Using line features for 3D face registration

Fabian Brix

18.06.2013

Abstract

In this bachelor thesis we attempt to modify the existing face registration pipeline for the morphable face model of Prof. Thomas Vetter by using a registration algorithm developed by PD Marcel Lthi at the University of Basel. ALTERNATIVE: In this bachelor thesis we discuss the construction of a face registration pipeline. The using an algorithm based on a vector-valued gaussian process and at the same time attempting to ensure registration quality through the use of contours marking important parts of the face - referred to as line features.

The algorithm is capable of mapping any two shapes on to one another. All that is needed is a set of corresponding points on the two shapes. Different constraints to the displacement field can be applied through regularisation.

The aim of this bachelor thesis is more specifically to apply this general algorithm for point correspondences to scanned face data, that is to implement feasible registration of face scans onto the mean face of the morphable model. In order to achieve this we mark important parts of the face meshs not only with point landmarks, but also structures and organs (eyebrows, eyes, ears) with lines - line features - and thereby to create further correspondences for the algorithm to perform better by. Instead of using sparse points of key features points of the face we mark complex features, e.g. the eyes, with contour lines - line features in order to create further correspondences

These line features are marked by hand using bzier curves on three 2D images to the front, left and right of the 3D face. In order to utilize them, however, they have to be projected on to the computed mesh of the face that was recorded by a 3D scanner. These meshs have holes in the region of the eyes and the ears rendering the projected line features useless at first. This thesis

first gives an overview over the morphable model and the face registration pipeline, then goes on to obtaining 3D points from the 2D line features, to explain the theory behind the general algorithm and in the main part discusses the problems and solutions we encountered trying to optimize the algorithm for and without line features for the face registration process.

Contents

Chapter 1

Introduction

1.1 Problem Statement

Beschreibung des Problems chronologische, kurze Beschreibung des Vorgehens Fachterminologie so allgemein wie mglich whlen

1. Use Gaussian Processes - 2. Use Line Features =¿ prepare for Gaussian Process Regression In this bachelor thesis Implement 3D face registration using Gaussian Processes and Line Features. One part of the problem is to sample equidistant 3D points from 2D line features marked on images of a 3D face scan. These line features should then be used as an additional input to a registration algorithm which is based on Gaussian Process Regression. The aim is to build a pipeline which starts off with the raw scan data as well as the landmarks and line features. The feature points are used to register the mean face of the MM/BFM (Basel Face Model) on to/with the raw scan thereby obtaining a fully defined and textured 3D model representation of the face in 3D. Registration is the technique of aligning to objects using a transformation, in this case the registration is performed by adding displacements to every points in the mean face model. A model is represented as vector N*d. What is a model? A vector representation of a 3D scan? For the morphing a Posterior Shape Model is used in combination with a Gaussian Process. Image registration is a process of aligning two images into a common coordinate system thus aligning.

(gaussian process + line features for accurate, reproducable registration)

1.2 Review Literature

2. Definition of terms (morphable model, 3D face registration, Gaussian Process regression, posterior shape models) 3. Review of literature (papers)

Chapter 2

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 new faces can be generated by calculating linear combinations of the set of 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 a large set of m 3D face scans. A vector space can be constructed from the available data set where each face is represented by a shape-vector $S \in \mathbb{R}^{3n}$ that contains a stack representation of its n vertices. The texture-vector $T \in \mathbb{T}^{3n}$ contains the corresponding RGB values. New shapes and textures can now be computed with a linear model parametrized by barycentric shape $\vec{\alpha} \in \mathbb{R}^m$ and texture coefficients $\vec{\beta} \in \mathbb{R}^m$.

However, the goal of such a 3D face model is not just to construct arbitrary faces, but plausible faces. This is achieved by estimating two multivariate normal distributions for the coefficients in $\vec{\alpha}$ and $\vec{\beta}$. By observing the likelihood of the coefficients it is now possible to find out how likely the appearance of a corresponding face is. The multivariate normal distributions are constructed from the average shapes $\overline{S} \in \mathbb{R}^{3N}$ and textures $\overline{T} \in \mathbb{R}^{3N}$ of the datasets and the covariance matrices K_S and K_T, which are defined over the differences between each example and the average in both shape and texture. The covariance matrices are then used to perform a Principal Component Analysis which defines a basis transformation to an orthogonal coordinate system the axis of which are the eigenvectors of the respective covariance matrices.

$$\mathcal{S}(\vec{\alpha}) = \overline{S} + S\vec{\alpha}, \quad \mathcal{T}(\vec{\beta}) = \overline{T} + T\vec{\beta} \tag{2.1}$$

In (2.1) the $N = m$ principal eigenvectors of K_S and K_T respectively are assembled column-wise in S and T and scaled in a way such that the prior distribution over the shape and texture parameters is given by a multivariate normal distribution with unit covariance (Amberg).

$$p(\vec{\alpha}, \vec{\beta}) = \mathcal{N}(\vec{\alpha}||\mathbf{0}, \mathbb{I})\mathcal{N}(\vec{\beta}||\mathbf{0}, \mathbb{I}) \tag{2.2}$$

