For this reason, the meshes first have to be brought into correspondence. In fact, we need dense correspondence between all mesh points. This is the case when vertices at the same semantical position in different meshes have the same vertex number, i.e the left corner of the left eye. The process of calculating a dense point-to-point correspondence between two meshes is called registration. The training data used for learning a 3D Morphable Models consists solely of registered examples of the 3D shape and texture of faces.

Registration Algorithms When performing registration, we start with a template shape that is a prototype of the semantic parametrization. In order to compute the parametrization for another shape we deform the template to match this shape. The shape being matched is referred to as the target. Both are considered as matched if the semantic face regions are mapped onto each other as closely as possible by satisfying a specified similarity measure. If dense correspondence is achieved, the target shape can be replaced with the deformed template, which will further be used in the process of building a model.

Today's registration algorithms use prior information in the form of manually clicked feature points at prominent points of the face, so called landmarks. These are important for the alignment of template and target as well as achieving an accurate match.

2.3 Available Data

For testing the registration algorithm we have a large set of face scans given - male as well as female - all of them recorded at the University of Basel by an active stereo vision system using structured light. The scanner records four 3D surfaces, called shells, and further records three high resolution color images for every face. The 3D

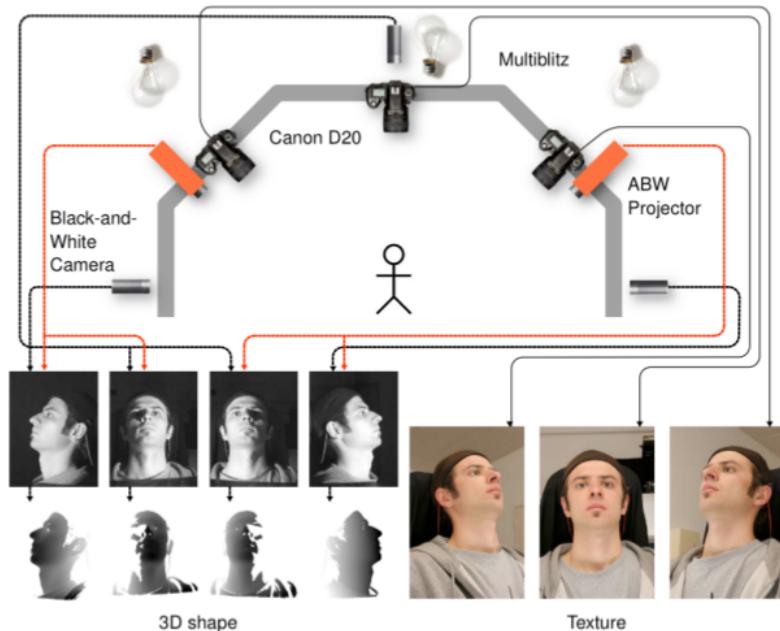


Figure 2.1: schema of the structured light 3D scanning system employed to record the facial scans. It consists of two projectors that project triangle gradient patterns on to a face from opposing angles, three black-and-white cameras used for the triangulation and three digital cameras which record the facial skin.

shells are obtained through triangulation. Each of the four black-and-white cameras 2.1 records the triangle gradient patterns - stripe patterns - emitted from the two projectors. The patterns allow for the extraction of depth information by the four neighbouring projector-camera-pairs using triangulation. The eyes and hair can not be captured due to their reflection properties. Therefore the shells only cover the front and the side of the head and have holes in the eyes and instead of parts

of the ears. The face textures are obtained from the color images. To obtain a



Figure 2.2: In a) the four separate shells of a scan are displayed in the b) they are merged in a one coherent mesh of the scan surface.

whole mesh the 3D shells first have to be cleaned and then merged. The scans are further manually annotated with a set of so called landmarks 2.3. Furthermore, they

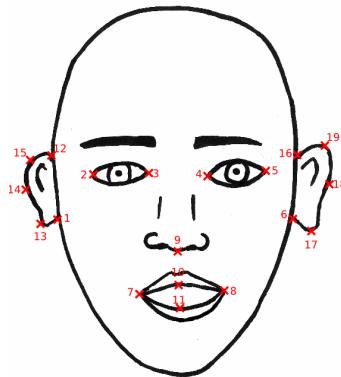


Figure 2.3: This schema shows the feature location of the landmarks that are given in the data for every scan.

are used to create 2D parametric curves that describe key regions of the face, i.e. the eyes. The line features will be fully introduced together with the registration pipeline. In the next chapter we will elaborate on the approach of using Gaussian Processes to solving the problem 3D face registration and then describe how to incorporate the annotations.

Gaussian Processes in 3D Face Registration

The first of our two objectives is to build a face registration pipeline. In this context we use a Gaussian Process as the registration algorithm. In this chapter we explain the theory behind the implementation of the registration pipeline. To begin with, we recapitulate the definition of stochastic processes and extend it to the definition of Gaussian processes. In the next step we introduce Gaussian Process Regression and finally explain how it can be applied 3D face mesh registration.

3.1 Stochastic Processes

In probability theory a stochastic process consists of a collection of random variables $\{X(t)\}_{t \in \Omega}$ where Ω is an index set. It is used to model the change of a random value over time. The underlying parameter time is either real or integer valued (discrete). A generalization of a stochastic process, which is defined over space on an index set of d-dimensional vectors, is called a random field.

3.2 Gaussian Processes

A Gaussian process is a stochastic process in which each finite collection $\Omega_0 \subset \Omega$ of random variables has a joint normal distribution. More formally, we define each random variable in the collection of random variables $\{X(t)\}_{t \in \Omega}$ to have a d-dimensional normal distribution if the collection $\{X(t)\}_{t \in \Omega_0}$ - for any finite subset Ω_0 - has a joint $d \times |\Omega_0|$ -dimensional normal distribution with mean $\mu(\Omega_0)$ and covariance $\Sigma(\Omega_0)$. If $\Omega \subseteq \mathbb{R}^d, d > 1$ holds, the process is a Gaussian random field. In the further proceedings we will use the term “Vector-valued Gaussian Processes” to refer to Gaussian random fields. Defining the random variables on an index set in a d-dimensional space, allows for spatial correlation of the resulting values, which is an important aspect of the algorithm discussed later on.

3.3 1-dimensional Gaussian Processes

In order to familiarize the reader with Gaussian Processes, we start explaining the concept on the basis of an index set \mathcal{X} of 1-dimensional inputs. To further simplify the concept it is possible to consider a Gaussian Process, or even any stochastic process, as a distribution over functions. Given a dataset, this circumstance allows us to look for inference in a space of functions. Each random variable $f(x)$ yields the value of some underlying function at a location $x \in \mathcal{X}$ in the index set of possible inputs. The index set is denoted by \mathcal{X} to stress that we are discussing Gaussian Processes defined over a space. In this function-space view a Gaussian Process at location x is thus $f(x) \sim GP(\mu(x), k(x, x'))$ defined by its mean $\mu : \mathcal{X} \rightarrow \mathbb{R}$ and covariance $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ functions which in turn are defined over the set of inputs. With $\mu(\mathcal{X}) = (\mu(x))_{x \in \mathcal{X}}$ and $\Sigma(\mathcal{X}) = (k(x, x'))_{x, x' \in \mathcal{X}}$ we obtain the full distribution of the process $GP(\mu(\mathcal{X}), \Sigma(\mathcal{X}))$. For the purpose of simplifying calculations we may assume that every random variable has zero mean without a loss of generality. (GP book)

Covariance Functions The key feature of a Gaussian Process is its covariance function also known as “kernel”. It specifies the covariance $\mathbb{E}[f(x)f(x')]$ between pairs of random variables for two inputs x and x' , allowing us to make assumptions about the input space by defining the spatial co-dependency of the modelled random variables. Note, that when assuming zero mean we can completely define the process’ behaviour with the covariance function.

A simple example of a covariance function is the squared exponential covariance function, defined by $cov(f(x), f(x')) = k(x, x') = \exp(-\frac{(x-x')^2}{2l^2})$. (derivation Rasmussen et al. p.83) It is possible to obtain different prior distributions by deploying different covariance functions. In our case, we use a stationary - meaning invariant to translation - isotropic, exponential covariance function - Squared Exponential Covariance Function (p. 38)

Gaussian Process Prior The specification of the covariance function implies that a GP is a distribution over functions, because the covariance matrix is defined over function evaluations. To illustrate this one can draw samples from a prior distribution of functions evaluated at any number of points, \mathcal{X}_* . The Gaussian Process Prior is solely defined by the covariance matrix made up of the covariance function evaluations of the input points.

