UNIVERSITY OF COPENHAGEN

Bayes reconstruction of missing teeth

Sporring, Jon; Jensen, Katrine Hommelhoff

Published in:
Journal of Mathematical Imaging and Vision

Publication date:
2008

Document version
Early version, also known as pre-print

Citation for published version (APA):
Sporring, J., & Jensen, K. H. (2008). Bayes reconstruction of missing teeth. *Journal of Mathematical Imaging and Vision*, *13*(2-3), 245-254.

UNIVERSITY OF COPENHAGEN

Bayes Reconstruction of Missing Teeth

Jon Sporring · Katrine Hommelhoff Jensen

Abstract We propose a method for restoring the surface of
tooth crowns in a 3D model of a human denture, so that the
pose and anatomical features of the tooth will work well for
chewing. This is achieved by including information about
the position and anatomy of the other teeth in the mouth.
Our system contains two major parts: A statistical model of
a selection of tooth shapes and a reconstruction of missing
data.

We use a training set consisting of 3D scans of dental cast
models obtained with a laser scanner, and we have build a
model of the shape variability of the teeth, their neighbors,
and their antagonists, using the eigenstructure of the covari-
ance matrix, also known as Principle Component Analysis
(PCA). PCA is equivalent to fitting a multivariate Gaussian
distribution to the data and the principle directions consti-
tute a linear model for stochastic data and is used both for
a data reduction or equivalently noise elimination and for
data analysis. However for small sets of high dimensional
data, the log-likelihood estimator for the covariance matrix
is often far from convergence, and therefore reliable models
must be obtained by use of prior information. We propose a
natural and intrinsic regularization of the log-likelihood es-
timate based on differential geometrical properties of teeth
surfaces, and we show general conditions under which this
may be considered a Bayes prior.

J. Sporring (✉) · K. Hommelhoff Jensen
Department of Computer Science, University of Copenhagen,
Universitetsparken 1, 2100 Copenhagen, Denmark
e-mail: sporring@diku.dk

Finally we use Bayes method to propose the reconstruc-
tion of missing data, for e.g. finding the most probable shape
of a missing tooth based on the best match with our shape
model on the known data, and we superior improved recon-
structions of our full system.

Keywords Principle component analysis · Bayes method ·
Missing data · Reconstruction of teeth

1 Tooth Reconstruction

In the dental industry, the design and construction of restora-
tions to be inserted in a patient's mouth is carried out by
dental technicians, that are highly trained experts in tooth
anatomy and the function of the bite. The task can be to
model the missing part of a broken tooth crown, model the
crown of a whole missing tooth or even several missing
teeth. The restorations are traditionally constructed directly
from the materials by hand, but the use of software to model
the construction elements of a restoration has been grow-
ing rapidly the last couple of years [1–4]. Other than saving
money on the temporary building materials, the software so-
lution saves time, as some of the traditional production steps
can be skipped and proper customized 3D modeling tools
and automatic routines can speed up the construction work.
The existing dental software systems combines the software
with a scanning device to produce a 3D surface model of the
patients remaining teeth, on which the restoration is to be
designed. The modeled restoration can then be exported as a
3D surface model and milled or printed directly in the final
material.

One of the most challenging steps towards an automa-
tion of dental restoration modeling software is the anatom-
ical deformations of the tooth crowns to be reconstructed.

Fig. 1 Cast models scanned by the 3Shape laser scanner, aligned (**a**) and the lower jaw (**b**), and a triangulated surface mesh of a scanned cast model (**c**). Notice on (**c**) that the mesh has been decimated to a much lower solution than the usual quality

(a)   (b)   (c)
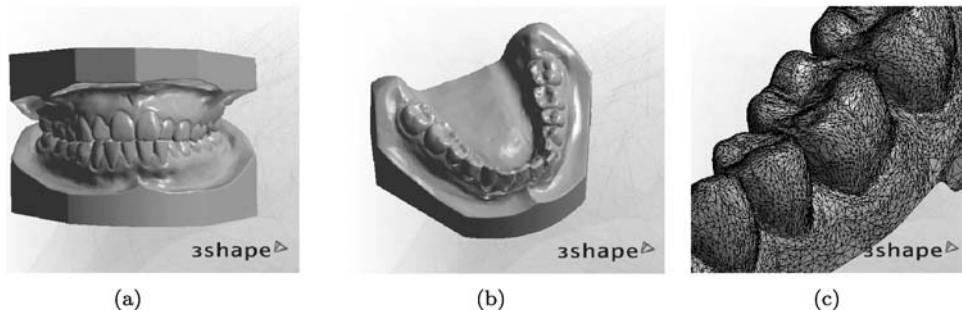
An anatomical correct deformation of a tooth crown surface cannot be calculated exclusively from the size and location of the surrounding surfaces of the scanned data. Some prior knowledge must be added to the system, which describes the shapes and legal deformations of the teeth. It is our goal to develop a system that can learn and describe the complex shape system of the bite, and with this knowledge reconstruct the surfaces of missing tooth crown parts, whole teeth, or several teeth from information extracted from scanned data. The plaster casts have been scanned by a 3Shape laser scanner, some of which are illustrated in Fig. 1. The work reported in this article is an extension of [5].

In the following we will discuss shape variation of landmarks on teeth using Bayes Principle Component Analysis [4, 6–9], and we will suggest to estimate of the covariance matrix as,

$$\mathbf{C} = \frac{1}{m} \mathbf{X} \mathbf{X}^T + \alpha \mathbf{W}, \tag{1}$$

where $\mathbf{X} = [\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \ldots \mid \boldsymbol{x}_m]$ is a matrix of m concatenated vectors $\boldsymbol{x}_i \in \mathbb{R}^{3n}$, $\mathbf{X}^T$ is the transpose of $\mathbf{X}$, and α controls the amount of regularization. It was observed that the generalization error was reduced in terms of the leave-one-out error, when matrix $\mathbf{W}$ had the following form,

$$\mathbf{W} = \boldsymbol{\Omega} \otimes \mathbf{1}_{3\times3}, \tag{2a}$$

$$\Omega_{ab} = \frac{1}{2} f(d_{ab})(\boldsymbol{n}_a \cdot \boldsymbol{n}_b + 1), \tag{2b}$$