2.2 Achieving Correspondence through Registration

be as specific to say there are triangulated meshs? In order for a 3D Morphable Model to generate plausible faces we have to make sure that all faces in the example set are parametrized equally. For this reason the meshs first have to be brought into correspondence, which is the case when the vertices of different meshs which are at the same semantical position, i.e the left corner of the left eye, have a similar vertex number. A dense point-to-point correspondence between two meshs is accomplished through the process of registration. The training data used for learning a 3D Morphable Models consists solely of registered examples of the 3D shape and texture of faces.

incorporate *WE WANT POINT TO POINT CORRESPONDENCE BETWEEN THE TWO FACES in general: point to point correspondence between to images Are scans already in semantical correspondence? No semantical correspondence FINDING CORRESPONDENCE IS EXACTLY THE AIM OF REGISTRATION =¿ HAVING SAME POINTS AS CLOSE TO ONE ANOTHER AS POSSIBLE Now in order to obtain a 3D representations of the face we need to transform the mean face so that it fits a particular 3D face scan. To find the transformation, however, we first have to find feature points in both 3D representations which correspond to the same semantical structure. Previous work has shown that point landmarks are not sufficient to preserve the level of detail which is imminent in the regions of the eyes, ears and lips and that the computed transformations are not able to preserve* *these regions. For this reason, additional line features have been introduced. In order to relate these How registration works so far What we want to change*

Registration Algorithm Registration is the task of parametrizing one shape in terms of another shape so that the points which are semantically correspondent are mapped onto each other. From a different viewpoint the parametrization can be viewed as a deformation. The shape which is deformed is called the template or reference shape, while the goal shape of the mapping is called the target shape. Registration achieved using a Registration algorithm. Such an algorithm uses prior information in the form of manually clicked feature points, so called landmarks, on all of the face meshs. Correspondence in-between these points is defined through smooth deformations of the template mesh which match the surface and feature points of the target. In this thesis we introduce ***"use" better? It is already academically introduced, just not in this context*** a registration algorithm which is novel to the problem of 3D face registration in two ways: the use of prior information is extended to whole contours of complex regions of the face, refered to as line features and the deformation is modeled using Gaussian Process Regression, a method from the field of Machine Learning.

2.3 Prerequisite Data

image with landmarks and line features a short overview what data we have given

Facial Scans: face scans given as point clouds The data we have given is a set of about 300 face scans that have had a set of key points marked. Furthermore important and detailed regions like the eyes, ears and lips have been marked by contour lines known as line features. The scans have been obtained with a scanner. The surface is very detailed, however the eyes and the nostrils are not recorded. From these scans we want to create fully textured 3D faces, which can be used to build a new face model.

Mean Face: The mean face has been derived from a collection of 100 male and 100 female 3D face models. Describe data and scanner given + Camera model? In the next chapter we will elaborate on the approach of using Gaussian Processes to solving the problem 3D face registration.

