Reconstructing the Complete 3D Shape of Faces from Partial Information

Computer Graphics Technical Report No. 1

Volker Blanz, Thomas Vetter

Based on the assumption that a class of objects or data can be represented as a vector space spanned by a set of examples, we present a general method to estimate vector components of a novel vector, given only a subset of its dimensions.
We apply this method to recover 3D shape of human faces from 2D image positions of a small number of feature points. The application demonstrates two aspects of the estimation of novel vector components: (1) From 2D image positions, we estimate 3D coordinates, and (2) from a small set of points, we obtain vertex positions of a high-resolution surface mesh. We provide an evaluation of the technique on laser scans of faces, and present an example of 3D shape reconstruction from a photograph.
Our technique involves a tradeoff between reconstruction of the given measurements, and plausibility of the result. This is achieved in a Bayesian approach, and with a statistical analysis of the examples.

1 Introduction

Arguments by analogy are a useful mode of reasoning if we lack sufficient information about a problem for a rigirous conclusion, but are provided with many instances of solutions for similar settings. In this paper, we address the problem of estimating the components of a vector, given only some of the components' values. More generally, the input may be the result of any linear mapping to a lower dimensional space. The prior knowledge that helps to solve this ill-posed problem is represented by a set of examples of vectors, and the assumption that any novel solution is in the span of these examples. Moreover, we exploit the statistical properties of the examples to obtain an estimate of prior probability. The correlation of vector components within the set of examples is the core property that makes an estimate of unknown vector components possible.

As an example of a vector space of objects, we apply our method to the geometry of faces. The morphable face model approach [1] provides a representation of facial shapes in terms of shape vectors, such that any linear combination of vectors describes a realistic face. Shape vectors are defined by concatenating the x, y, and z coordinates of a large set of surface points to a single, high-dimensional vector. The technique for selecting these surface points on individual faces ensures that each component of the shape vector refers to corresponding points on all faces, such as the tip of the nose.

In this paper, we estimate full 3D structure of a face from 2D image positions of a subset of the morphable model's vertices. Image positions are taken from a front view of the face, and with orthographic projection. However, the system can also be applied to any other viewing direction, or a combination of views. Restricted to linear mappings of the original data, the system cannot handle perspective projection from close viewpoints. For larger distances, the difference between perspective and orthographic projection decreases, and our technique provides realistic results.

The morphable face model has previously been used to estimate 3D shape from a single image [1]. Comparing color values of the image with those obtained from the model, this system iteratively matches the morphable model to the image. Similar to the approach presented here, the system relies on the vector space structure of faces for estimating 3D structure. However, it also exploits shading information from the image. Matching the entire facial surface to the image, the result recovers many facial details. In contrast, the method presented here relies only on a relatively small set of feature points provided by the user. However, the matching problem solved here is computationally much simpler, and can be solved in a single step in a robust way. Therefore, the algorithm is considerably faster and may be applied in interactive tools for face reconstruction.

We extend and generalize a method that has been applied to estimate dense optic flow fields in image data [4], using a data set of flow vectors obtained by a 2D projection of a 3D morphable face model. The modifi-

University of Freiburg, Georges-Köhler-Allee 52, 79110 Freiburg, http://graphics.informatik.uni-freiburg.de

cation presented in this paper makes the system more robust, which proves to be crucial to achieve high overall quality of the estimate.

The problem addressed in this study is related to the statistical problem of regression. In regression, a set of measurements (x_i, y_i) of a random variable y for different values of the known parameter x is used to estimate the expectation value $y(x)$ at any x. Regression techniques select a function y from a family of functions, which can be linear mappings, polynomials, or any other function space. If the capacity of the function space is too large, some methods produce overfitting effects (see [2]): the function fits the measurements precisely, but varies drastically in between, rather than being smooth. The desired generalization of $y(x)$ to novel values of x tends to be poor.

As we demonstrate in Section 6, a similar effect may occur here, if the low-dimensional input vector is subject to noise or other sources of error, or if the desired solution cannot be entirely captured by the model.

To overcome the problem of overfitting, most regression techniques impose a smoothness constraint on the solution, or restrict the family of functions [7, 2]. In our approach, we restrict solutions to the span of a set of examples, and impose an additional penalty on solutions far from the observed average. The result will be a trade-off that is both plausible a priori, and still fits the given measurements well.