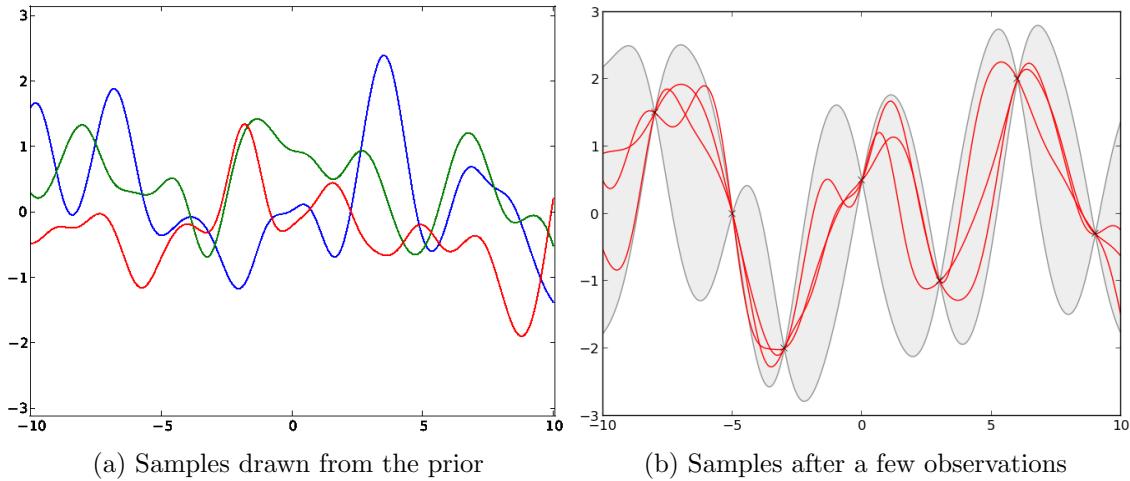


Figure 3.1: In figure a) three functions have been drawn from the GP prior, each has a sample size of 1000 points. Figure b) again shows three functions, but this time the prediction incorporates the information of random values observed at seven points in the input space. The grey zone depicts the area of two standard deviations from the posterior mean.

$$\Sigma(\mathcal{X}_*) = \begin{bmatrix} k(x_{*1}, x_{*2}) & \cdots & cov(f(x_{*1}), f(x_{*n})) \\ \vdots & \ddots & \vdots \\ cov(f(x_{*n}), f(x_{*1})) & \cdots & cov(f(x_{*n}), f(x_{*n})) \end{bmatrix} \in \mathcal{M}^{|\mathcal{X}_*| \times |\mathcal{X}_*|} \quad (3.1)$$

A sample is a random Gaussian vector $\mathbf{f}_* \sim \mathcal{N}(0, \Sigma(\mathcal{X}_*))$ containing a function value for every given input point. Plotting random samples above their input points is a nice way of illustrating that a GP is indeed a distribution over functions, see figure 3.1. The GP Prior forms the basis for inference in Gaussian Process Regression.

3.4 Gaussian Process Regression

The task of registering a 3D template mesh with a scanned target mesh can be treated as a regression problem in which the goal is to predict the deformation field for all vertices in the template mesh, given the displacements of the template landmarks on to the target landmarks. Trying to fit an expected function - be it linear, quadratic, cubic or nonpolynomial - to the data is, however, not a sufficiently elaborated approach to our problem. Fortunately, using a Gaussian Process disposes of the need to describe the data by a specific function type, the response for every input point is instead represented by a normally distributed random variable, in turn governed by the specification of the covariance function. In short, we are

modelling the space of possible regression functions through an infinite number of random variables in our input space. For the purpose of simplifying the concept of Gaussian Process Regression we will continue to stick with the exemplory case of 1-dimensional inputs.

Regression Problem Assume a training set $\mathcal{S} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ where $x \in \mathbb{R}$ and y_i is a scalar output or target. The task is now to infer the conditional distribution of the targets for yet unseen inputs x_* given the training data $p(\mathbf{f}_*|x_*, \mathcal{S})$

Noise-free Prediction First, we assume the observations from the training data to be noise-free so that we can fix the training data to these observations \mathbf{y} without complicating the model. The joint prior distribution with training \mathbf{f} and test \mathbf{f}_* outputs indicated is the following:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma(X) & \Sigma(X, X_*) \\ \Sigma(X_*, X) & \Sigma(X_*) \end{bmatrix} \right) \quad (3.2)$$

We obtain the posterior samples illustrated in 3.1 b) by conditioning the above joint Gaussian prior distribution on the observations $\mathbf{f}_*|\mathbf{f} = \mathbf{y}$ which results in the following distribution:

$$\mathbf{f}_*|X_*, (X, \mathbf{f}) \sim \mathcal{N} (\Sigma(X_*, X)\Sigma(X)^{-1}\mathbf{f}, \Sigma(X_*) - \Sigma(X_*, X)\Sigma(X)^{-1}\Sigma(X, X_*)) \quad (3.3)$$

Prediction with Gaussian Noise Model In most real world applications, however, observations from the training data are not free of noise. The landmarks clicked on the 3D face meshes, for example, can never be marked at the exact same feature location. These circumstances call for the incorporation of a noise model. We specify a simple additive i.i.d Gaussian noise model $y = f(x) + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ for every input vector x .

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma(X) + \sigma^2 \mathcal{I}_{|X|} & \Sigma(X, X_*) \\ \Sigma(X_*, X) & \Sigma(X_*) \end{bmatrix} \right) \quad (3.4)$$

The distribution - now conditioned on the noisy observations - is thus

$$\mathbf{y}_*|\mathbf{f} = \mathbf{y} \sim \mathcal{N} (\bar{\mathbf{y}}_*, \Sigma(\mathbf{y}_*)) \quad (3.5a)$$

where the mean depends on the observed training targets

$$\bar{\mathbf{y}}_* = \Sigma(X_*, X) (\Sigma(X) + \sigma^2 \mathcal{I}_{|X|})^{-1} \mathbf{y} \quad (3.5b)$$

whilst the covariance depends only on the input points

$$\Sigma_* = \Sigma(X_*) - \Sigma(X_*, X) (\Sigma(X) + \sigma^2 \mathcal{I}_{|X|})^{-1} \Sigma(X, X_*) \quad (3.5c)$$

We have thus defined a Gaussian Process Posterior Model that contains the 1-dimensional function space, fixed at the outputs of the training input points, between the training input points, as can be viewed in fig. 3.1 b).

3.5 Application to 3D Face Meshes

In this section we adapt the above presented GP regression theory to our case of 3D face mesh registration. The task at hand is to register a reference or template face mesh with a scanned face mesh. We therefore strive to predict a deformation field $\mathcal{D} : \mathcal{M} \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$ which assigns a displacement vector to every vertex in the template mesh. During registration we refer to the template as the moving mesh \mathcal{M} . Adding the deformation field to the moving mesh should then provide an accurate mapping to the target mesh \mathcal{T} and thereby perform the registration. Our objective is to register the template with the set of scanned faces.

Von skalarwertiger Funktion zu vektorwertiger Funktion, siehe Block!

Vector-valued Gaussian Processes In order to use Gaussian Processes to model deformation fields of three dimensional vectors as intended, there is the need for a generalization of the definition from the function-space view. The random variables $X_1, X_2, \dots, X_k, \dots, X_n$ are now d -dimensional vectors, yielding a vector-valued covariance function of the form $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^{d \times d}$ and $k(x, x') = \mathbb{E}[X_k(x)^T X_k(x')]$ and mean function $\mu : \mathcal{X} \rightarrow \mathbb{R}^d$. Mention vector field

Reference Mesh Prior As defined by the deformation field the output the regression problem is in \mathbb{R}^3 calling for the use of a Vector-valued Gaussian Process with random variables $\delta \subseteq \mathbb{R}^3$ where δ stands for deformation. After the template and target have been aligned ?? a Vector-valued Gaussian Process can be initialized by constructing a GP Prior of smooth deformations over all n vertices of the template mesh. For this purpose the covariance function has to be redefined to handle

3-dimensional vectors.

$$k \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} \right) = xy^T \in M^{3 \times 3} \quad (3.6)$$

Each covariance entails 9 relationships between the different components of the vectors, yielding a 3×3 matrix. The covariance matrix then grows to become:

$$\Sigma_{\mathcal{X}} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \in M^{3n \times 3n} \quad (3.7)$$

The template mesh is defined by a finite set of vectors $\mathcal{X} \in \mathbb{R}^3$ and a set of landmarks $L_{\mathcal{M}} = \{l_{\mathcal{M}1}, \dots, l_{\mathcal{M}n}\} \subset \mathbb{R}^3$. The mean vector μ is made up of the component-wise listing of vectors so that it has dimensionality $3n$. Setting the whole mean vector to zero, as discussed before, implies a mean deformation of zero and makes perfect sense in this setting, because we are modelling deformations of the template surface. The prior distribution over the template mesh is therefore defined as

$$\mathcal{D} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathcal{X}}) \quad (3.8)$$

meaning that a sample deformation field can be directly drawn from the prior distribution of the template mesh.

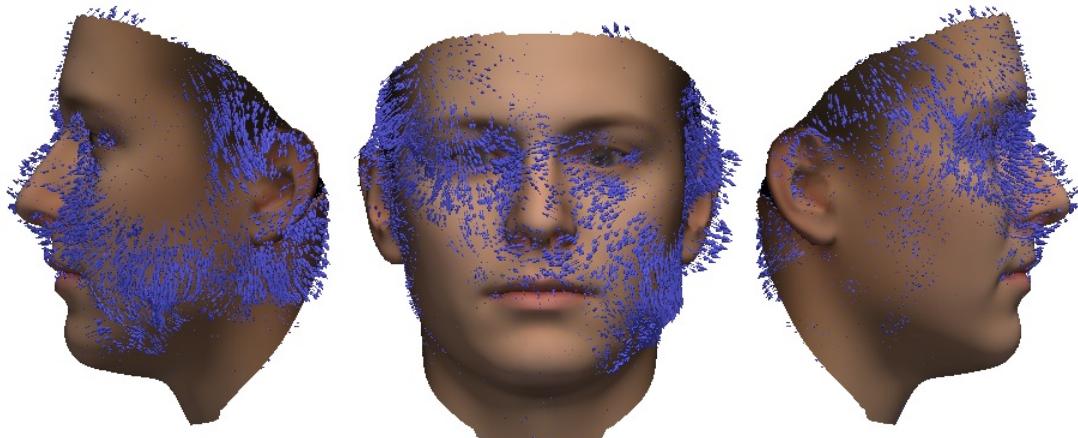


Figure 3.2: A sample **prior** deformation field - depicted with a set of **blue arrows** on the textured template face.