shells from the scanner are cleaned

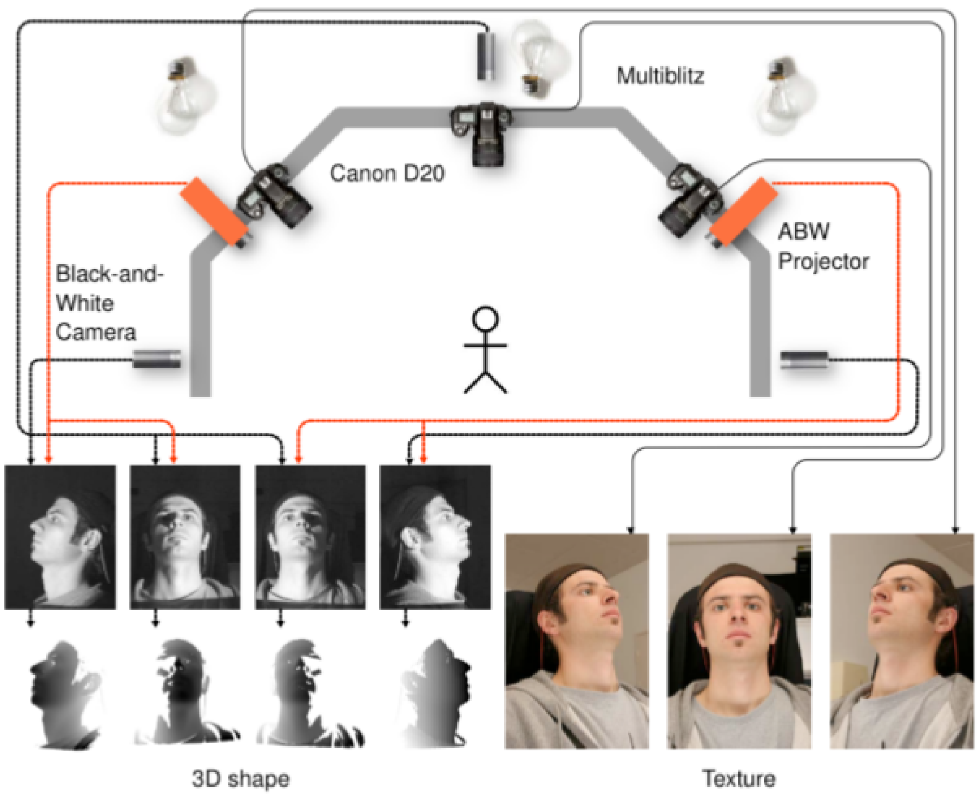


Figure 2.1: 3D scanner

Chapter 3

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 stochastic process, more specifically a vector-valued Gaussian process or Gaussian random field as the registration algorithm. 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 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. A generalization of a stochastic process, which can handle multidimensional 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 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 $\Omega \subseteq \mathbb{R}^n, n > 1$ holds, the process is a Gaussian random field. In the further proceedings the term "Vector-valued Gaussian Processes" will be used to refer to Gaussian random fields. Defining the random variables on an index set in an n-dimensional space, allows for spatial correlation of the resulting values, which is an important aspect of the algorithm discussed later on.

An alternative way of viewing a Gaussian process is to consider it as a distribution over functions. This allows us to look for inference in the space of these functions given a dataset, specifically to find the deformation function given a 3D face mesh. Each random variable now yields the value of a function $f(x)$ at a location $x \in \mathcal{X}$ in the index set of possible inputs. We now denote the index set by $\mathcal{X}$ to stress that we are ceasing to discuss Gaussian processes defined over time. 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 $\mu : \mathcal{X} \to \mathbb{R}$ and covariance $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ functions which in turn are defined over the set of input vectors. 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. When modeling a deformation field with a Gaussian process this circumstance implies that the expected deformation is itself zero.

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 input vectors 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.

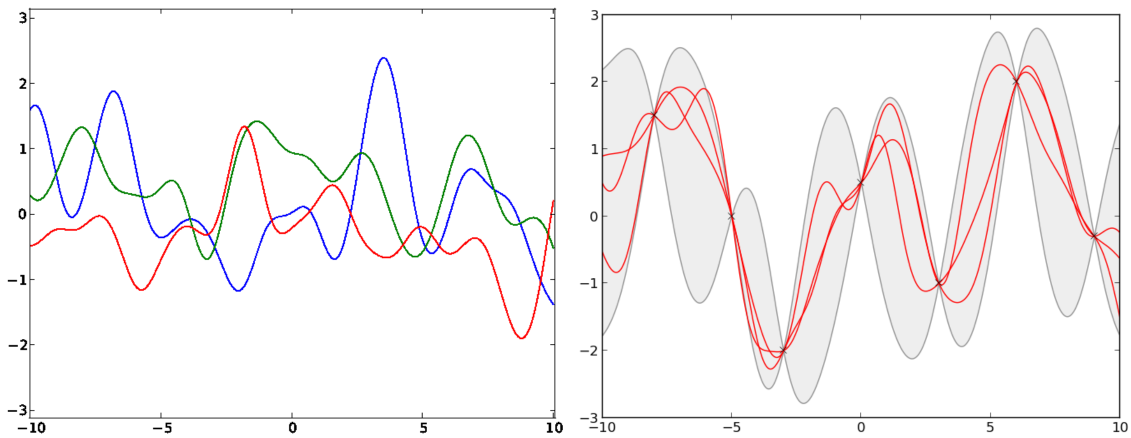
A simple example of a covariance function is the squared exponential covariance function, defined by $cov(f(x), f(x')) = k(x, x') = \exp(-\frac{(x-x')}{2l^2})$. (derivation Rasmussen et al. p.83) *still to be continued and refined...*

It is possible to obtain different prior models by using different covariance functions. In our case, we use a stationary (x-x, 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. To illustrate this one can draw samples from a prior distribution of functions evaluated at any number of points, X_*. The Gaussian Process Prior is solely defined by the covariance matrix made up of the covariances of the input points.

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

A sample is a random Gaussian vector $f_* \sim \mathcal{N}(0, \Sigma(X_*))$ 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 indead a distribution over functions, see figure 3.1. The GP Prior forms the basis for inference in Gaussian Process Regression.



(a) Samples drawn from the prior (b) Samples after a few observations

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

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 above definition from the function-space view. The random variables $X_1, X_2, \ldots, X_k, \ldots, X_n$ are now d-dimensional vectors, yielding a covariance function of the form $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{d \times d}$ and $k(x, x') = \mathbb{E}[X_k(x)^T X_k(x')]$. *Should this paragraph be continued?*

3.3 Gaussian Process Regression

The task of registering two 3D face meshs can be treated as a regression problem in which the goal is to predict the deformation of all floating mesh points, given

the displacement of the landmarks present in both meshs. Trying to fit an expected function - be it linear, quadratic, cubic or nonpolynomial - to the data is not a sufficiently elaborated approach to our problem. Using a Gaussian Process disposes of the need to describe the data by a specific function type, because the response for every input point is now represented by a normally distributed random value, in turn governed by the specification of the covariance function.
Key assumption: data can be represented as a sample from a multivariate gaussian distribution P

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

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:

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

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:

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

Prediction with Gaussian Noise Model In most real world applications

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

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

$$\mathbf{y}_* | \mathbf{f} = \mathbf{y} \sim \mathcal{N}\left(\overline{\mathbf{y}}_*, \Sigma(\mathbf{y}_*)\right) \tag{3.5a}$$

where the mean depends on the observed training targets

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

whilst the covariance depends only on the input points

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

Conclusion, how does this help us to proceed?