In the following section, we give a definition of object classes in terms of a probabilistic criterion for class membership. Section 3 presents a direct approach to estimating vector components from sparse data. Section 4 derives a framework to avoid overfitting and accomodate noisy measurements. Section 5 discusses a special case that relates the theory to a straightforward projection into the span of examples. In Section 6, we present results obtained with 3D models of faces.

2 Representation of Class-Specific Knowledge

We assume that the examples of class elements

$$\mathbf{v}_i \in I\!R^n, \qquad i = 1, \ldots, m \tag{1}$$

are given in a vector space representation such that linear combinations

$$\mathbf{v} = \sum_{i=1}^{m} a_i \mathbf{v}_i \tag{2}$$

describe new elements of the class. However, the coefficients of the linear combinations must be restricted by additional conditions to ensure realistic results.[1]

An estimate of the prior probability of vectors within the span of examples can be obtained by a Principal Component Analysis (PCA, see [3]). The original data are centered around the origin by subtracting the arithmetic mean

$$\mathbf{x}_i = \mathbf{v}_i - \overline{\mathbf{v}}, \qquad \overline{\mathbf{v}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{v}_i, \tag{3}$$

and concatenated to a data matrix

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) \in I\!R^{n \times m}. \tag{4}$$

The covariance matrix of the data set is given by

$$\mathbf{C} = \frac{1}{m} \mathbf{X}\mathbf{X}^T = \frac{1}{m} \sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j^T \in I\!R^{n \times n}, \tag{5}$$

PCA is based on a diagonalization of the covariance matrix,

$$\mathbf{C} = \mathbf{S} \cdot diag(\sigma_i^2) \cdot \mathbf{S}^T. \tag{6}$$

Since $\mathbf{C}$ is symmetrical, the columns $\mathbf{s}_i$ of $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots)$ form an orthogonal set of eigenvectors. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m$ are the standard deviations within the data along each eigenvector $\mathbf{s}_i$. The diagonalization of $\mathbf{C}$ can be calculated by a Singular Value Decomposition (SVD, [5]) of $\mathbf{X}$.

Having subtracted the arithmetic mean, the m vectors $\mathbf{x}_i$ are linearly dependent, so their span is at most $m' = (m-1)$ dimensional, and the rank of $\mathbf{X}$ and $\mathbf{C}$ is at most m'. Therefore, $\sigma_m = 0$, and $\mathbf{s}_m$ is irrelevant.

In the following, we use the eigenvectors as a basis,

$$\mathbf{x} = \sum_{i=1}^{m'} c_i \sigma_i \mathbf{s}_i = \mathbf{S} \cdot diag(\sigma_i)\mathbf{c}. \tag{7}$$

An important property of PCA is that variations along the eigenvectors are uncorrelated within the set of examples. Assuming a normal distribution in each of the directions, the probability density at $\mathbf{x}$ is

$$p(\mathbf{x}) = \prod_{i=1}^{m'} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2}\langle \mathbf{s}_i, \mathbf{x} \rangle^2} = \prod_{i=1}^{m'} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}c_i^2} \tag{8}$$

$$= \frac{1}{(2\pi)^{m'/2} \prod_i \sigma_i} e^{-\frac{1}{2}\|\mathbf{c}\|^2}. \tag{9}$$

[1] Coefficients might be constrained to the convex hull by $a_i \in [0, 1]$ and $\sum_{i=1}^{m} a_i = 1$. The first constraint is replaced here by a probabilistic measure. The second is enforced implicitly by forming linear combinations relative to $\overline{\mathbf{v}}$: Any linear combination $\mathbf{v} = \sum_{i=1}^{m} b_i \mathbf{x}_i + \overline{\mathbf{v}}$ can be shown to satisfy $\sum_{i=1}^{m} a_i = 1$ in terms of (2) and (3).

The probability density for $\mathbf{c}$ can be rescaled to

$$p(\mathbf{c}) = \nu_c \cdot e^{-\frac{1}{2}\|\mathbf{c}\|^2}, \qquad \nu_c = (2\pi)^{-m/2}. \tag{10}$$

The exponent $\|\mathbf{c}\|^2$ is often referred to as Mahalanobis Distance.

3 Incomplete Measurements

Given a measurement $\mathbf{r} \in I\!R^l$, $l < n$, we would like to find the full vector $\mathbf{x} \in I\!R^n$ such that

$$\mathbf{r} = \mathbf{L}\mathbf{x} \tag{11}$$

with a mapping $\mathbf{L} : I\!R^n \mapsto I\!R^l$. $\mathbf{L}$ can be any linear transformation, and does not need to be a projection.