where $\mathbf{R} \otimes \mathbf{S} = \{R_{ij}\mathbf{S}\}$ is the Kronecker product, $\mathbf{1}_{3\times3}$ is a 3×3 matrix of ones, the function f is an exponentially decreasing function such that $f(0) = 1$, d_{ab} is the geodesic distance between two points on the surface of a tooth $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$, and $\boldsymbol{n}_a$ and $\boldsymbol{n}_b$ are the corresponding outward normals. The estimate in (1) corresponds to a Bayes estimate of the covariance matrix with a Gaussian Likelihood and an inverted Wishart distribution as Prior,

$$P(\mathbf{C}) = k \exp\left(-\frac{m}{2} \operatorname{tr}(\mathbf{W}\mathbf{C}^{-1})\right), \tag{3}$$

for a suitable normalization constant k, arbitrary constant matrix $m > 0$, and when $\mathbf{W}$ is positive definite. We will prove that $\mathbf{W}$ is positive definite, when $f(d_{ab})$ is diagonally dominant.

2 Shape Modeling

We operate with the notion of shape as that which is left, when translation, rotation, and scaling is removed. To reconstruct the shape of an incomplete object, we need a model description of the object shape and variability. The classical approach is based on representing and modeling shape as a set of *landmarks* (see [8, 10–13] and references herein). On each training shape, a finite number of landmarks are located on surface features that corresponds between the shapes. This representation is directly applicable in the *Active Shape Model* (ASM). The biggest disadvantage is the time-consuming manual labor needed. Alternative approaches, that carry a smooth surface implicitly in the shape description, are the *Level-set function* representation [14] and the *Medial representation* (M-rep) [15]. However, both representations are problematic, since they cannot robustly handle non-closed shape surfaces such as scans of plaster casts of teeth.

Shape Warping has also been studied in the literature: In [16], each training shape landmark is warped to a template shape, where a template shape mesh is projected onto the shape before warping landmarks and mesh vertices back. Based on this idea, [17] introduced a 3D morphable model that use 3D meshes as training data rather than images. Each mesh vertex achieves the same status as a landmark, and the correspondence between the training meshes and the template mesh is estimated from a sparse set of correspondence points, manually marked by the user on each training mesh. This work was later extended to include Bayes priors using a the exponential of the Mahalanobis distance between landmarks [4]. In [3] the reconstructions were adjusted for neighboring teeth using local mathematical morphology. The quality of the warping depends on both the complexity of the surfaces and the unknown variation of the samples. Further, when only extracting data from the occlusal, frontal, and back sides of the teeth, a significant amount of surface is left unknown. We are thus lead to the idea of keeping the surface representation separated from the data. Further, the reconstruction situations depends on the amount of a tooth that must be reconstructed. If the areas of reconstruction are somehow marked in the process, a better

approach is to let the surface conform to the reconstructed data, while respecting the borders between the reconstructed parts and the original data. The task is to create a mesh that can be guided through the landmarks, while keeping the surface smooth and respecting the local and global constraints. This coupling we implement using Variational Implicit Surfaces [18] interpolated between landmarks, and thus we obtain a flexible method, that may be extended with additional information about the anatomical features, and allow us to focus on landmarks only.

A shape will be described as a collection of n points $p_i \in \mathbb{R}^3$, and we represent a shape as the concatenation of these points into a single $3n$ dimensional vector $x = [p_1, p_2, \ldots, p_n]^T$. We assume that a population of m different examples of a shape is given, and we analyze the shape variability in this population up to the similarity group of actions: translation, rotation, and scaling. Following [8] we normalize each shapes by fist translating it to origo, scale it each to unit size, and rotate it to a common reference shape, e.g. x_1. We will misuse notation and use x for the normalized shape coordinates in the remainder of this article. We now build a linear model of the population by calculating the population mean $\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$, produce a matrix of concatenated differences of each shape to the common mean as,

$$\mathbf{X} = [x_1 - \overline{x} \mid x_2 - \overline{x} \mid \ldots \mid x_m - \overline{x}]. \tag{4}$$

Then we compute the covariance matrix as

$$\mathbf{C} = \frac{1}{m}\mathbf{X}\mathbf{X}^T, \tag{5}$$

and its eigenvalues λ_i and eigenvectors ϕ_i,

$$\mathbf{C}\mathbf{\Phi} = \mathbf{\Lambda}\mathbf{\Phi}, \tag{6}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues and the columns of the matrix $\mathbf{\Phi}$ contains the corresponding eigenvectors or *modes*. Thus any shape x from the training set can be reproduced by a linear computation of the mean and the principal components as

$$x = \overline{x} + \mathbf{\Phi}b, \tag{7}$$

where b is the vector of shape model parameters. The strength of the model is that the eigenvectors corresponding to the largest eigenvalues model the training set with an error equal to the sum of the neglected eigenvalues.

Our training set for each tooth consist of 12 samples. The landmarks were set by a non-expert with expert assistance and are primarily anatomical, with the exception of some pseudo-landmarks. 11 principal components were calculated for each shape model. The relative small training set size could potentially introduce problems regarding the
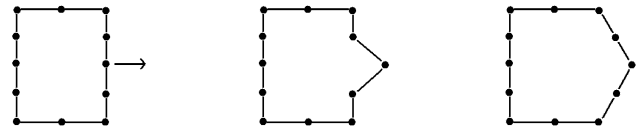

Fig. 2 The original 2D shape (*left*), the effect on the neighbors when moving a point with $\mathbf{C}_{Id}$ as covariance matrix (*middle*) and with $\mathbf{C}_{Id}$ augmented with a small value in the covariances of neighboring points (*right*)

generality, if the dependence on the model statistics is not relaxed in the reconstruction procedure. We will attempt to add artificial eigenmodes to the models, to improve flexibility without hazarding the object shape or overruling the existing eigenmodes. The rationale is that landmarks on the same side of a surface are expected to be correlated proportional with their distance. Thus, if one landmark were to be moved, then we expect that the neighboring landmarks will be effected. We will refer to this as *elasticity*, which will be described in the following.