3.4 Application to 3D Face Meshs

In this section of we adapt the above presented 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. ***registration and correspondence already explained in model building, deformation field bold instead of calligraphic?*** We therefore strive to predict a deformation field $\mathcal{D} : \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}^3$ 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 displacement 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 multiple meshs of scanned faces. ***Andreas: don't refer to 3DMM mean, because we haven't built a model yet! Leave out "triangulated", kind of mesh topology is not important in this thesis***

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 $d \subseteq \mathbb{R}^3$ where d stands for deformation. After the template and target have been aligned **??** a Vector-valued Gaussian Process can be initialized by defining the prior over all vertices of the template mesh. For this purpose the covariance function has to be redefined to handle 3-dimensional vectors. ***Prior consists of smooth deformations of the mean face***

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

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

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

The template mesh is defined by a set of vectors $\mathcal{X} \in \mathbb{R}^3$ and a set of landmarks $L_{\mathcal{M}} = \{l_1, \cdots, l_n\} \subset \mathbb{R}^3$. ***Introduce landmarks in model building*** 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

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

meaning that a deformation field can be directly drawn as a sample from the prior distribution of the vertices of the template mesh. ***show two or three samples of prior here, next to template/mean mesh***

Reference Mesh Posterior The target landmarks also consist of a set $L_{\mathcal{T}} = \{l_{\mathcal{M}1}, \cdots, l_{\mathcal{M}N}\} \subset \mathbb{R}^3$. Fixing the prior output to the deformation vectors $\mathcal{Y} = \{\mathbf{t} - \mathbf{m} | \mathbf{t} \in L_{\mathcal{T}}, \mathbf{m} \in L_{\mathcal{M}}\}$ defined by the distance between the template and target landmarks and assuming additive i.i.d Gaussian noise the resulting posterior distribution is

$$\begin{bmatrix} \mathcal{Y}_\varepsilon \\ \mathcal{D} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma(\mathcal{Y}) + \sigma^2 \mathcal{I}_{3|\mathcal{Y}|} & \Sigma(\mathcal{Y}, X_*) \\ \Sigma(X_*, \mathcal{Y}) & \Sigma(X_*) \end{bmatrix}\right) \tag{3.9}$$

Is this a correct definition for the distribution?

The deformation model is now rendered fixed at certain landmark points in the target mesh and the goal is to find valid deformations through the set of fixed targets, analogous to the case of eq. 3.5a. The posterior model is defined as the joint distribution of all template mesh points and the template landmarks, conditioned on the output deformation vectors for every template landmark with added noise.

$$\mathcal{D} | \mathcal{X} \rightarrow \mathcal{Y}_\varepsilon. \tag{3.10}$$

We now have defined a distribution over our template mesh. ***mean/template is now max aposteriori solution*** Sampling the conditional distribution creates deformed 3D surfaces of the mean mesh which are fixed at the target landmarks. ***show images of mean, prior and posterior with added landmarks***

3.5 Fitting & Optimization

Due to the fact that the posterior model is fixed at the target landmarks it is possible to perform the registration by drawing a shape from the posterior model. The sample, however, has to be optimized according to the shape of the target face. In order to find the deformation corresponding to the optimal fit d_* a linear optimization with the posterior process as a constraint is be employed/ regularization term. (small lambda) a bit of the posterior mean)

$$d_* = \underset{d \in \mathcal{D}}{\textbf{arg min}} \quad L[O_{\mathcal{T}}, O_{\mathcal{M}} \circ d] + \lambda R[d] \tag{3.11}$$

Minimizing a loss function L - mean square distance for example - on the target and the deformed mean provides a feasible deformation field. D denotes the space of possible deformations in the posterior model. The information needed is held by the covariance matrix of the posterior process. However, using Mercer's theorem (in short what it does, yada yada yada) a basis of functions corresponding to all possible deformations can be extracted. The kernel is thus described as a linear model of these basis functions. ***whole matrix or simple kernel? refer to paper "a unified formulation of statistical model fitting and non-rigid registration***

$$k(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x') \tag{3.12}$$

λ_i are the eigenvalues and ϕ_i the eigenvectors of K. They denote the deformation directions while the eigenvalues ... We are looking for a finite linear combination of eigenvectors that form a deformation field with $\exists \alpha_1 \cdots \alpha_n \in \mathbb{R}$ as linear parameters.