If $\mathbf{L}$ is not a one-to-one mapping, the solution (11) is not uniquely defined. Therefore, we restrict the admissible solutions to the span of $\mathbf{x}_i$. As we cannot expect to find a linear combination of the examples that solves (11) exactly, we compute a vector $\mathbf{x}$ that minimizes

$$E(\mathbf{x}) = \|\mathbf{L}\mathbf{x} - \mathbf{r}\|^2. \tag{12}$$

Let $\mathbf{q}_i = \sigma_i \mathbf{L}\mathbf{s}_i \in I\!R^l$ be the reduced versions of the scaled eigenvectors, and

$$\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, ...) = \mathbf{L}\mathbf{S} \cdot diag(\sigma_i) \in I\!R^{l \times m'}. \tag{13}$$

In terms of model parameters c_i, Equation (12) is

$$E(\mathbf{c}) = \|\mathbf{L}\sum_i c_i \sigma_i \mathbf{s}_i - \mathbf{r}\|^2 = \|\sum_i c_i \mathbf{q}_i - \mathbf{r}\|^2 \tag{14}$$

$$= \|\mathbf{Q}\mathbf{c} - \mathbf{r}\|^2. \tag{15}$$

The optimum can be found by a Singular Value Decomposition [5]

$$\mathbf{Q} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{16}$$

with a diagonal matrix $\mathbf{W} = diag(w_i)$, and $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = id_{m'}$.

The pseudoinverse (see [6]) of $\mathbf{Q}$ is

$$\mathbf{Q}^+ = \mathbf{V}\mathbf{W}^+\mathbf{U}^T, \tag{17}$$

$$\mathbf{W}^+ = diag\begin{pmatrix} w_i^{-1} & \text{if } w_i \neq 0 \\ 0 & \text{otherwise} \end{pmatrix}. \tag{18}$$

To avoid numerical problems, the condition $w_i \neq 0$ may be replaced by a threshold. The minimum of (15) is

$$\mathbf{c} = \mathbf{Q}^+\mathbf{r}, \tag{19}$$

which is optimal in two respects [6]:

1. $\mathbf{c}$ minimizes E, so for all $\mathbf{c}'$, $\quad E(\mathbf{c}') \geq E(\mathbf{c})$.
2. Among the set of solutions $\{\mathbf{c}' | E(\mathbf{c}') = E(\mathbf{c})\}$, $\mathbf{c}$ has minimum norm $\|\mathbf{c}\|$ and thus maximum prior probability (Equation 10).

By Equation (3) and (7), $\mathbf{c}$ is mapped to $I\!R^n$:

$$\mathbf{v} = \mathbf{S} \cdot diag(\sigma_i)\mathbf{c} + \overline{\mathbf{v}}. \tag{20}$$

For solving Equation (11), it might seem more straightforward to compute the pseudoinverse of $\mathbf{L}$ and set $\mathbf{x} = \mathbf{L}^+\mathbf{r}$. However, among vectors with equal error $\|\mathbf{L}\mathbf{x} - \mathbf{r}\|$, this method would return the solution with minimum $\|\mathbf{x}\|$ rather than minimum $\|\mathbf{c}\|$. Vector components x_i that do not affect $\mathbf{L}\mathbf{x}$ would be zero, and the result would not be in the span of the examples.

4 Prior Probability versus Matching quality

The previous solution will always ensure that E is minimized, and in particular that $E = 0$ whenever this is possible. Prior probability is only considered within solutions of equal $E(\mathbf{c})$.

However, it may well be that the measurement $\mathbf{r}$ cannot be fully accounted for by an element $\mathbf{v}$ of the object class. First, $\mathbf{r}$ may be subject to noise or other sources of error, such as wrong assumptions on $\mathbf{L}$. Moreover, we cannot expect to cover the full range of the object class with the set of examples.

Therefore, minimizing $E(\mathbf{x}) = \|\mathbf{L}\mathbf{x} - \mathbf{r}\|^2$ may lead to model coefficients far from the average, and a heavily distorted vector $\mathbf{v}$. To avoid this overfitting, we propose a tradeoff between matching quality and prior probability of the solution. This tradeoff will be derived from a Bayesian approach in the following section.