The modes of variations were calculated from the covariance matrix of the combined data samples. Consider the general $3n \times 3n$ covariance matrix with covariances C_{ij}. If the covariance matrix of the data sample matrix was replaced by the identity covariance matrix $\mathbf{C}_{Id} = \mathbf{I}_{3n}$, then all landmark coordinates would be independent of each other. This means that combinations of the resulting eigenmodes could move the landmarks of a modeled shape in any direction. In that sense, $\mathbf{C}_{Id}$ defines an under-constrained, lower limit to the shape models. In [19] it has been shown how to add smoothness constrained deformations to a shape model by increasing the correlation between neighboring points in 2D shapes. The idea is, that when adding a small value to neighboring points in $\mathbf{C}_{Id}$, a covariation between the points is artificially created. Visually, moving a point in the shape will have an elastic effect on the neighbors. The effect of moving a point in a 2D shape with $\mathbf{C}_{Id}$ and $\mathbf{C}_{Id}$ augmented with a positive value, e.g. 0.5, in the covariances between neighboring points, is illustrated in Fig. 2.

The actual smoothness used in [19] was, however, not controlled in the relation to the model statistics, they were implemented to *substitute* model statistics. We need to control the amount of smoothness so that its function is a deformation *supplement*. Furthermore, we need to re-think the concept of neighbors in 3D, so that the elastic deformation added makes sense and respect the object shape.

In order to relax the tooth shape models we add a small value to all the neighbor-landmark covariances in the covariance matrix. Defining 'neighborhood' is a little more difficult in 3D, though. Neighborhood should be defined more in terms of distance than a number of closest neighbors, and should furthermore be measured over the surface and not necessarily as the shortest distance between two landmarks. As the solid objects teeth are, the smaller artifacts

on one side of a given tooth type doesn't have any effect on the smaller surface variations on the other side, disregarding the distance from landmarks on one side to landmarks on the other. In fact, when teeth have more unusual artifact's on the surface and is thus a difficult subject for an over-constrained shape model, it is usually due to abnormal chewing or smaller damages, both of which only have a local effect. The neighborhood of a landmark p could be measured as the landmarks within some fixed distance from p. This, however, leaves us with the problem of deciding such a distance. Furthermore, it could introduce some problems regarding the scale of the individual tooth samples—we must choose a method that determines a well defined neighborhood of all shape samples in a model. Therefore, along with each shape sample we calculate a matrix of landmark pair geodesic surface distances. These distances are then 'aligned' along with the shape samples, by scaling them according to the corresponding sample. This way we achieve a mean distance for each landmark pair, a value which should support in an expression of the effect of the position of one landmark, on others. Furthermore, we want to restrict the effect of a landmark on one side of a tooth 'cusp' feature on landmarks on the other side, referring to the above philosophy and thus gaining an elasticity with an even more local effect. As illustrated on Fig. 3 this can be done by calculating a mean surface normal in each landmark and adding to our expression a weight of the angular difference between each landmark pair. In effect, the smaller the geodesic distance, and the smaller the angular difference between two landmarks, the bigger a value we should add to the covariance matrix and the bigger effect will the landmarks thus have on each others positions in the later reconstruction. To restrict the expression even more, so only landmarks close to each other, i.e. in the 'local' neighborhood, should have increased covariance, we apply an exponential weighting of the distances and angular differences rather than linear, and find an appropriate locality parameter experimentally. Let d_{ab} be the mean surface-geodesic distance between landmark p_a and p_b over the training data. Then, a general way of calculating how much they should affect each other is to calculate this as a weight $0 < \Omega_{ab} < 1$, where it is our experience that the following function is useful,

$$\Omega_{ab} = \frac{1}{2} f(d_{ab}) (\boldsymbol{n}_a \cdot \boldsymbol{n}_b + 1), \tag{8a}$$

$$f(d_{ab}) = 3^{1 - \frac{\beta d_{ab}}{d_{\max}}} - 1 \tag{8b}$$

and where $\boldsymbol{n}_a$ are the normal at point p_a as shown in Fig. 3, $d_{\max}$ is the maximum surface-geodesic distance between points on a shape over the training set, and β is a locality parameter, controlling the increasing of the weight for decreasing distance between the landmarks. For a given pair of

Fig. 3 (Color online) Three points on a tooth (*green*) together with their surface normal (*red*), and angle difference to the rightmost normal. The neighbors to the rightmost landmark is a weighted sum of the surface-geodesic distance and the angle differences

Table 1 Leave-one-out experiments without and with elasticity of $v = 0.2$ and locality $\beta = 2$

Shape model	Mean residual error	Mean residual error with elasticity
Upper 1st molar	0.011	0.010
Upper 2nd premolar	0.018	0.016
Upper 1st premolar	0.016	0.015
Upper canine	0.024	0.022
Lower 1st molar	0.012	0.011

hypothetical landmarks p_a and p_b, where $a \neq b$, we modify the 9 corresponding entries of the covariance matrix, $\mathbf{C}_{\text{elastic}} = \{C_{ij}^{\text{elastic}}\}$, as follows

$$C_{ij}^{\text{elastic}} = C_{ij} + W_{ij}\alpha, \tag{9}$$

where we have found it useful to let $\alpha = \frac{vm}{n}$ be controlled by the new parameter v. The elasticity influences the least significant eigenmodes the most, and should be kept sufficiently small in order not to destroy the statistical properties of the training data.

The parameters will experimentally be found by performing the leave-one-out experiments on the corresponding shape models based on PCA of $\mathbf{C}_{\text{elastic}}$. The goal is thus to find a set of parameters that decreases the residual error for all models in the experiment. Table 1 demonstrates the generally better results. With higher values of v than $v = 0.2$, the most significant eigenvectors slowly started changing direction, until the corresponding eigenmode changed significance at around $v = 0.9$, and therefore we accept $v = 0.2$ as the maximum. The amount of non-zero eigenmodes created from $\mathbf{C}_{\text{elastic}}$ are typically as many as kn, but a big amount of the least significant eigenmodes can be removed while still keeping the model more general than in the pure statistical model. The locality parameter β has given the best results for values between 2 and 3.