Reference Mesh Posterior The joint distribution of template mesh vertices and template landmarks is given by the following equation where \mathcal{D}_L is the vector returning the prior deformations of the template landmarks.

$$\begin{bmatrix} \mathcal{D}_L \\ \mathcal{D} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma(L_{\mathcal{M}}) + \sigma^2 \mathcal{I}_{3|L_{\mathcal{M}}|} & \Sigma(L_{\mathcal{T}}, X_*) \\ \Sigma(X_*, L_{\mathcal{M}}) & \Sigma(X_*) \end{bmatrix} \right) \quad (3.9)$$

We assume additive i.i.d Gaussian noise to model the ambiguity in the placement of the landmarks. The target landmarks also consist of a set $L_{\mathcal{T}} = \{l_{\mathcal{T}1}, \dots, l_{\mathcal{T}n}\} \subset \mathbb{R}^3$ and the corresponding deformation vectors $\mathcal{Y} = \{\mathbf{t} - \mathbf{m} | \mathbf{t} \in L_{\mathcal{T}}, \mathbf{m} \in L_{\mathcal{M}}\}$ - are defined by the difference vectors between the template and target landmarks. The goal is to find valid deformations through this set of fixed targets, analogous to the case of eq. 3.5a. These deformations are obtained by approximately fixing the joint distribution at the landmarks in the target mesh. In other words, the known respective displacements at the template landmarks co-define the distribution of possible deformations of all vertices in the template mesh. The posterior model

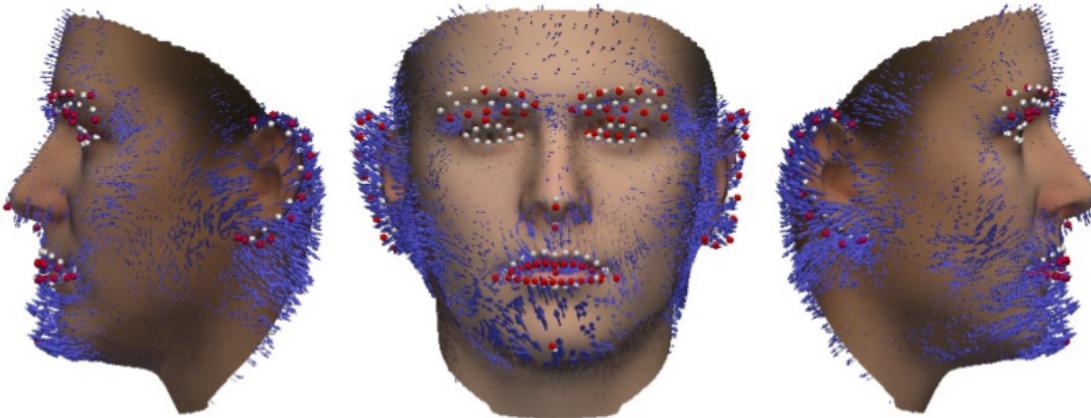


Figure 3.3: a sample **posterior** deformation field depicted with a set of [blue arrows](#) on the textured template face. The white spheres are the landmarks of the template mesh which are mapped on to the [red landmarks](#) of the target.

make same images without line features and display instead

is defined as the joint distribution of all template mesh points and the template landmarks, conditioned on the respective displacement vectors for every template landmark with added noise.

$$\mathcal{D} | \mathcal{D}_L = \mathcal{Y}_{\varepsilon} \quad (3.10)$$

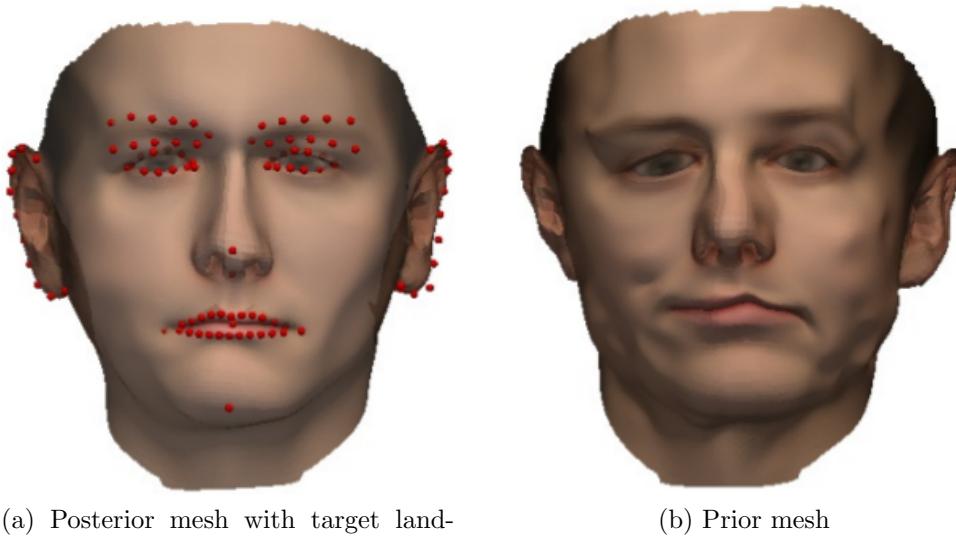


Figure 3.4: a) is a posterior mesh, obtained by adding a sample deformation field to the template. In comparison to the mesh in b) the mesh in a) is much more face-like. The reason for this is that the displacement of the landmarks has constricted the space of deformations to solely admissible deformations.

We now have defined a distribution over our template mesh. Sampling the conditional distribution and subsequently adding the resulting deformation onto the template creates deformed 3D surfaces of the template mesh which are fixed at the target landmarks 3.4.

3.6 Fitting & Optimization

In the previous paragraph we modelled the space of feasible deformations as a vector-valued Gaussian Process. We now want to perform registration by optimizing a deformation field according to the shape of the target face. In order to find the deformation corresponding to the optimal fit d_* a linear optimization is employed, where \mathbb{D} denotes the space of possible deformations in the posterior model.

$$d_* = \arg \min_{d \in \mathbb{D}} L[\mathcal{T}, \mathcal{M} \circ d] + \lambda R[d] \quad (3.11)$$

Therein the deformation vectors of the landmarks and the posterior process act as constraints in the regularization term R . Minimizing the loss function L - for example the mean squared error (MSE) - on the target and the deformed template should provide us with the optimal deformation under the given constraints.

The problem with the above approach to optimization is that the deformations are modelled by a distribution and are therefore non-parametric. In order to properly perform optimization, however, we need a representation that can be controlled through / is governed by a set of parameters. For this purpose we would like to specify a parametric model $\mathbb{M}[\alpha](x) = \mu(x) + \sum_{i=1}^n \alpha_i \lambda_i \phi_i(x)$ which is defined over a finite set of basis functions. The question remains how the GP Posterior Distribution can be expressed as a linear model. Fortunately, the answer lies in Mercer's Theorem. It states that a symmetric, non-negative definite, continuous function k can be expressed by the following linear combination of a countable sequence of functions $\{\phi_i\}_{i \in \mathbb{N}}$

$$k(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x')^T \in \mathbb{R}^{3 \times 3} \quad (3.12)$$

Through the theorem we can define the covariance function of the Gaussian Process Posterior in an infinite dimensional space made up of its eigenvalues λ_i and eigenvectors $\phi_i : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ which are functions in this case. This infinite basis of functions corresponds to all possible deformations at a specific vertex in the template mesh. By evaluating a basis function at every vertex we receive a basis deformation field. We can further use Mercer's theorem to obtain the sought parametric model by estimating the covariance function. In order to achieve this, a low-rank approximation of the covariance function - in terms of the first leading terms of its Mercer expansion - is performed $k(x, x') = \sum_{i=1}^n \lambda_i \phi_i(x) \phi_i(x')$. For this to work, the modelled deformations have to be sufficiently smooth, because we are only using a fixed number of basis functions corresponding to the highest eigenvalues and disregard the rest of basis functions. Smoothness is guaranteed by the square exponential covariance function. We can further express the deformation at a given mesh vertex by a linear combination of the selected basis functions and their eigenvalues with $\alpha_1 \dots \alpha_n \in \mathbb{R}$ as the linear parameters

$$f(x) = \sum_{i=1}^n \alpha_i \lambda_i \phi_i(x) \quad (3.13)$$

In effect, we have formulated the Gaussian Process Registration problem in parametric form. Next, we want to minimize the residuals for all the points on the template surface according to optimal parameter values α_i

$$\arg \min_{\alpha \in \mathbb{R}^n} \Sigma_{x_i \in \mathcal{M}} (f(x_i) - \varphi_T(x_i))^2 \quad (3.14)$$

where $f(x_i)$ is the deformation function and $\varphi_T(x_i)$ returns the nearest point on the

target mesh. The optimization problem therefore becomes

$$\alpha_* = \underset{\alpha \in \mathbb{R}^n}{\arg \min} \quad L[\mathcal{T}, \mathcal{M} \circ \mathbb{M}[\alpha]] + \lambda R[||\alpha||^2] \quad (3.15)$$

For finding the parameters' minimum any standard optimization scheme can be employed.

Registration Pipeline using Line Features

In this chapter we describe how the Vector-valued Gaussian Model is utilized in our implementation of a 3D Face Registration Pipeline. To enhance the registration outcome of this pipeline we use additional annotations to extend the prior correspondences given initially by the landmarks described in ???. The additional annotations are parametric curves, we call line features, that mark key regions on the 2D colored images of a face scan. In the following we provide a description of line features and their use as well as a description of the registration pipeline.

4.1 Line Features

Line features serve the purpose of augmenting the quality of registration by initiating it with a set of corresponding lines. They define correspondence in feature regions of the given face scans, i.e. the eyes and ears and lips. These regions have to be accurately mapped on to one another for the registration to produce acceptable results. For this purpose we mark their contour lines with parametric curves.

We have 8 line features given for every scan we want to register. These have been marked on three images of every face, see 2.1. They are made up of a set of segments, each of which is modelled with a **Bézier curve** of a specified order, see 4.1. Bézier curves are often used in Computer Graphics for modelling smooth curves of varying order. Given a set of control points $\mathcal{P} = \{P_0, P_1, P_2, \dots, P_n\}$ the Bézier curve through these points is given by

$$C(t) = \sum_{i=0}^n P_i B_{i,n}(t) \quad (4.1)$$

where $B_{i,n}(t)$ is a Bernstein polynomial

$$B_{i,n}(t) = \binom{n}{i} (1-t)^{n-i} t^i \quad (4.2)$$

and $t \in [0, 1]$ is the curve parameter. The line features consists of multiple segments of Bézier curves. Due to the nature of the objects depicted there are open as well as closed curves, see fig. ??.