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

f GP(0, K) we take our gaussian process ¿ f — x=y, ***ask Marcel for a helping hand with the theory?***

Next, we want to minimize residuals for the whole of our surface according to optimal parameter values α_i

$$\underset{\alpha \in \mathbb{R}^n}{\textbf{arg min}} \quad \Sigma_{x_i \in \mathcal{T}_R} (f(x_i) - \phi_T(x_i))^2 \tag{3.14}$$

where $f(x_i)$ is the deformation function and $\varphi_T(x_i)$ returns the nearest point on the target mesh. Yields the overall loss function Φ_L

$$\Phi_L(f(x_i) - \varphi_T(x_i)) \tag{3.15}$$

The eigen vectors - which are deformation vectors defining a deformation for every model vertex - of the covariance matrix define a basis space? Shape Modell =¿ select best eigenvectors via PCA in order to simplify computation.

=¿ Vorstellen wie wenn mehrere Wellbleche durch die Target"_"landmarks gelegt werden und dann mit bestimmten parametern alpha zwischen ihnen interpoliert wird Alternative way to understand basis functions for gaussian process: sample from the GP(0, K) and then build a linear model from the functions, f(x) = sum(i, n) alpha(i) si(x) Posterior Distribution of Landmarks Defining the Gaussian Process Posterior Distribution - Landmarks (Referenz deformieren From Gaussian Processes to Shape Models =¿ by selected principal components of the covariance matrix

Chapter 4

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 additionally enhance the registration outcome of this pipeline we use face data where the key regions have been marked with contour lines. In the following we provide a description of line features and their use as well as a specification of the registration pipeline. ***pipeline type important?***

4.1 Line Features

Line features serve the purpose of augmenting the quality of registration by initiating it with a larger set of corresponding points, by sampling points from the lines themselves. They are used to mark complex regions of the face, i.e. the eyes and ears, so that the registration process produces an accurate mapping of the contours of these organs which would otherwise not be possible. Without the prior information provided by line features, an accurate mapping in these regions is hard to achieve, because they have a dense abundance of points, while regions like the cheeks are scarcely defined by points.

For every scan we want to register, we have 8 contours given. These have been marked on three images of every face, see Fig section 1.3, with a special GUI for marking points and lines on images. The contours we call line features depict the eyebrows, eyes, ears and lips of a face. They are made up of a set of segments, each of which is modelled with a **Bézier curve** of a specified order. Bézier curves are often used in Computer Graphics for modelling smooth curves of varying order. Given a set of control points $\mathcal{P} = \{P_0, P_1, P_2, \ldots, P_n\}$ the Bézier curve through these points is given by

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

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

$$B_{i,n}(t) = \begin{pmatrix} n \\ \\ i \end{pmatrix} (1-t)^{n-i} t^i \tag{4.2}$$

and $t \in [0, 1]$ is the curve parameter. The Bernstein polynomials of degree n form a basis for the power polynomials of degree n. Due to the nature of the objects depicted, there are open as well as closed curves.

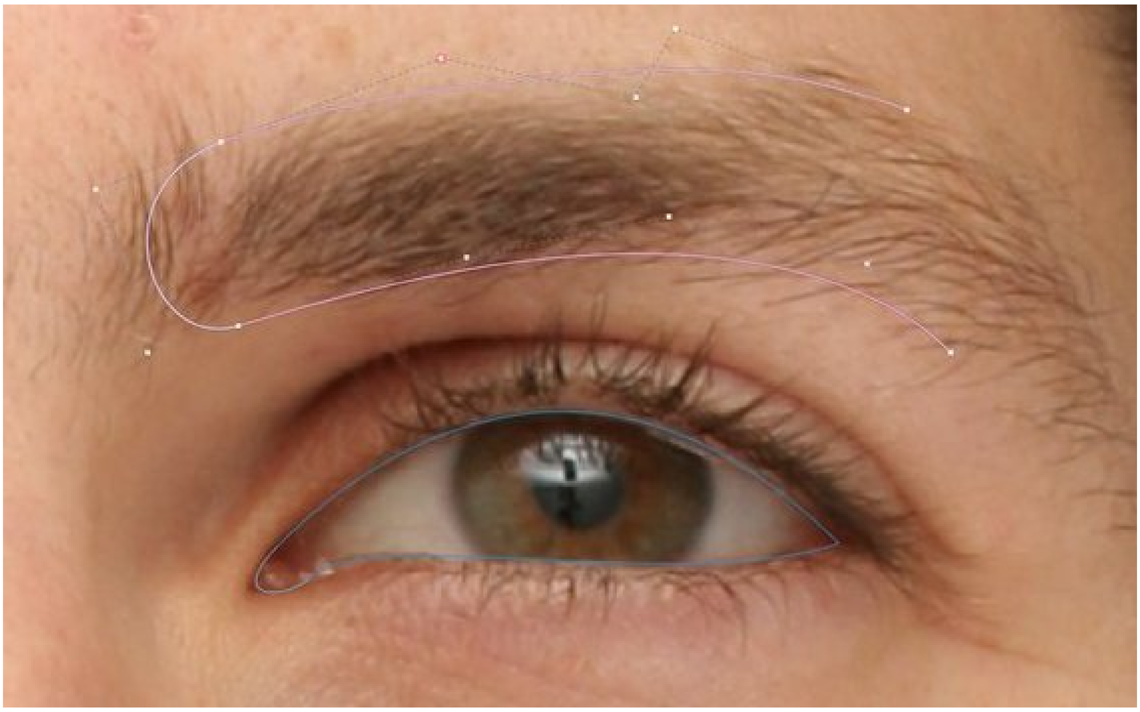


Figure 4.1: Line Features of the left eyebrow and eye, consisting of bézier curves defined by visible control points (white)