3 Bayes Estimation of the Covariance Matrix

In the following we will show that (9) is a Bayes estimate for the covariance matrix, where $W_{ij}\alpha$ is the prior on the covariance matrix, when W_{ab} is diagonally dominant. We use

the matrix differential calculus according to [20], and for expository reasons we will first give a derivation of the classical log-likelihood estimates of the covariance matrix for random data, and this material will be used as a reference point for deriving expressions for a Bayes estimate. Derivation of the log-likelihood estimate for the mean is similar. Following this derivation we will show that the prior must be of Inverse Wishart type, and we will prove the conditions on W_{ab} for this to be a distribution.

Consider an $3n$ dimensional Euclidean space and m sample points in this space, $\boldsymbol{x}_m \in \mathbb{R}^{3n}$. We will assume that the sample points are identically, independently, and normally distributed in $\mathbb{R}^{3n}$ according to, $P(\boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) = (2\pi)^{(\frac{3n}{2})} \det \mathbf{C}^{(\frac{1}{2})} \exp(-\frac{1}{2}(\boldsymbol{x}_m - \overline{\boldsymbol{x}})^T \mathbf{C}^{-1}(\boldsymbol{x}_m - \overline{\boldsymbol{x}}))$, for unknown covariance matrix $\mathbf{C} \in \mathbb{R}^{3n \times 3n}$ and mean $\overline{\boldsymbol{x}} \in \mathbb{R}^{3n}$. Hence, the joint distribution is given as,

$$
\begin{aligned}
P(&\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) \\
&= \left((2\pi)^{(\frac{3n}{2})} \det \mathbf{C}^{(\frac{1}{2})} \right)^{-m} \\
&\quad \times \exp\left(-\frac{1}{2} \sum_{m=1}^{m} (\boldsymbol{x}_m - \overline{\boldsymbol{x}})^T \mathbf{C}^{-1}(\boldsymbol{x}_m - \overline{\boldsymbol{x}}) \right).
\end{aligned} \tag{10}
$$

To estimate $\mathbf{C}$ from a set of samples, we seek the maximum point of $P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C})$, and since the logarithm function is strictly monotonic, this point coincide with the maximum point of $\log P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C})$. For practical reasons we also rewrite the sum under the exponential function as, $\sum_{m=1}^{m} (\boldsymbol{x}_m - \overline{\boldsymbol{x}})^T \mathbf{C}^{-1}(\boldsymbol{x}_m - \overline{\boldsymbol{x}}) = \mathrm{tr}(\mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T)$, and rewrite the logarithm of the Gaussian distribution as,

$$
\begin{aligned}
L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) &= \log P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) \tag{11a} \\
&= -\frac{3nm}{2} \log(2\pi) - \frac{m}{2} \log \det \mathbf{C} \\
&\quad - \frac{1}{2} \mathrm{tr}(\mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T). \tag{11b}
\end{aligned}
$$

This is the same as minus the optimal code length [21] of the total set of data points from our assumed Gaussian source, and where $\mathbf{C}$ and $\overline{\boldsymbol{x}}$ is fixed and known to both the sender and receiver.

The differential $\mathrm{d}L$ of L varying only $\mathbf{C}$ is found to be,

$$
\mathrm{d}L_{\mathbf{c}} = -\mathrm{tr}\left(\frac{m}{2} \mathbf{C}^{-1} \mathrm{d}\mathbf{C} + \frac{1}{2}(\mathrm{d}\mathbf{C}^{-1})\mathbf{X}\mathbf{X}^T \right). \tag{12}
$$

To identify the partial derivatives of $\mathrm{d}L$ we isolate $\mathrm{d}\mathbf{C}$ using, $\mathrm{d}\mathbf{C}^{-1} = -\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}$, and we find that

$$
\begin{aligned}
0 &= \mathrm{d}L_{\mathbf{C}} \tag{13a} \\
&= -\mathrm{tr}\left(\frac{m}{2} \mathbf{C}^{-1} \mathrm{d}\mathbf{C} - \frac{1}{2}\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T \right) \tag{13b} \\
&= -\frac{1}{2} \mathrm{tr}(\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})(m\mathbf{I} - \mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T)). \tag{13c}
\end{aligned}
$$

A non-trivial solution is $0 = m\mathbf{I} - \mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T$ or equivalently,

$$
\mathbf{C} = \frac{1}{m}\mathbf{X}\mathbf{X}^T, \tag{14}
$$

which may be recognized as the log-likelihood estimate of the covariance matrix [22, Theorem 3.2.1].

Applying Bayes theorem on Mean vectors and Covariance matrices we find that

$$
P(\overline{\boldsymbol{x}}, \mathbf{C} \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \frac{P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) P(\overline{\boldsymbol{x}}, \mathbf{C})}{P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)}, \tag{15}
$$

where we denote $P(\overline{\boldsymbol{x}}, \mathbf{C} \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ the posterior, $P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C})$ the likelihood, $P(\overline{\boldsymbol{x}}, \mathbf{C})$ the prior, and $P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ the evidence. The point of Maximum Posterior, also known as Maximum A Posteriori (MAP), is found as the maximum as the point of zero partial derivatives by the differential of the log-Posterior w.r.t. $\overline{\boldsymbol{x}}$ and $\mathbf{C}$,

$$
\begin{aligned}
\mathrm{d} &\log P(\overline{\boldsymbol{x}}, \mathbf{C} \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \\
&= \mathrm{d}\log P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) + \mathrm{d}\log P(\overline{\boldsymbol{x}}, \mathbf{C}) \\
&\quad - \mathrm{d}\log P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \tag{16a} \\
&= \mathrm{d}\log P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \mid \overline{\boldsymbol{x}}, \mathbf{C}) + \mathrm{d}\log P(\overline{\boldsymbol{x}}, \mathbf{C}). \tag{16b}
\end{aligned}
$$