Figure 4.1: Line Features of the left eyebrow and eye, consisting of b^ezier curves defined by visible control points (white). The line features are marked on 2D color images taken during scanning procedure.

4.2 Sampling 3D Points from 2D Line Features

The line features provide us with additional prior information about the nature of the deformation field. In order to incorporate the line features in the Vector-valued Gaussian Model, we transform them to additional point-wise information. The points are obtained by sampling the line features at discrete intervals resulting in a set of additional landmarks $L_{Add} = \{l_1, \dots, l_N\}$. These define the mapping $\Omega : L_{Add\mathcal{M}} \rightarrow L_{Add\mathcal{T}}$ of the line features in the template face mesh on to those in the target face mesh.

Establishing correspondence for a pair of ears for example is very difficult, even for experts, because of their very different topology. The structure of the concha is very complex. For this reason, we define correspondence between lips, eyes and ears through the contour lines of their rims. Further modelling the ratio between the

different parts of these contours would again be tedious. Therefore, we simplify the assumptions for similarity to an equidistant parametrization of the line features so that the mapping Ω is approximately plausible. In effect, when a curve is sampled at n points, these n points are all at equal parametric intervals.

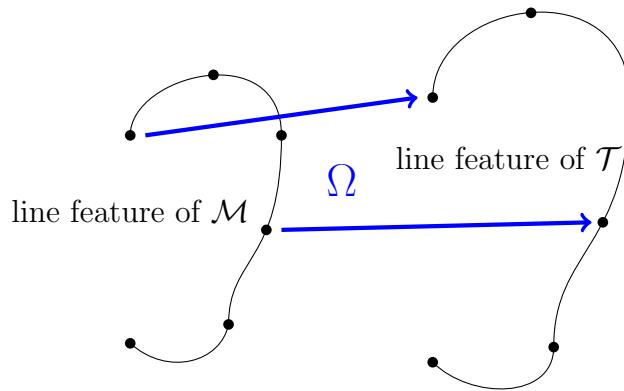


Figure 4.2: Mapping of equidistant samples of **ear** line features from the reference (left) on to the target (right)

Arc Length Parametrization

The underlying parameter $t \in \mathbb{R}$ of a bézier curve segment of a single line features is dictated by the curve's velocity. In order for us to be able to sample the line feature in equidistant intervals, however, the parameter has to be linear in respect a the curve segment's length. In consequence, we have to compute an equidistant parametrization for each of the curve segments of a line feature before sampling it. The solution to this problem is based on the respective curve's arc-length, which is defined as the length of the rectified curve. The underlying parameter t must then correspond - at every point of the curve - to the ratio between the length of a fraction of the curve l and the total curve length L : $t = \frac{l}{L}$.

In theory It is possible to get the arc length $L(t) = \int_{t_0}^{t_1} |C'(t)| dt$ for given parameters t_0, t_1 where $C'(t)$ is the derivative of the curve $C : t \in [0, 1] \rightarrow \mathbb{R}^2$. We, however, want to find a reverse mapping from the length of a fraction of the curve l to the curve parameter $t = L^{-1}(l)$. Instead of deriving this function analytically we decided to use a numeric approximation.

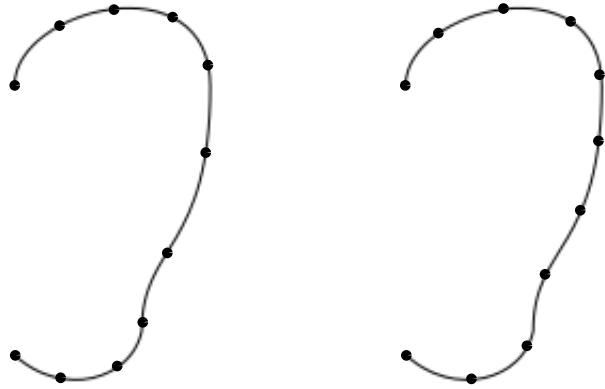


Figure 4.3: a simplified - the depicted **ear** line feature is treated as one sole b  zier segment - illustration of the difference between b  zier (**left**) and equidistant (**right**) parametrization.

In practice As we are not in need of a subpixel resolution, we can skip the formal math and use a lookup table to compute the arc-length. First, we calculate a large number of points on each segment of the curve using the parametrization of the corresponding B  zier curve. For each sampled point, we save a key-value-pair to a new entry of the lookup table. The approximate distance of the point from the origin of the segment is inserted as the key, while the point's coordinates are stored as the value. The distance is approximated by summing up the euclidean distances

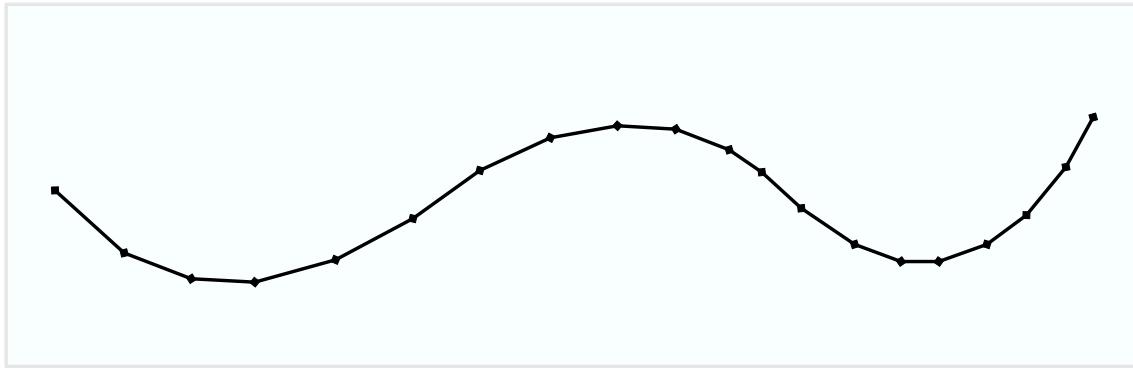


Figure 4.4: Visualization of euclidean distances between points computed with a b  zier curve parametrization. This is an arbitrary line feature used solely for demonstration purposes.

of each point to its respective predecessor starting from the origin. The resulting lookup table contains the approximate distances and coordinates of a large number

of points from the origin of a curve segment. Assembling the segments' lookup results in the lookup table of the whole curve, where the last key represents the curve's arc-length. Through this procedure we have gained an indirect mapping from the length of a fraction of the curve to the parameters of the Bézier curve segments via the point coordinates.

In a second step, the curve can easily be sampled by computing the length of parametric intervals $\frac{L}{N}$ for a specified number of points N .

$l = k \cdot \frac{L}{N}$ returns the current length of the curve for the sampling point of index k , where $k = [0, \dots, N]$ for open curves and $k = [0, \dots, N - 1]$ for closed curves. To get the point coordinates for a fraction of the curve, we now simply perform a binary search on the lookup table for its approximated distance from the origin l . We choose the index which returns the coordinates for the exact fraction length and if this requirement can't be met the index with the next smaller length. The returned point coordinates are used as the approximate coordinates of the sought sample point.

3D Mesh Projection of Sampled Points

Having implemented arc length parametrization it is possible to draw an arbitrary amount of sample points $p \in \mathbb{R}^2$ from the 2D line features. These are defined as a set of points $S \subset \mathbb{R}^2$. Our goal is to use these additional points as landmarks describing the features on the 3D mesh of the corresponding face. To achieve this we have to project the sampled points of each line feature on to the corresponding face mesh in order to obtain their approximate 3D representation. We use this method because we have no information on the depth of the respective line feature. In the camera model - derived from the calibration of the scanner - the image is located on the viewing plane or viewport along the z-axis from the focal point. The idea is to follow the direction of a point on the line feature - which lies in the viewport - from the focal point and obtain a projection on to the face mesh along this direction. We want to use the discrete mesh of a given face scan to create a new vertex that is the most accurate representation of the sample point on the line feature. In order to find vertices with similar directions to the sample direction, we save all the distances of mesh vertices in a list for which a similarity measure is higher than a specified threshold. As the similarity measure we use the dot product of normalized vertex and sample directions. We then select the distance of the vertex with the highest similarity. Finally, we project the distance of this vertex onto the direction of our

sample point and thereby obtain an approximation of the points position in the mesh. A new vertex is created at this location.