4.1 Bayesian Approach to Reconstruction

For an element of the model that is defined by model parameters $\mathbf{c}$, a noiseless measurement would be

$$\mathbf{r}_{model} = \mathbf{L}\sum_i c_i \sigma_i \mathbf{s}_i = \sum_i c_i \mathbf{q}_i = \mathbf{Q}\mathbf{c} \tag{21}$$

We assume that each dimension j of the measured vector $\mathbf{r}$ is subject to uncorrelated Gaussian noise with a variance σ_N^2. Then, the likelihood of measuring $\mathbf{r} \in I\!R^l$ is given by

$$P(\mathbf{r}|\mathbf{r}_{model}) = \prod_{j=1}^{l} P(r_j | r_{model,j}) \tag{22}$$

$$= \prod_{j=1}^{l} \nu_N \cdot e^{-\frac{1}{2\sigma_N^2}(r_{model,j} - r_j)^2} \tag{23}$$

$$= \nu_N^l \cdot e^{-\frac{1}{2\sigma_N^2}\sum_j (r_{model,j} - r_j)^2} \tag{24}$$

$$= \nu_N^l \cdot e^{-\frac{1}{2\sigma_N^2}\|\mathbf{r}_{model} - \mathbf{r}\|^2} \tag{25}$$

with a normalization factor ν_N. In terms of the model parameters $\mathbf{c}$, the likelihood is

$$P(\mathbf{r}|\mathbf{c}) = \nu_N^l \cdot e^{-\frac{1}{2\sigma_N^2}\|\mathbf{Qc}-\mathbf{r}\|^2}. \quad (26)$$

Given an observed vector $\mathbf{r}$, we are looking for the estimate $\mathbf{c}$ with maximum probability. According to Bayes Rule [2], this posterior probability is given by

$$P(\mathbf{c}|\mathbf{r}) = \nu \cdot P(\mathbf{r}|\mathbf{c}) \cdot p(\mathbf{c}). \quad (27)$$

with a constant factor $\nu = (\int P(\mathbf{r}|\mathbf{c}') \cdot p(\mathbf{c}')d\mathbf{c}')^{-1}$.
Substituting (10) and (26) yields

$$P(\mathbf{c}|\mathbf{r}) = \nu \cdot \nu_N^l \cdot \nu_c \cdot e^{-\frac{1}{2\sigma_N^2}\|\mathbf{Qc}-\mathbf{r}\|^2} \cdot e^{-\frac{1}{2}\|\mathbf{c}\|^2}, \quad (28)$$

which is maximized if the cost function

$$E = -2 \cdot logP(\mathbf{c}|\mathbf{r}) = \frac{1}{\sigma_N^2}\|\mathbf{Qc}-\mathbf{r}\|^2 + \|\mathbf{c}\|^2 + const. \quad (29)$$

is minimized.

4.2 Combined Cost Function

In this section, we show that the cost function (29) can be minimized in a single step. To simplify the calculation, we introduce a weight factor $\eta = \sigma_N^2 \geq 0$ and minimize

$$E = \|\mathbf{Qc}-\mathbf{r}\|^2 + \eta \cdot \|\mathbf{c}\|^2. \quad (30)$$

This can be expanded to

$$E = \langle \mathbf{Qc}, \mathbf{Qc} \rangle - 2\langle \mathbf{Qc}, \mathbf{r} \rangle + \|\mathbf{r}\|^2 + \eta \cdot \|\mathbf{c}\|^2 \quad (31)$$

$$E = \langle \mathbf{c}, \mathbf{Q}^T\mathbf{Qc} \rangle - 2\langle \mathbf{c}, \mathbf{Q}^T\mathbf{r} \rangle + \|\mathbf{r}\|^2 + \eta \cdot \|\mathbf{c}\|^2 \quad (32)$$

In the optimum,

$$0 = \nabla E = 2\mathbf{Q}^T\mathbf{Qc} - 2\mathbf{Q}^T\mathbf{r} + 2\eta\mathbf{c}, \quad (33)$$

so

$$\mathbf{Q}^T\mathbf{Qc} + \eta\mathbf{c} = \mathbf{Q}^T\mathbf{r}. \quad (34)$$

Singular Value Decomposition $\mathbf{Q} = \mathbf{UWV}^T$ yields[2]

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{VWU}^T\mathbf{UWV}^T = \mathbf{VW}^2\mathbf{V}^T. \quad (35)$$

From (34), we obtain

$$\mathbf{VW}^2\mathbf{V}^T\mathbf{c} + \eta\mathbf{c} = \mathbf{VWU}^T\mathbf{r}. \quad (36)$$

[2] The matrix $\mathbf{U} \in I\!R^{l \times m'}$ computed by SVD has the following property [5]: If $m' \leq l$, $\mathbf{U}^T\mathbf{U} = id_{m'}$. If $m' > l$, only the first m' columns of $\mathbf{U}$ are orthogonal, while the others are 0, so $\mathbf{U}^T\mathbf{U}$ is not the full identity matrix. However, $w_i = 0$ for $i > m'$, so $\mathbf{WU}^T\mathbf{UW} = \mathbf{W}^2$ still holds.