Again this has an information theoretical equivalent as the differential of minus the two-parts Minimum Description Length (MDL) [23]. We will impose independence on our prior, such that $P(\overline{\boldsymbol{x}}, \mathbf{C}) = P(\overline{\boldsymbol{x}})P(\mathbf{C})$, in which case for fixed $\overline{\boldsymbol{x}}$ we have that

$$
\mathrm{d}\log P(\overline{\boldsymbol{x}}, \mathbf{C}) = \mathrm{d}\log P(\mathbf{C}). \tag{17}
$$

Returning to the regularization term defined in (8) and (9), we will now interpret the addition of the standard log-likelihood estimate (14) of the covariance matrix by a constant matrix $\mathbf{W}$,

$$
\mathbf{C} = \frac{1}{m}\mathbf{X}\mathbf{X}^T + \alpha\mathbf{W}, \tag{18}
$$

as Bayes estimate of the covariance matrix, where $\mathbf{W}$ is the essential part of a prior. The parameter $\alpha > 0$ controls the amount of regularization, but has no impact on the interpretation of (18) as a Bayes estimate, and henceforth we will assume that $\alpha = 1$.

Targeting a Gaussian likelihood we pre multiply (18) with $\frac{1}{2}m\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}$, rearranging terms, and taking the trace, which gives

$$
\begin{aligned}
0 &= -\frac{1}{2} \mathrm{tr}(m\mathbf{C}^{-1}\mathrm{d}\mathbf{C} - \mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T \\
&\quad - m\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}\mathbf{W}) \tag{19a} \\
&= -\frac{1}{2} \mathrm{tr}(m\mathbf{C}^{-1}\mathrm{d}\mathbf{C} - \mathbf{C}^{-1}\mathbf{X}\mathbf{X}^T\mathbf{C}^{-1}\mathrm{d}\mathbf{C} \\
&\quad - m\mathbf{C}^{-1}(\mathrm{d}\mathbf{C})\mathbf{C}^{-1}\mathbf{W}). \tag{19b}
\end{aligned}
$$

Comparing with (12) we realize that the first two terms may be attributed the Gaussian likelihood, and we thus choose to attribute the last term to the prior. This yet unknown prior must have the differential,

$$d \log P(\mathbf{C}) = -\frac{1}{2} \operatorname{tr}(-m\mathbf{C}^{-1}(d\mathbf{C})\mathbf{C}^{-1}\mathbf{W}) \quad (20a)$$

$$= -\frac{1}{2} \operatorname{tr}(-m\mathbf{W}\mathbf{C}^{-1}(d\mathbf{C})\mathbf{C}^{-1}) \quad (20b)$$

$$= -\frac{1}{2} \operatorname{tr}(m\mathbf{W}d\mathbf{C}^{-1}). \quad (20c)$$

Thus, the prior must be a linear function in $\mathbf{C}^{-1}$, and we conclude that

$$P(\mathbf{C}) = \exp\left(-\frac{1}{2} \operatorname{tr}(m\mathbf{W}\mathbf{C}^{-1} + \mathbf{B})\right) \quad (21a)$$

$$= k \exp\left(-\frac{m}{2} \operatorname{tr}(\mathbf{W}\mathbf{C}^{-1})\right), \quad (21b)$$

for a suitable, constant matrix $\mathbf{B}$ normalizing the integral of $P(\mathbf{C})$ by $k = \exp(-\frac{1}{2} \operatorname{tr}(\mathbf{B}))$. When $\mathbf{W}$ is positive semi-definite, this is identified to be an inverted Wishart distribution [22, Chap. 7.7]. A Wishart distribution $W(\Sigma, n)$, is the distribution of matrices $\mathbf{X}^T\mathbf{X}$, where the p columns of $\mathbf{X}$ are random n-dimensional vectors identically but independently drawn from a normal distribution with mean value $\overline{x}$ and covariance Σ, $N(\overline{x}, \Sigma)$. If $\mathbf{C}$ is distributed as $W(\Sigma, n)$, then $\mathbf{C}^{-1}$ is distributed as (21), where $\mathbf{W} = \Sigma^{-1}$.

Our matrix $\mathbf{W}$ defined in (2) has non-negative eigenvalues and therefore is positive semi-definite, when $f(d_{ab})$ is diagonally dominant. This is proven by the following 3 steps:

1. If Ω has non-negative eigenvalues then so does $\mathbf{W}$: Given two matrices $\mathbf{R}$ and $\mathbf{S}$ with eigenvalues λ_i and μ_j, their Kronecker product $\mathbf{R} \otimes \mathbf{S}$ have eigenvalues $\lambda_i \mu_j$ [20, Theorem 2.1]. Since $\mathbf{1}_{3\times3}$ has eigenvalues $\mu_1 = 3, \mu_2 = \mu_3 = 0$, then $\mathbf{W} = \Omega \otimes \mathbf{1}_{3\times3}$ has non-negative eigenvalues, when Ω has non-negative eigenvalues.
2. The matrix Ω has positive eigenvalues when it is diagonally dominant: Since Ω has positive diagonal, and according to Gershgorin's theorem [24, as presented in [25]] all diagonally dominant matrices with positive diagonals have positive eigenvalues.
3. The matrix Ω is diagonally dominant, when $f(d_{ab})$ is diagonally dominant, i.e. for all $a = 1 \ldots n$

$$|f(d_{aa})| \geq \sum_{a=1, a\neq b}^{m} |f(d_{ab})|: \quad (22)$$

Using the Hadamard product, $\mathbf{R} \odot \mathbf{S} = \{R_{ab}S_{ab}\}$, and $F_{ab} = f(d_{ab})$, we may rewrite (2) as,

$$\Omega = \frac{1}{2}\mathbf{F} \odot (\mathbf{N}^T\mathbf{N} + \mathbf{1}_{n\times n}), \quad (23)$$

where $\mathbf{N} = [n_1 \mid n_2 \mid \ldots \mid n_n]$. The matrix $\mathbf{F}$ has a positive diagonal, since $f(d_{ii}) = 1$. Further, assuming that $\mathbf{F}$ is diagonally dominant, we find that $\mathbf{F} \odot \mathbf{N}^T\mathbf{N}$ is diagonally dominant, since the element of $|n_i \cdot n_j| \leq 1$ and 1 along the diagonal. The sum, $\mathbf{F} \odot (\mathbf{N}^T\mathbf{N} + \mathbf{1}_{n\times n})$, has 2 along the diagonal, and its off diagonal is at most $f(d_{ij})(|n_i \cdot n_j| + 1)$, hence the sum is also diagonally dominant, and hence Ω is diagonally dominant.