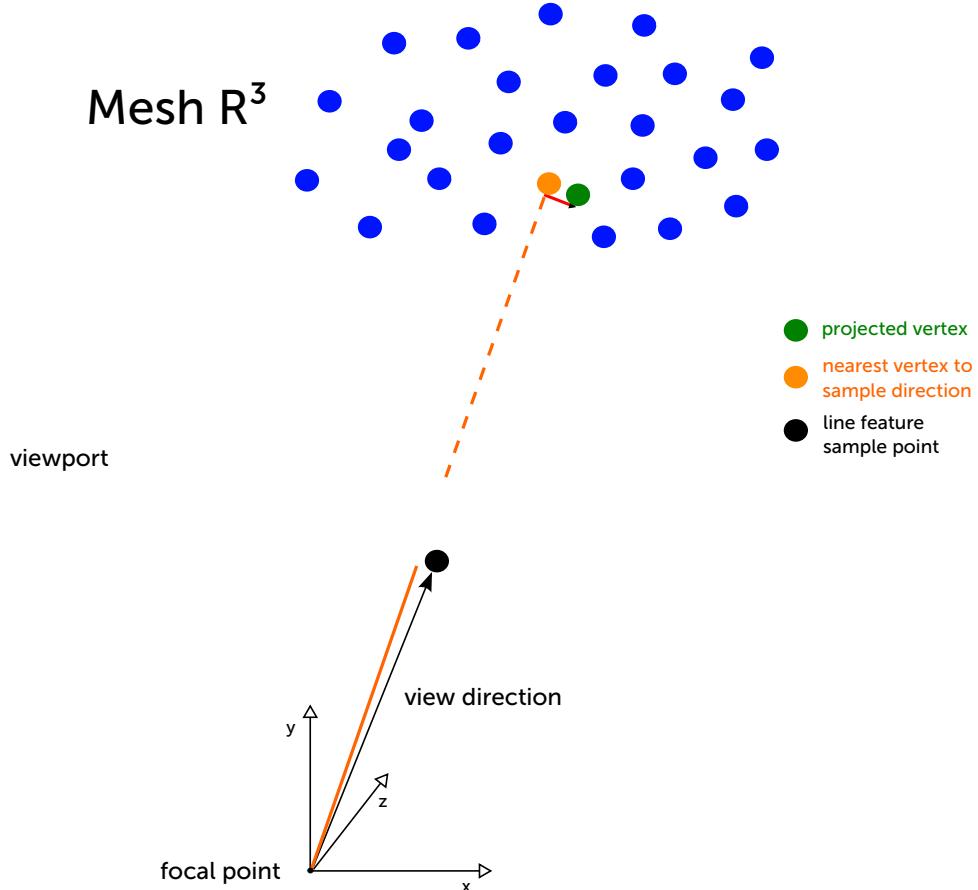


Figure 4.5: shows the projection of a sample point - on a 2D line feature - from the viewport onto a triangulated 3D mesh. The distance vector of the vertex with the most similar direction vector (orange) is projected (red) onto the direction of the sample point (black arrow), resulting in the projected vertex (green)

Preparing the Template Mesh

For the line features projected on to the scans to be of use, line features also have to be marked on the template mesh. We use the mean mesh of a 3D Morphable Model [<http://gravis.cs.unibas.ch/publications/2009/BFModel09.pdf>] as the template mesh.

We render three different images, similar to those taken during the scanning of the faces, for marking the line features. The 2D line features can then be projected

back onto the mesh using the rendering parameters. With the projection we obtain the corresponding, equidistant samples of the line features on the template. They obtained additional prior information is used in the registration algorithm to define the posterior deformation model and to produce better registration results.



Figure 4.6: Mean mesh of the Basel Face Model



Figure 4.7: Overlap of textured mean template with a target scan (gray surface)

4.3 Rigid Mesh Alignment

Before we can start creating Gaussian Models, we first have to ensure that each target is aligned with the template. After all, we want to model the variability of different faces without incorporating an additional offset. We therefore have to perform a rigid transformation consisting of a translation and a rotation to align the template and target meshes according to their landmarks. To compute the transformation we use the set of landmarks pairs whose identifiers exist for both meshes.

The targets are clipped at the neck and around the ears where the scanner has produced artifacts. The computed transformation is then applied to all vertices of the respective face scan to align it with the template. The aligned mesh pair serves as the starting point for the actual registration, see fig. 4.7.

We are now set for the actual registration involving the computation of the Prior and Posterior Gaussian Models and the Fitting of the Posterior Model to a face scan. For the definition of the Gaussian Process distributions we use statismo ?? . It is a framework for PCA based statistical models. These are used to describe the variability of an object within a population, learned from a set of training samples. We use it to generate a statistical model of the template mesh equaling the parametric representation of the Gaussian Process Posterior described in sec. ?? .

4.4 Prior Model

After aligning the target with the template mesh, we define a distribution of possible deformations of the template mesh, as described in 3. This Gaussian Process (GP) Prior distribution is represented as a statistical model. This allows for samples to be drawn from this representation. These samples are 3D face meshes that are the result of the deformations generated by the GP Prior and added to the template mesh for visualizations, some of these prior faces are displayed in fig. ?? . The deformations are up until now defined solely on the ground of the covariance of the template mesh points.

4.5 Posterior Model

The deformations generated by the GP Prior are of course not admissible. We receive admissible deformations by adding the target landmarks and the additional set of sampled line features to the GP Prior distribution. Using statismo we create a statistical model, which is partially fixed at the target landmarks, from the template mesh. Now we have incorporated the given prior information, we obtain a model that leads to meshes that are more face-like.

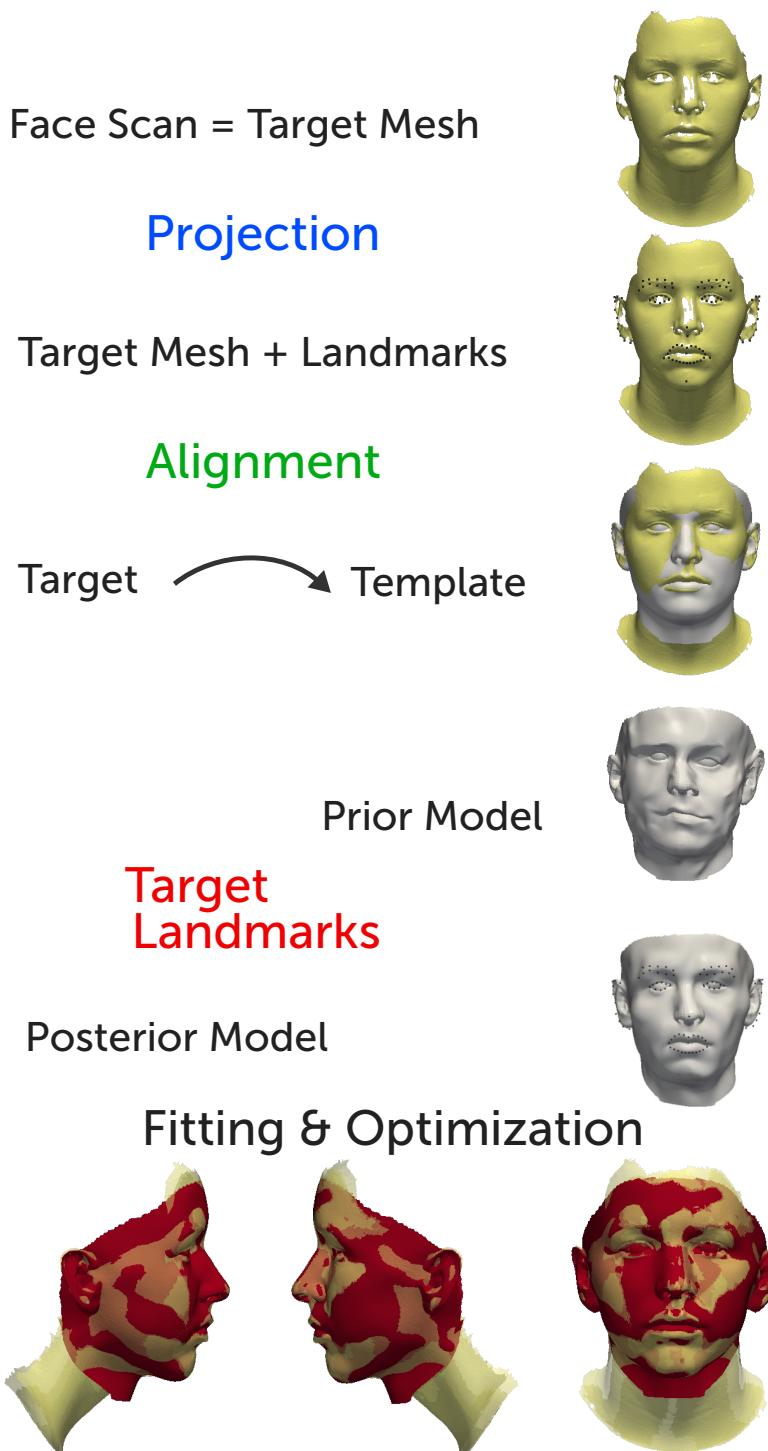


Figure 4.8: An overview of the different stages of the registration pipeline with an exemplary visualization of face meshes at every stage.



(a) sample of the GP Prior

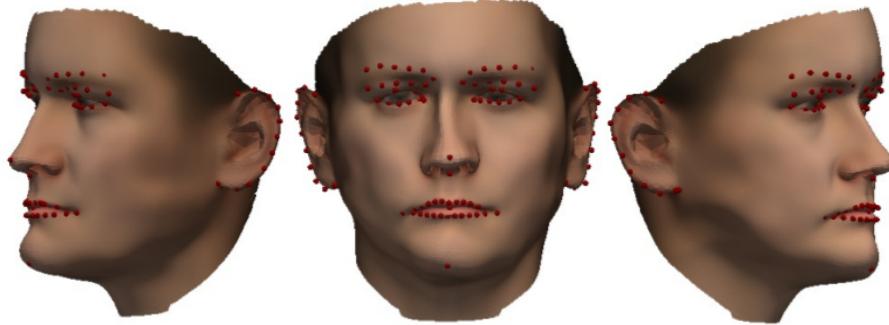


(b) sample of the GP Prior

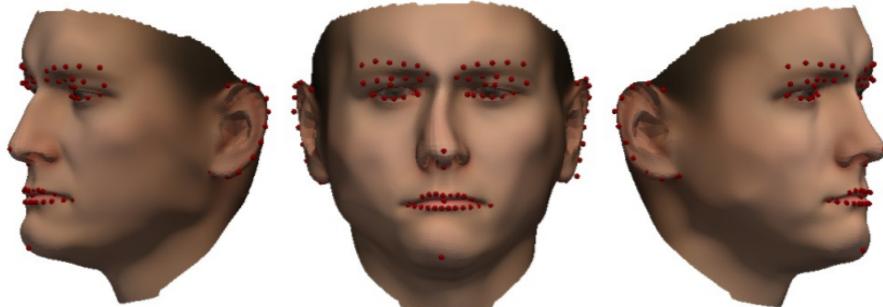
Figure 4.9: Prior faces computed from samples drawn from the Gaussian Process Prior Model. The deformations are not truly face-like, because they are only modeled through the covariance of the template mesh vertices.