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 agreed to use them as additional point-wise information. In order to get this point-wise information the line features are sampled at discrete intervals resulting in a set of additional landmarks $L_{Add} = \{l_1, \cdots, l_N\}$. These define the mapping $\Omega : L_{Add\mathcal{M}} \rightarrow L_{Add\mathcal{T}}$ of the contours - describing the different important features present in the faces - in the template face mesh on those of the target face mesh. In order for the mapping Ω to be approximately plausible, we choose an equidistant parametrization. In effect, when a curve is sampled at N points, these N points are all at equal parametric intervals. ***equidistant parametrization is not a fact, it is a choice. Different ears have different topology?***

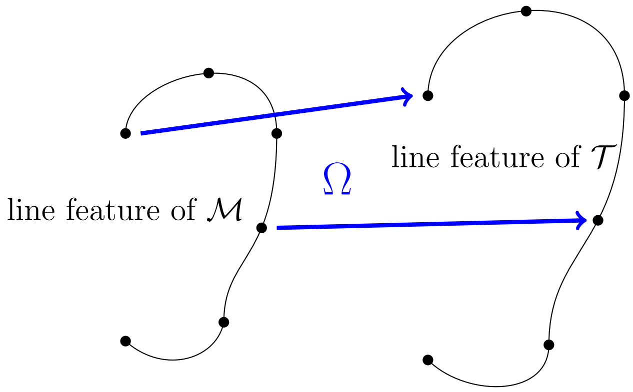


Figure 4.2: Mapping of equidistant samples of **ear** line features from the reference (left) on to the target (right)

Arc Length Parametrization

The first problem which becomes apparent when trying to sample the line features is that the bézier curve segments don't allow for equidistant parametrization, because the underlying parameter $t \in \mathbb{R}$ is not linear in respect to the length of the curve. The growth of the parameter of a bézier curve is instead dictated by velocity.
Consequently, the imperative must be instead to evaluate the curves based on their arc-length, which is defined as the length of the rectified curve. The underlying parameter must then correspond - at every point of the curve - to the ratio between the length of the part of the curve that has been traversed and the total curve length.

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 to

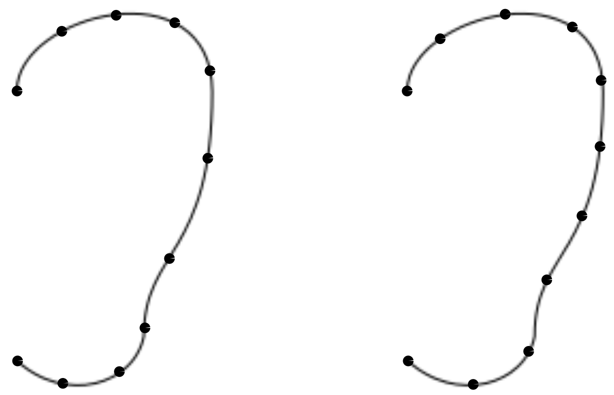


Figure 4.3: a simplified - the depicted **ear** line feature is treated as one sole bézier segment - illustration of the difference between bézier (**left**) and equidistant (**right**) parametrization.

the curve parameter $t = L^{-1}(l)$. This mapping can of course be derived analytically, but it is far easier to implement it using a numeric approximation.

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 bézier curve. For each point we save its approximate distance from the origin of the segment as a key into a new slot in the lookup table of this segment, while its coordinates act as the slot value. The distance is approximated by summing

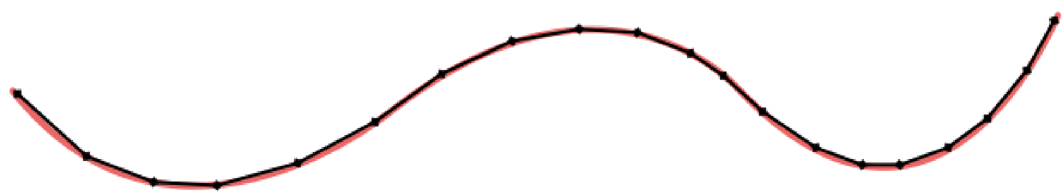


Figure 4.4: Visualization of euclidean distances between points computed with bézier curve parametrization

up the euclidean distances of each point to its respective predecessor starting from the origin.

The resulting lookup table contains the approximative distances and coordinates of a large number of points from the origin of a curve segment. Assembling the segments' lookup tables gives us the table table for the whole curve with the last key representing its arc-length.