Multiplying by $\mathbf{V}^T$, this can be solved for $\mathbf{c}$:[3]

$$\mathbf{W}^2\mathbf{V}^T\mathbf{c} + \eta\mathbf{V}^T\mathbf{c} = \mathbf{WU}^T\mathbf{r} \quad (37)$$

$$diag(w_i^2 + \eta) \cdot \mathbf{V}^T\mathbf{c} = \mathbf{WU}^T\mathbf{r} \quad (38)$$

$$\mathbf{V}^T\mathbf{c} = diag(\frac{w_i}{w_i^2+\eta})\mathbf{U}^T\mathbf{r} \quad (39)$$

$$\mathbf{c} = \mathbf{V}diag(\frac{w_i}{w_i^2+\eta})\mathbf{U}^T\mathbf{r} \quad (40)$$

Note that in the special case $\eta = 0$, this equivalent to Equation (19).
The overall result is

$$\mathbf{x} = \sum_i c_i\sigma_i\mathbf{s}_i = \mathbf{S}diag(\sigma_i)\mathbf{c} \quad (41)$$

$$= \mathbf{S}diag(\sigma_i)\mathbf{V}diag(\frac{w_i}{w_i^2+\eta})\mathbf{U}^T\mathbf{r}. \quad (42)$$

and

$$\mathbf{v} = \mathbf{x} + \overline{\mathbf{v}}. \quad (43)$$

5 Special case $\mathbf{L} = id_n$

In some applications it may be desirable to find the closest element of the span of examples from a vector $\mathbf{x}$ that is entirely known, or to approximate a given element of the span by a more plausible solution. Both cases are covered by the previous results if we set $\mathbf{L} = id_n$.
If $\mathbf{L} = id_n$, the Singular Value Decomposition of $\mathbf{Q}$ is trivial

$$\mathbf{Q} = \mathbf{S} \cdot diag(\sigma_i) = \mathbf{UWV}^T \quad (44)$$

with the orthogonal matrix $\mathbf{U} = \mathbf{S}$, the diagonal matrix $\mathbf{W} = diag(\sigma_i)$, and $\mathbf{V} = id_{m'}$. Then, Equation (42) reduces to

$$\mathbf{x} = \mathbf{S} \cdot diag(\sigma_i)id_{m'} \cdot diag(\frac{\sigma_i}{\sigma_i^2+\eta})\mathbf{S}^T\mathbf{r} \quad (45)$$

$$= \mathbf{S} \cdot diag(\frac{1}{1+\frac{\eta}{\sigma_i^2}})\mathbf{S}^T\mathbf{r} \quad (46)$$

$$= \sum_i \frac{1}{1+\frac{\eta}{\sigma_i^2}} \langle \mathbf{s}_i, \mathbf{r} \rangle \mathbf{s}_i. \quad (47)$$

The most relevant dimensions $\mathbf{s}_i$ with large standard deviation σ_i are affected less by η than those with small σ_i. In the special case $\eta = 0$, $\mathbf{x}$ is given by a simple projection

$$\mathbf{x} = \sum_i \langle \mathbf{s}_i, \mathbf{r} \rangle \mathbf{s}_i. \quad (48)$$

[3] If $(w_i^2 + \eta) = 0$, which only occurs if $w_i = \eta = 0$, we replace $\frac{w_i}{w_i^2+\eta}$ by 0, as we did for the pseudoinverse.