4.6 Optimization of the Posterior

In order to optimize the respective GP Posterior Models of the template according to the target scans, we obtain the parametric representation of the GP Posterior Model using statismo. We then use the limited-memory BFGS optimization build into the Insight Toolkit (ITK). Optimize parameters for the set of chosen basics functions in the parametric model. First we used the Mena Squared Error as the loss function. However, this estimator did not yield desirable results, as can be seen in ???. The reason for the bad performance of MSE is that after the alignment of template and target mesh, the template protrudes over the target on both sides of the head, consult in fig. ???. Performing optimization as described in ?? using the simple MSE as a distance measure between the template and target mesh penalizes the protruding regions of the template with a strong gradient towards the rims of the template and therefore causes strong distortions.



(a) sample of GP Posterior



(b) sample of GP Posterior

Figure 4.10: Posterior faces computed from samples drawn from the GP Posterior. The resulting faces are considerably more face-like than those in fig. ??

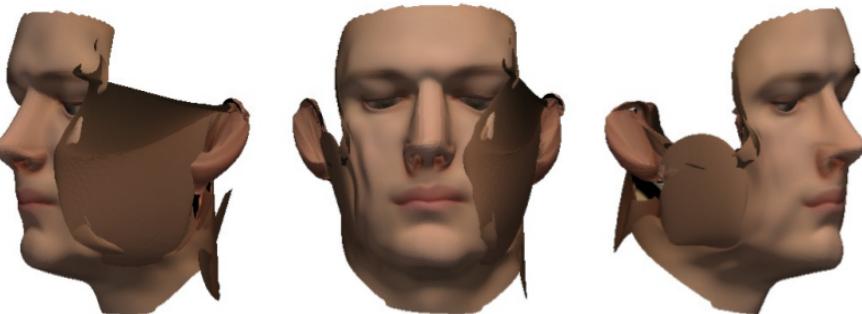


Figure 4.11: A fit of the template mesh obtained using the Mean Squared Error as the distance metric. The deformation in the regions where template and target do not overlap cause major distortions in the resulting fit.

4.7 Robust Loss Functions

To overcome the problems described above, we tried a range of different robust estimators, namely the Tukey, Huber, and Fair estimators. The advantage of these estimators lies therein that they are less sensitive to outliers, reducing registration artifacts considerably. Outliers are in this case template mesh points that are farther

away than a certain threshold from the next point on the target mesh. However, these techniques require parameter tuning to first which produce reasonable/acceptable visual results.

Fair

$$\rho(x) = c^2 \left[\frac{|x|}{c} - \log\left(1 + \frac{|x|}{c}\right) \right] \quad (4.3a)$$

$$\psi(x) = \frac{x}{1 + \frac{|x|}{c}} \quad (4.3b)$$

Huber

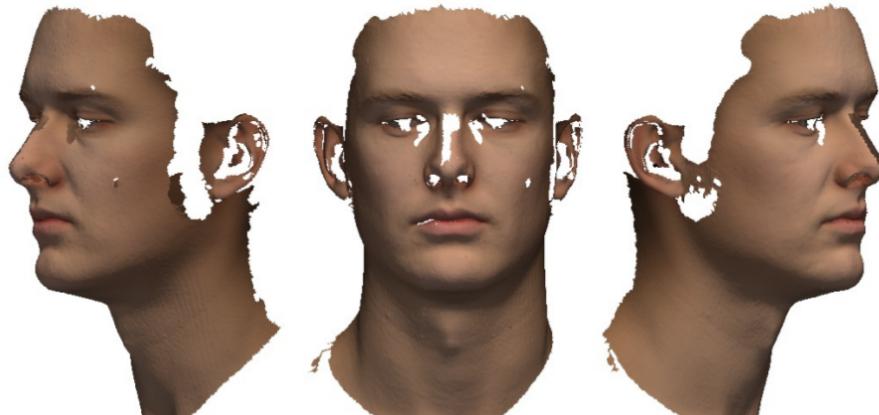
$$\rho(x) = \begin{cases} \frac{x^2}{2} & \text{if } |x| < k \\ k(|x| - \frac{k}{2}) & \text{if } |x| \geq k \end{cases} \quad (4.4a)$$

$$\psi(x) = \begin{cases} x & \text{if} \\ ks \operatorname{sgn}(x) & \text{if} \end{cases} \quad (4.4b)$$

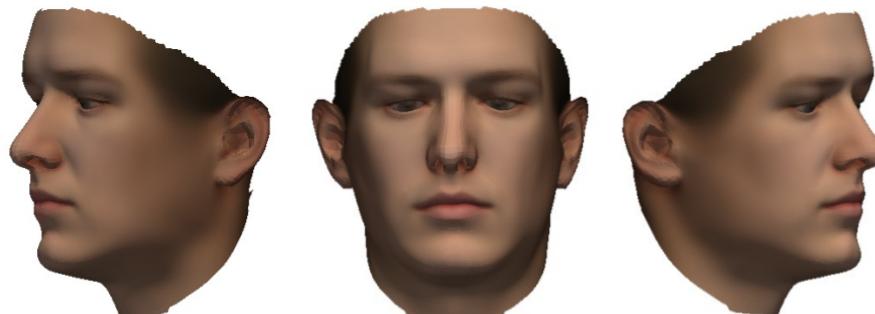
Tukey

$$\rho(x) = \begin{cases} \frac{c^2}{6} \left(1 - \left[1 - \left(\frac{x}{c} \right)^2 \right]^3 \right) & \text{if } |x| \leq c \\ \frac{c^2}{6} & \text{if } |x| > c \end{cases} \quad (4.5a)$$

$$\psi(x) = \begin{cases} x \left[1 - \left(\frac{x}{c} \right)^2 \right]^2 & \text{if} \\ 0 & \text{if} \end{cases} \quad (4.5b)$$



(a) target with mean texture applied



(b) optimized mean posterior fit

Figure 4.12: The fitting result in a) was obtained with the Tukey estimator, parameter $c=7.0$. The texture of the mean mesh has been applied to the target to make the meshes comparable. The fit in a) has considerable likeness to the target b), although it lacks expressiveness

Results

5.1 Line Feature Sampling Inaccuracy

The faces in the set of given scans all contain large holes in the regions of the eyes and the ears. Some also have holes on top of the nose. Due to this circumstance the vertices with the most similar directions to that of the sample point are farther away a suboptimal location. Consequently, the projected sample points of the line features may be off target in cases where the holes are very large. In effect, the projected line is distorted, as can be seen in 5.1 on the right side. On the different

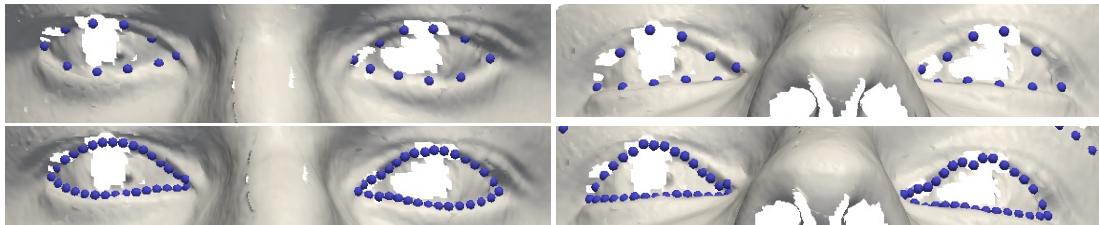


Figure 5.1: Line Features

data sets quality of the projected line features varies significantly, because each scan has different holes. If a scan contains large holes one can say overall that the distortion of a projected line features increases with the amount of sampled points. On different data sets the performance of the projection of the line features for a large number of samples, i.e. 30, varied significantly. An easy workaround to this problem is to reduce the amount of sampled points. By using only 5-10 sample points per curve most datasets rendered near perfect results. However, when the holes are too large this workaround also fails, slight distortions can be made out in 5.1 top right. If the distortions are not too big they are accounted for by the additive Gaussian Noise in the joint distribution ???. However, as long as the method is dependent on the data from the scans - the size of the holes in the meshes - it lacks generality and

generality is exactly the basis for feasible and reproducible registration results.

5.2 The effect of Line Features

In this section we illustrate the effect of the line features on the registration results. For this purpose we compare the feature regions of registered templates with and without line features with the respective target. We use up-close images of the various feature regions for a qualitative comparison. Target with mean texture projected on to it.



Figure 5.2: A comparison of the fitting results concerning the eyes. The first image displays the eyes in the registered template without line features. The second image has been registered with line features. As one can see the eyes in the top image are in the right position, because the displacement of the corners is defined by the given landmarks. However, compared to the middle image the contours of the eyes are very different. The incorporation of line features in the middle image has lead to a much better correspondence with the eyes of the target mesh, displayed at the bottom.