Second, the curve can now 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 $k = [0, \cdots, N]$ for open curves and $k = [0, \cdots, N-1]$ 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 this distance. We choose index which returns the coordinates for the exact fraction length and if that is not the case the index with the next smaller length. The coordinates of the point either exactly or approximately computed for the given fraction length of the curve is used as the sampled point.

3D Mesh Projection of Sampled Points

Having implemented arc length parametrization it is possible to draw an arbitrary amount of samples $x \in \mathbb{R}^2$ from the line features. They are then defined as a set of points $S \subset \mathbb{R}^2$. Our goal is, however, to have these additional landmarks describing the features on the mesh of a face itself and not a 2-dimensional snapshot thereof. Because we have no information on the depth of the line features, we have to project the sampled points of each line feature onto a face mesh in order to obtain their approximate 3D representation. ***Introduce camera model with schema here?*** In the camera model the image is located on the viewing plane or viewport opposite of the focal point. (computes 3D direction of 2D sample point) The direction of the 3D representation of a point on a curve is given by the (normalized) vector defining the position of the point on the viewing plane from the perspective of the focal point of the camera. Given: diskrete mesh, how to choose point? We now a seek a mesh vertex that is the most accurate representation of a sample point on the 2D curve. The dot product of their normalized directions is used as a similarity measure. In order to find a corresponding vertex, we save all the distances of mesh vertices in a list which have a similarity measure that is higher than a specified threshold. We then select the distance of the vertex with the maximum 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.

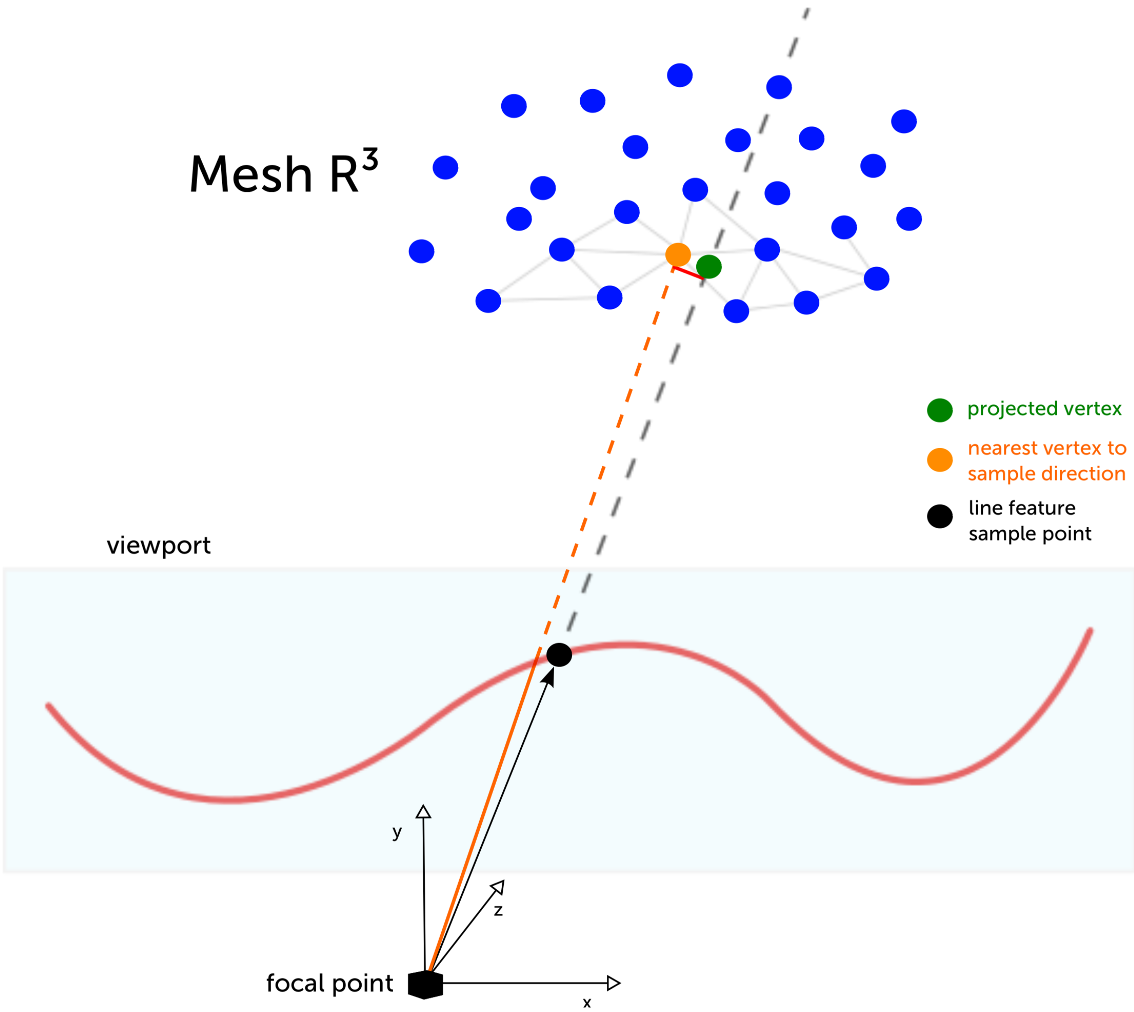


Figure 4.5: shows the projection of a sample point - on a 2D line feature - from the viewport onto a 3D mesh. The distance vector of the vertex with the most similar direction vector (orange) is projected (red) onto the direction of the sample point, resulting in the projected vertex (green)