We conclude, that Ω is positive-definite, when $\mathbf{F}$ is diagonally dominant, in which case $\mathbf{W}$ is positive semi-definite.

4 Reconstruction with Elasticity

For the practical construction of a missing tooth, e.g. when a dentist needs to design a crown, we need to align the patients teeth with our statistical model for the non-missing teeth. This is a missing data problem, which we solve again using Bayes estimation.

Let y be an incomplete shape vector with $l < n$ points, and x be the corresponding full shape. Then, we wish to find a linear transformation $\mathbf{L} : \mathbb{R}^n \mapsto \mathbb{R}^l$ such that

$$y = \mathbf{L}x. \quad (24)$$

This system is an overdetermined system of equations. Since we cannot expect to find a linear combination of the training samples that solves (24) exactly. Instead the values of x can be found by minimizing an energy functional,

$$E(x) = \|\mathbf{L}x - y\|^2, \quad (25)$$

which may be solved using the linear least squares method. Assume now that y has been subtracted with the (dimension reduced) mean $\overline{x}$, so that a model approximation can be calculated as $x = \Phi b$. Inserting this into (25) we get

$$E(b) = \|\mathbf{L}(\Phi b) - y\|^2. \quad (26)$$

However, a fundamental problem with least square fitting of a model to data is that of *over-fitting*. Hence, we formulate the problem in the Bayes setting: Given an incomplete shape vector y, the reconstruction problem consist of finding the optimal model coefficients b for y. In terms of probability:

$$P(b \mid y) = \frac{P(y \mid b)P(b)}{P(y)} \propto P(y \mid b)P(b). \quad (27)$$

This states that the optimal coefficients b will be the ones with maximum probability, conditioned to y. Both the prior probability $P(b)$ and the likelihood $P(y \mid b)$ can be derived from the shape model definition. We use a normally distributed on b with a zero mean and covariance matrix equal to

the identity, $\boldsymbol{b} \sim N(0, \mathbf{I})$, the probability density can then be written as:

$$P(\boldsymbol{b}) = (2\pi)^{-\dim(\boldsymbol{b})/2} \exp\left(-\frac{\|\boldsymbol{b}\|^2}{2}\right), \tag{28}$$

where $\dim(\boldsymbol{b})$ is the dimensionality of vector $\boldsymbol{b}$, and the probability density of the likelihood

$$P(\boldsymbol{y} \mid \boldsymbol{b}) = (2\pi\sigma^2)^{-l/2} \exp\left(-\frac{\|(\mathbf{L}(\boldsymbol{\Phi}\boldsymbol{b}) - \boldsymbol{y}\|^2}{2\sigma^2}\right). \tag{29}$$

Using an eigenvalue decomposition $\mathbf{L}\boldsymbol{\Phi} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix, the point of maximum posterior is found to be [26]

$$\boldsymbol{x} = \overline{\boldsymbol{x}} + \boldsymbol{\Phi}\mathbf{V}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^2 + \sigma^2\mathbf{I})^{-1}\mathbf{U}^T(\boldsymbol{y} - \mathbf{L}\overline{\boldsymbol{x}}) \tag{30}$$

where $\mathbf{I}$ is the identity matrix of appropriate dimensions. The reconstruction error for various missing landmarks is shown in Table 2. The reduced elastic model results in a significantly better reconstruction, compared to the normal covariance reconstruction.

The reconstruction with the elastic models showed that some kind of regularization is necessary, due to the increased number of eigenmodes and the effect of the neighborhood relations on the reconstruction. As the complexity of the reconstruction problem increases, the resulting shape quickly becomes very distorted. This brought us to a general regularizing expression, where the number of least significant elastic eigenmodes to be removed increases with the

Table 2 Mean reconstruction error of molar training sample for the normal model and the elastic model that includes regularized elasticity

Removed landmarks out of 44 total	Normal model $E_{\mathbf{r}_{rec}}$	Elastic model $E_{\mathbf{r}_{rec}}$
1	0.0016	0.00075
5	0.022	0.0068
12	0.049	0.017
36	0.66	0.66
36 distr	0.20	0.12

number of landmarks to be reconstructed vs. the indata. Our experiments showed that only a small number of artificial eigenmodes should be used for optimal reconstruction, as illustrated in Fig. 4, and in complicated cases up to 90% are removed for the best result. With this approach we are now able to reconstruct hitherto unseen teeth structure such as shown in Figs. 5 and 6. Figure 6 illustrates the idea of adding more teeth to the shape model, and thus taking advantage of the bite information; For reconstruction a tooth, we can add landmarks from e.g. its antagonist, neighboring teeth and (in this case) opposite tooth, gaining a reconstruction that conforms better to the surroundings.

5 Bite Constrained Reconstruction

One of the limitations of a point-based shape model is the undefined limit of shape variation in each eigenmode. We assume limits of three standard deviations of the mean, assuming a Gaussian data distribution. In PCA, when fitting the data to an affine subspace, we cannot guarantee the displacement vectors in some eigenmode not to overlap. The approximated limits on the modes of variation makes it a common problem that the point-based models produces shapes with illegal border-overlaps. This is a problem of particular importance when modeling multiple tooth shapes with one model, since the surfaces of neighboring and antagonist teeth in a natural bite are not only close, they also share one or more contact points, and thus inter-model border overlapping is very likely.