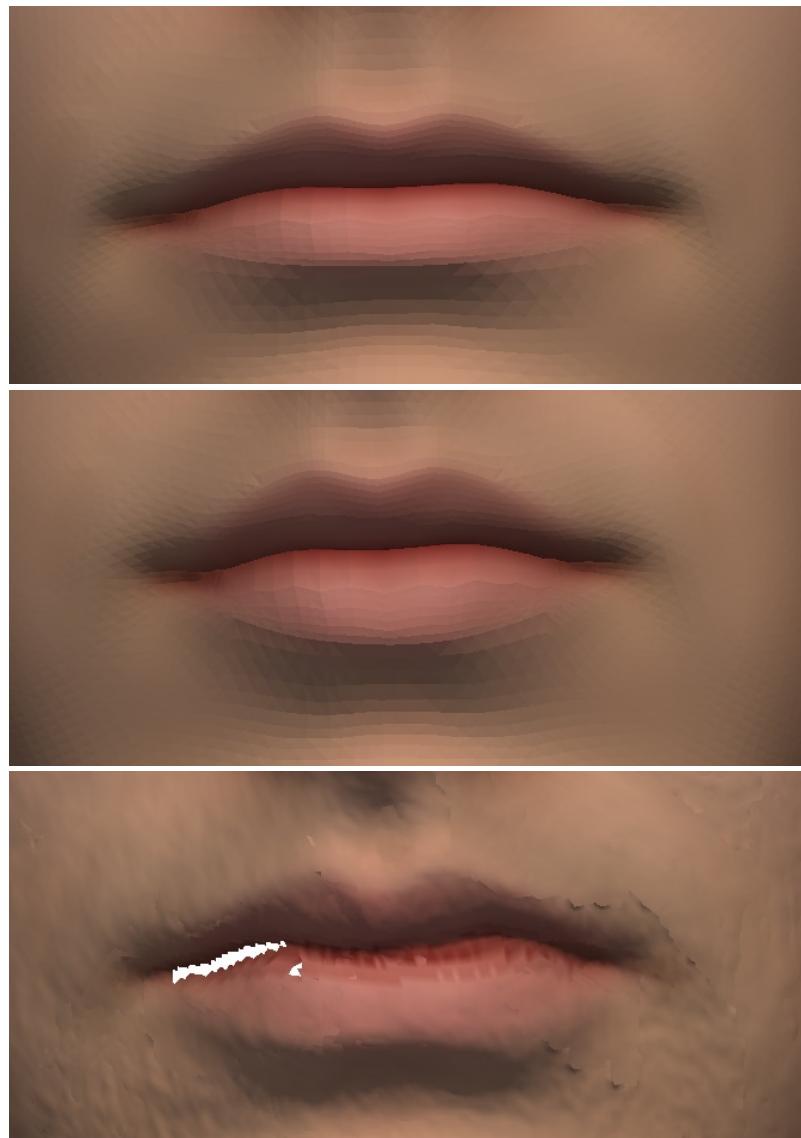


Figure 5.3: At the top again is the mouth fitted without line features and beneath the one with line features. The lips in the top image are considerably thinner although the position of the mouth is of course correct due to the displacement of the landmarks in marked at the corners of the mouth. The lips in the middle image are fuller and match the mouth of the target better in this respect. However, the lower lip of the target appears to be bimodal while the lip of the fit is still unimodal as in the template mesh.



Figure 5.4: Ears

Figure 5.5: A comparison of the fitting results of the ears. The image on the left hand side is the fit without line features and image on the right hand side is the fit using line features. The target is displayed between the two. The ear on the left is narrower than the ear of the target and the ear of the fit using line features.

Visualized Distance

A further way to evaluate the matching of the contour lines is to analyse distance maps. These are obtained by taking the differences of the nearest points in the deformed template and the target mesh. Distance maps to show distance for semantic correspondences

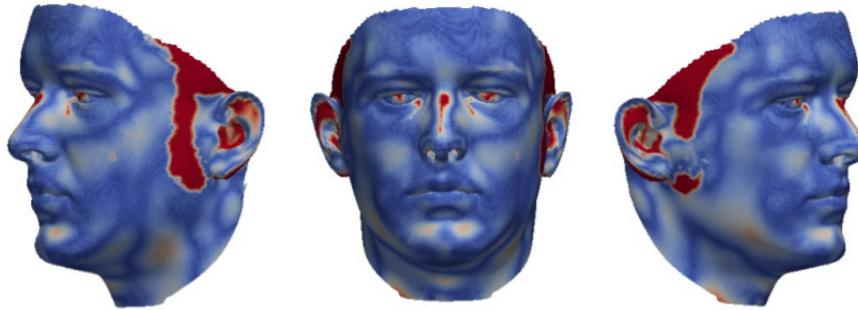
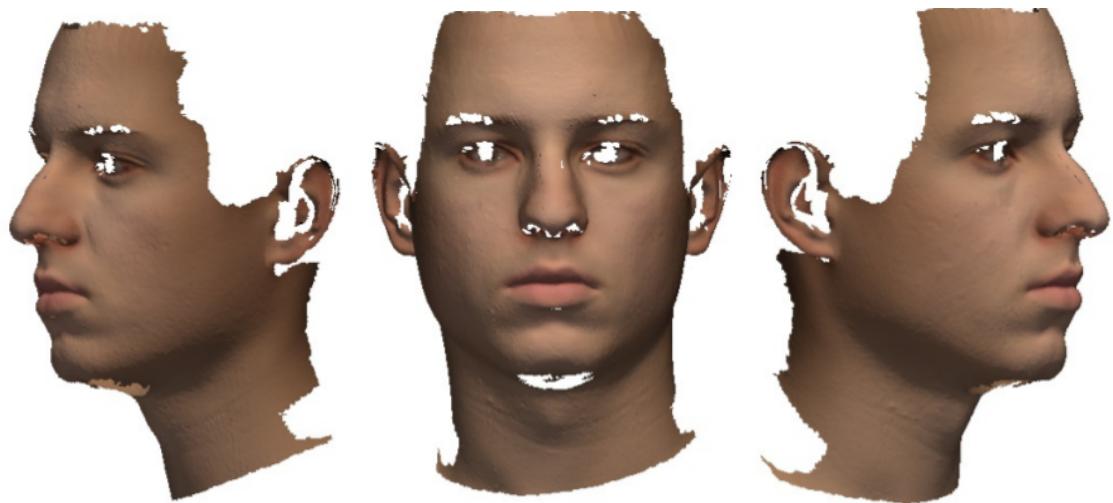
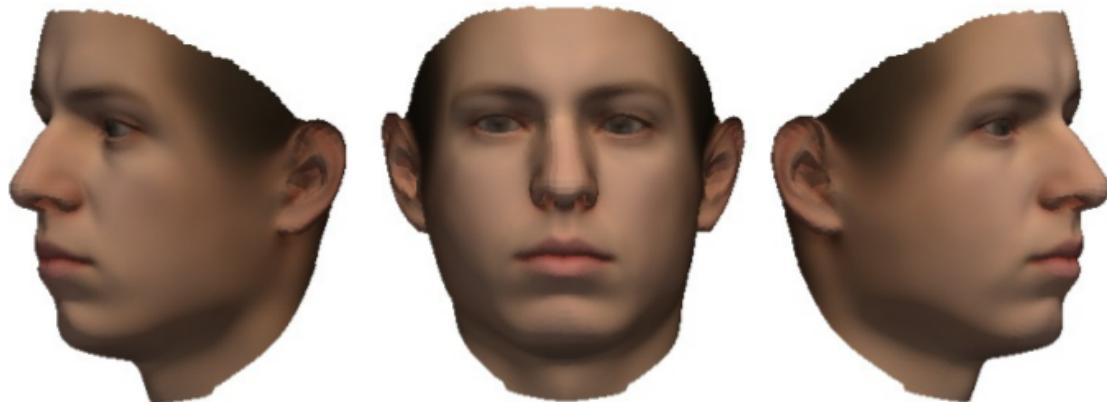


Figure 5.6: Distance face of fit displayed in fig. ??

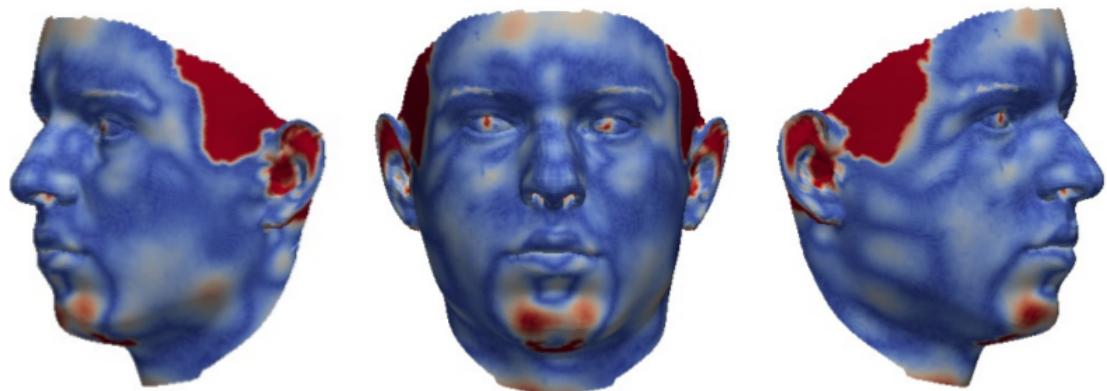
in range $[0, 3\text{mm}]$. Except for the holes in the target mesh all vertices in the template have less than 3mm distance to their nearest neighbours on the target mesh. The distance map doesn't show semantical correspondence.