Problems & Inacurracies The face meshs used in the face registration pipeline contain large holes around the ears and the eyes. In effect, the projected sample points are off target, because now the mesh vertex with the most similar direction is likely farther away at a suboptimal location. This circumstance leads to the projected line not clearly being distinguishable as a contour line. On different data sets the performance of the projection of the line features for a large number of samples, i.e. 30, varied significantly. Overall one can say that the distortion of a projected line increases with the amount of sample points. ***Compute some***

landmarks with 30 samples up close image of eye holes of face scan An easy workaround was to reduce the number/amount of sample points. By using only 5-10 sample points per curve some datasets rendered near perfect results on a “control dataset”. ***Compile list of datasets*** However, when the holes are too large this workaround also fails. Stoff for discussion what was not solved: when holes are too large, “method” fails –¿ treat projected points as outliers However, as long as the method is dependent on the data from the scans - the size of the holes in the meshs - it lacks generality and generality is exactly the basis for feasible and reproducable registration results.

4.3 Preparing the Mean Mesh

Rendering, marking =¿ Projection On top of that, another problem occured, because the mean face mesh of course doesn’t have any line features projected on to it either. Rendering, marking line features, projecting back possible, because we know direction

However, it contains about 60 feature points manually clicked, which are not present in newly scanned datasets. Eliminate the ones, which are not clicked on scans

Output pipeline specifications

4.4 Rigid Mesh Alignment

Rigid Alignment We have to perform a rigid transformation to align the meshs according to the feature/ landmark points.

simple rigid transformation of the scanned face onto the mean, transformation computed from landmark vectors. To begin the registration we first have to align the two meshs. The floating mesh has to be clipped at the neck and around the ears where the scanner has left artifacts. Furthermore the mouth cavity of the mean face has to be removed. We then selected the 11 feature points present in the floating mesh in the mean face from the abundant 60. To achieve this we wrote a python script loading the feature point files. A feature point is described by its 3D coordinates, a visibility parameter in the range [-1,1] and a label denoting its exact location (mouth.inner.upper). All we had do to now was to create to dictionaries label : (x,y,z) and to compare them for labels. Then we passed the resulting point correspondencies to the python vtk api for the mean of computing a transformation comprised of simple translation and rotation (no scaling, only 3 point correspondencies needed). Note, we are not trying to map the meshs on to one another here. We are simply trying to align them through the use of the feature points. The computed transformation we applied to all points in the floating mesh. The resulting mesh was written to a file and then opened in paraview. We now had the meshs in a position from where we could start the actual mapping. The mean face was broader in shape than the scan and was perfectly coated in texture for the simple reason that hours of manual labour have been invested to render this important piece of data a perfect reference. Now in order to receive a perfect mapping of the floating mesh on to the mean/reference mesh we have to allow for 3 degrees of freedom, that is in all 3 dimensions x,y and z, for every pixel in the floating mesh except for the reference points we have used as correspondencies. The parameters having the most influence to the mapping will be those specified in the constraints we introduced into the equation via regularization. The idea behind the use of sampled points from the line features was to have more point correspondencies in complex regions as for example the eyes and the ears where there is a great abundancy of pixels and the algorithm isnt likely to create a flow field which is accurate not enough to describe these regions, because of its smoothness constraint. For the actual registration we use the software framework statismo developed at the Computer Science Department of the University of Basel. 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 from the floating mesh. Furthermore we use the software package gpfitting for the actual fitting. We generate a infinite row of faces from the statistical model using gaussian processes and then sample out a fixed number. Then the faces are left. Carry on.

4.5 Prior Model

what to say here? describe programme?

4.6 Posterior Model

what to say here? describe programme?

4.7 Fitting

4.8 Robust Loss Functions

Optimizing the loss function? After the alignment of template and target mesh, the template protrudes over the target on the upper side of the head and the side of the neck. ***show an image with template and target on top of each other*** Performing optimization as described in ?? using a simple Mean Square Error(MSE) as a distance measure between the template and target mesh penalizes the portruding regions of the template with a strong gradient towards the rims of the template and therefore causes strong distortions. ***show image of failed fitting, next to target***

Our approach to tackling this problem was to try out 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 farther away than a certain threshold from the next point on the target mesh However, as can be seen from the formulas, these techniques require finding appropriate parameters first which produce reasonable/acceptable visual results.

Fair

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

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

Huber

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

$$\psi(x) = \begin{cases} x & \text{if} \\ ksgn(x) & \text{if} \end{cases} \tag{4.4b}$$

Tukey

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

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

for each estimator show a sequence of fits for different parameters and 3 different meshs?

4.9 Varying the Variances