In the reconstruction, we wish to maximize the posterior probability of the model parameters $\boldsymbol{b}$ given the incomplete shape vector $\boldsymbol{y}$, by minimizing $\|(\mathbf{L}\boldsymbol{\Phi})\boldsymbol{b} - \boldsymbol{y}\|^2$. Let $\boldsymbol{p}_y$ be the landmark in $\boldsymbol{y}$ that was just given a new position away from the collision, and let $\boldsymbol{p}_m$ be the corresponding landmark in the model to approximate $\boldsymbol{p}_y$. Let $\boldsymbol{n}$ be a normal vector in the direction of the penetration. Then, by minimizing the distance between the landmarks as

$$\|\boldsymbol{n}^T\boldsymbol{p}_m - \boldsymbol{n}^T\boldsymbol{p}_y\|^2 \tag{31}$$

the distance between the landmarks is only measured in the direction of $\boldsymbol{n}$. This expression increases the possibility of
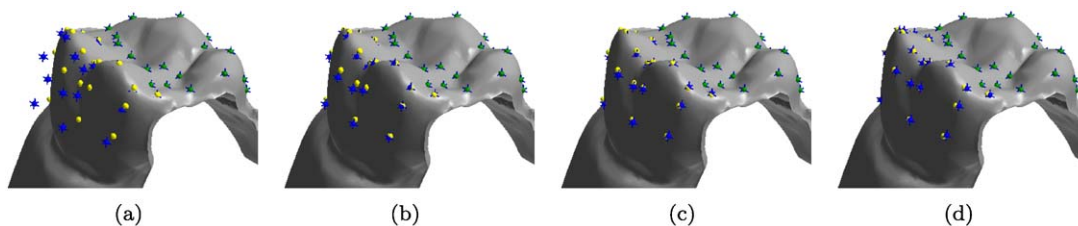

(a)　　　　　　　　(b)　　　　　　　　(c)　　　　　　　　(d)

Fig. 4 (Color online) Reconstruction molar training sample with elastic model, and different amount of elastic eigenmodes removed. Three sets of landmarks are shown: original (*green*), the reconstructed (*blue*), and the ground truth (*yellow*). The number of eigenmodes removed are: (**a**) 0%, (**b**) 50%, (**c**) 60%, and (**d**) 80%
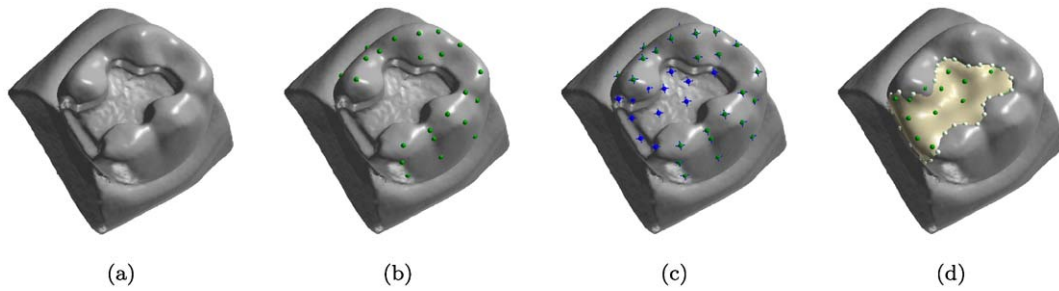
Fig. 5 (Color online) Reconstruction molar from elastic shape model. (**a**) Original tooth with surface part to be reconstructed. (**b**) Landmarks to reconstruct from. (**c**) Reconstructed landmarks (*blue*). (**d**) Reconstruction with surface mesh
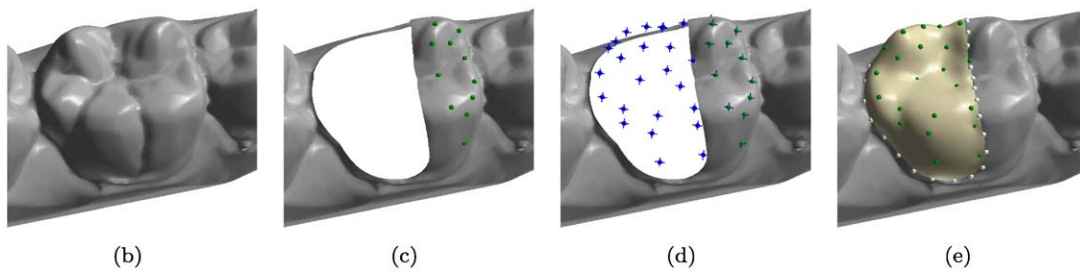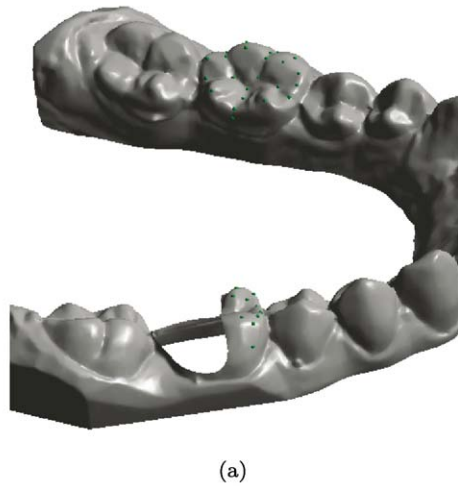

Fig. 6 (Color online) Reconstruction of part of the lower left molar (46) from the elastic model containing lower 1st molar (36) and (46), (36) supplying main cusp feature landmarks as illustrated. (**b**) shows the original "Ground Truth"; (**a**) and (**c**) the part removed including the landmarks used for reconstruction; (**d**) the guessed surface points; and (**e**) the reconstructed surface
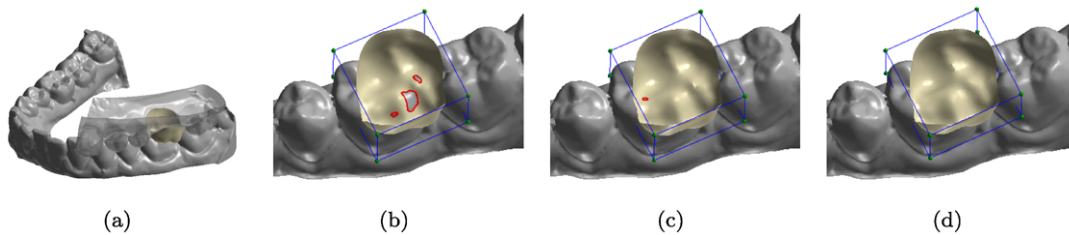

Fig. 7 (Color online) Reconstruction of upper right molar (26) from the elastic model containing the upper right premolar (25), molar (26) and lower right molar (36), handling collisions between antagonists (26) and (36). Iterations in the collision response algorithm. (**a**) Overview. (**b**)–(**d**) iterative reduction in collisions

estimating a p_m in a non-collision position, by relaxing the movement of p_m in the plane through p_y with the normal n, thus motivating the most probably p_m close to this plane. We will also refer to this as a *plane constraint*. To respond to collisions, we apply an iterative algorithm, where we in each step, seek the landmark of deepest penetration. This landmark is then pushed out along the surface normal and apply the plane constraint, i.e. require, that this landmark no longer can move in the normal direction. Steps from this algorithm is illustrated in Fig. 7.