(a) target mesh with projected template texture

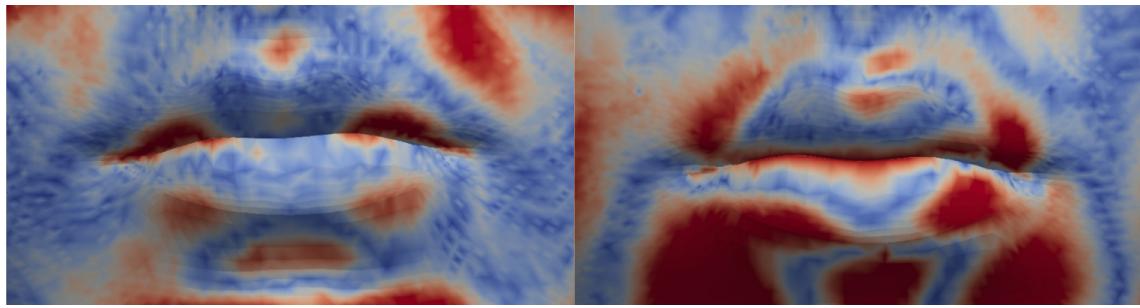


(b) deformed template fitted to the target



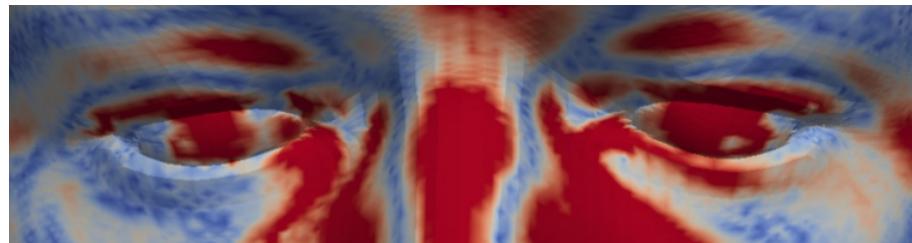
(c) distance map

Figure 5.7: This figure shows the fit of the template mesh to a different target face and the distance map in a range of 0-3 millimeters.

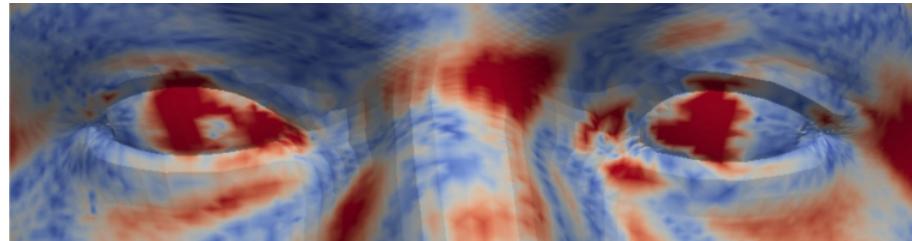


(a) Distance image of the mouth of fit in ?? (b) Distance image of the mouth of fit in ??

Figure 5.8: Both distance images are on a scale of 0-1mm [blue, red]. The fit of the lips is better in a) where the area along the contour lines of the lips is mostly light blue. In b), however, the distances in the region of the lips often exceeds 1mm and quality of the fit is not as good.



(a) 00029 eyes distance



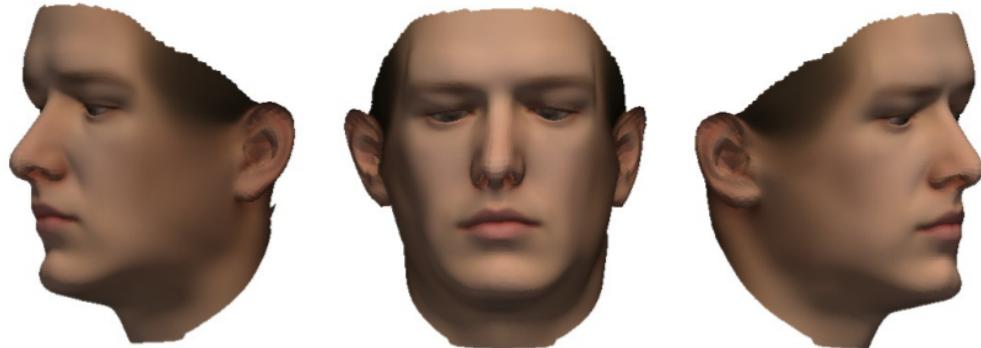
(b) 00303 eyes distance

Figure 5.9: Again both distance images are shown on a scale of 0-1mm [blue, red]. The fit in a) has a lot red areas, because the target mesh has large holes in the region of the eyes and along both sides of the nose and on top of the nose. The bottom inner contour of the left eye is however matched very well. The same can be said for b). Here, the bottom inner contour of the right eye is also matched fairly well.

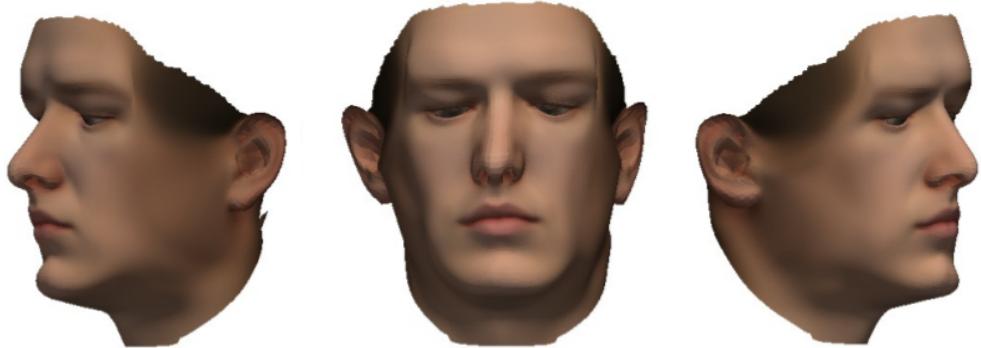
Caricatures

Caricatures of the deformed template are another way of inspecting the quality of the registration results. Points in the same feature regions should be displaced in a very similar manner. Therefore, by adding the deformation field to the target multiplied by a scaling parameter exceeding 1, these regions should ideally become

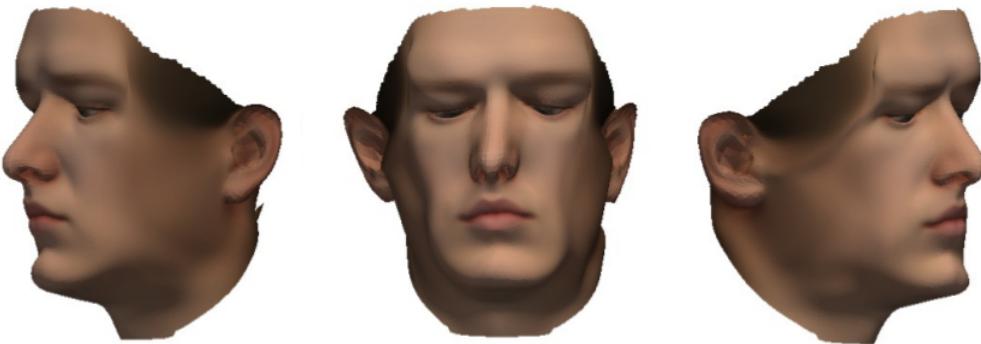
more prominent in the resulting mesh. The same rule applies to distortions/artifacts which also become more visible with an increase in the scaling factor.



(a) scaling factor 1.3

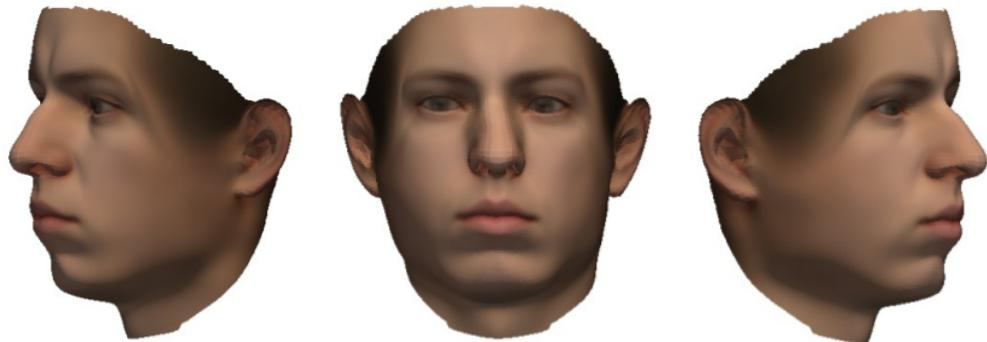


(b) scaling factor 1.6



(c) scaling factor 2

Figure 5.10: Caricatures of the fit in fig. ?? with different parameter values. With an increasing scaling factor the face is twisted more to the left, with the mouth and the nose while the ears become pointier. The eyelids are shut more and more with each step. Also a fold begins developing to both sides of the head. This artifact is caused by the increase of the minimal displacements of the template mesh vertices which have no corresponding vertices on the target mesh and are not completely ignored by the Tukey estimator.



(a) scaling factor 1.3



(b) scaling factor 1.6



(c) scaling factor 2

Figure 5.11: Caricatures of fit in ?? for three different scaling factors. Here, the exaggeration of a slightly hooked nose can be observed. In this case the face is bent slightly to the right and lips are pressed together and forward. The folds on both sides of the head are prominent than in the other example.

Discussion & Future Work

In this thesis we discussed a new approach to 3D face registration with Gaussian Process Regression incorporating line features. As demonstrated in chapter 5 the quality of the registration was significantly enhanced through the use of line features. The registration pipeline yields deformed templates that provide a good matching of the feature regions with those of the respective targets. This is largely due to the attempted mapping of feature regions between template and target using the line features. For transforming the 2D line features we were obliged to use an approximative approach where the different appearances of the holes in the respective scans carry the risk of influencing the results. The ideal case would of course be to find a parametric 3D representation of the line features that could be plugged directly in to the registration process. The constructed registration pipeline is in its current state a good basis to further build upon and refine the registration algorithm which already yields fits of the template mesh resulting in profound likeness with the respective targets. However, there is still room for improvement, particularly concerning the expressiveness and the exact modelling of the feature regions of the obtained registration results. The results lack in accuracy, because of the necessitated use of a robust estimator as the loss function which is due to the overhang of the template mesh and the face scans. This overhang also causes artifacts despite the fact that a robust estimator is deployed. An improvement of the registration scheme therefore would be to first detect the vertices in the template mesh corresponding with the holes in the target mesh by registering the template with it. With the outlier regions detected, the Mean Squared Error could be used to achieve very accurate registration results without the appearance of artifacts. The proposed approaches will be subject of future work on this topic.

In conclusion, the presented approach to 3D face registration successfully incorporated line features to establish correspondence which lead to more expressive and accurate registration results.