6 Conclusion

We have presented a system for reconstructing teeth based on an extension of the Principal Component Analysis on shape models containing selected teeth in the human bite. Our extensions include both an elasticity term for the covariance matrix and collision avoidance for antagonist teeth, and we have given conditions under which the elasticity is a Bayes estimate. The conclusion is, that the reconstruction generalize well in terms of missing data, collisions are minimized for improved biting, and preliminary clinical evaluation indicate that the resulting models visualized by variational implicit surfaces are more natural looking than standard reconstructions.

One difficulty in point-based shape model is implementing hard constraints such as collisions with antagonist teeth. We have extended our system with a simple collision detection mechanism, where we respond to collisions by applying an iterative algorithm, that seek the landmark of deepest penetration. This may be considered a projection method, and future work includes finding better ways to incorporate of hard constraints in the Principle Component Analysis reconstruction.

References

1. Modgil, S., Hutton, T.J., Hammond, P., Davenport, J.C.: Combining biometric and symbolic models for customised, automated prosthesis design. Artif. Intell. Med. **25**, 227–245 (2002)
2. Gürke, S.: Restoration of teeth by geometrically deformable models (1997). http://citeseer.comp.nus.edu.sg/gurke97restoration.html
3. Hayashi, T., Tsuchida, J., Kato, K.: Semi-automatic design of tooth crown using a 3-D dental CAD system, Vocs-1B. In: Proceedings of the 22nd Annual EMBS International Conference, Chicago IL, USA, July 2000, pp. 565–566 (2000)
4. Blanz, V., Mehl, A., Veter, T., Seidel, H.P.: A statistical method for robust 3D surface reconstruction from sparse data. In: Int. Symp.

on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece (2004)
5. Hommelhoff Jensen, K., Sporring, J.: Reconstructing teeth with bite information. In: Ersbøll, B., Pedersen, K.S. (eds.) Proceedings of the Scandinavian Conference on Image Analysis (SCIA '07). Lecture Notes in Computer Science, vol. 4522, pp. 102–111. Springer, Berlin (2007)
6. Pearson, K.: On lines and planes of closest fit to systems of points in space. Philos. Mag. **6**(2), 559–572 (1901)
7. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. **24**, 417–441 and 498–520 (1933).
8. Bookstein, F.L.: Shape and the information in medical images: A decade of morphometric synthesis. Comput. Vis. Image Underst. **66**(2), 97–118 (1997)
9. Cootes, T.F., Taylor, C.J.: A mixture model for representing shape variation. Image Vis. Comput. **17**, 567–573 (1999)
10. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. Wiley, New York (1998)
11. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comput. Vis. Image Underst. **61**(1) (1995)
12. Cootes, T.F., Taylor, C.J.: Statistical models of appearance for computer vision. Technical Report, University of Manchester (March 2004). http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf
13. de Bruijne, M., Lund, M.T., Tankó, L.B., Pettersen, P.C., Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. Med. Image Anal. (2007)
14. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
15. Pizer, S., Thall, A., Chen, D.: M-reps: A new object representation for graphics. Technical Report, University of North Carolina (1999)
16. Hutton, T.J., Buxton, B.F., Hammond, P.: Dense surface point distribution models of the human face. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), p. 153 (2001)
17. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proc. of SIGGRAPH '99, Los Angeles, August 1999, pp. 187–194 (1999)
18. Turk, G., O'Brien, J.F.: Variational implicit surfaces. Technical Report, Georgia Institute of Technology (May 1999). Tech Report GIT-GVU-99-15
19. Wang, Y., Staib, L.H.: Boundary finding with correspondence using statistical shape models. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 338–345 (1998)
20. Magnus, J.R., Neudecker, H.: Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York (1988)
21. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)
22. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley, New York (2003)
23. Rissanen, J.: Stochastic Complexity in Statistical Inquiry. World Scientific, Singapore (1989)
24. Gerschgorin, S.: Über die abgrenzung der eignewerte einer matrix. Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk 7, 749–754 (1931)
25. Gershgorin circle theorem. http://en.wikipedia.org/wiki/Gershgorin_circle_theorem, September 8 2007
26. Blanz, V., Vetter, T.: Reconstructing the complete 3d shape of faces from partial information. Technical Report, University of Freiburg (2001). Computer Graphics Technical Report No. 1

Jon Sporring received his Master and Ph.D. degree from the Department of Computer Science, University of Copenhagen, Denmark in 1995 and 1998, respectively. Part of his Ph.D. program was carried out at IBM Research Center, Almaden, California, USA. Following his Ph.D, he worked as a visiting researcher at the Computer Vision and Robotics Lab at Foundation for Research & Technology—-Hellas, Greece, and as assistant research professor at 3D-Lab, School of Dentistry, University of Copenhagen. Since 2003 he has been employed as associate professor at the Department of Computer Science, University of Copenhagen, and since 2008 he has been Vice-Chair for Research at Department of Computer Science. His main topics of research are medical image processing, medical computer graphics, and information theory.

Katrine Hommelhoff Jensen received her Master degree from the Department of Computer Science, University of Copenhagen, Denmark in 2006. During her studies she held several positions as software developer among others at 3Shape A/S, Denmark. Currently she is employed at FOLLOW-ME! Technology Systems GmbH, Germany. Her research interests include: Medical image analysis, machine learning and 3D computer graphics.