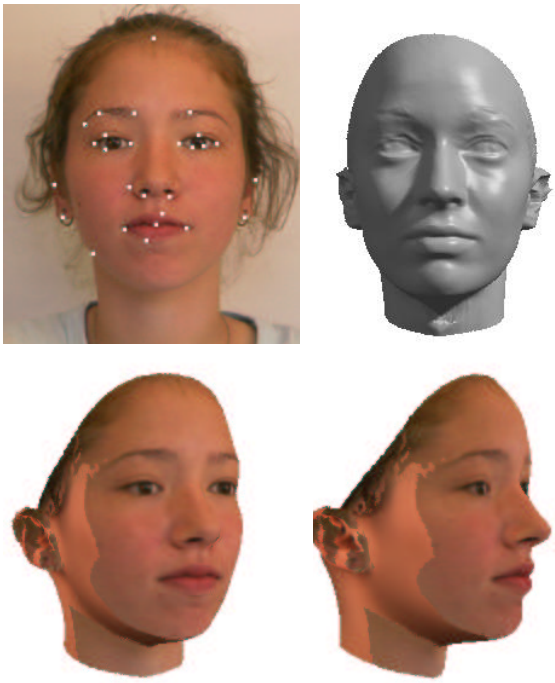

Figure 1: From 26 feature coordinates manually defined on the original image (top left), the system recovered the overall shape of the face (top right). With an additional texture extraction, color information can be transferred to the 3D model (bottom line) to generate new views. Vectors for translation and scaling (Equation 50) were added to the 99 principal components.

6 Application to Face Data

In the morphable face model [1], facial surface data that were recorded with a laser scanner are represented in shape vectors that combine x, y, and z coordinates of all vertices:

$$\mathbf{v} = (x_1, y_1, z_1, \ldots x_p, y_p, z_p)^T \in I\!R^n,\ n = 3 \cdot p \qquad (49)$$

Sampled at a spacing of less than 1mm, surface is represented by $p = 75972$ vertices. Linear combinations of shape vectors will only produce realistic novel faces if corresponding points, such as the tip of the nose, are represented by the same vector components across all individual shape vectors. This is achieved by establishing dense correspondence between different scans, and forming vectors $\mathbf{v}_i$ in a consistent way.

Along with shape, the morphable face model also represents texture. In this study, texture is not considered, and all images are rendered with the average texture. The method described in this paper could also be applied to texture vectors, filling in occluded regions of the face.

The database of 200 individual faces used in this study has been randomly split into a training set and a test set of $m = 100$ faces each. The training set provides the examples $\mathbf{v}_i$ that are available to the system. From these, we computed $m' = 99$ principal components which are

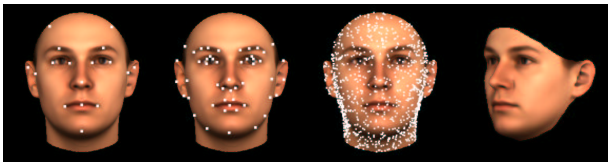

Figure 2: The first three images show the sets of 17, 50, and 1000 feature points used for evaluation. The image on the right illustrates where the error of 3D shape reconstruction was evaluated.

used throughout the following evaluation. The test set provides data for performance assessment on novel faces.

From the vertices of the full model, we selected sets of $f = 17$, 50, or 1000 vertices (Figure 2). The smaller sets are formed by salient points such as the corners of the mouth, that can be identified in an image. The set of 1000 vertices was selected at random.

Computed by orthographic projection in a frontal orientation, the image plane coordinates of these feature points form the vectors $\mathbf{r} \in I\!R^l$, $l = 2 \cdot f$, that are used for evaluation.

Projection and orientation also define the mapping $\mathbf{L}$, which is assumed to be known. For real images, it is important that the system can automatically adapt at least to translation and scaling. This is achieved if vectors

$$\mathbf{s}_{tx} = (1, 0, 0, 1, 0, 0, \ldots)^T, \quad \mathbf{s}_{ty}, \quad \mathbf{s}_{tz}, \quad \text{and} \qquad (50)$$

$$\mathbf{s}_s = \overline{\mathbf{v}} \qquad (51)$$

are added to the principal components in $\mathbf{S}$.

The evaluation of the algorithm is based on the following quantities, which are averaged across all 100 training or test faces:

- $E_r = \|\mathbf{Qc} - \mathbf{r}\|$, the image plane matching error for all feature points, measured in units of pixels in a 300x300 image.
- $\|\mathbf{c}\|$, the Mahalanobis distance of the resulting face from the average.
- The per-vertex average of distances in 3D space between reconstruction and original, computed over the entire set of vertices:

$$E_{full} = \frac{1}{p} \sum_{i=1}^{p} \| \begin{pmatrix} x_{i,reconst.} \\ y_{i,reconst.} \\ z_{i,reconst.} \end{pmatrix} - \begin{pmatrix} x_{i,orig.} \\ y_{i,orig.} \\ z_{i,orig.} \end{pmatrix} \| \qquad (52)$$

  The neck and the top of the forehead are ignored in this measure, as shown in Figure 2.

For 99 principal components and 50 feature points, the computation takes 1.6 seconds on an SGI O_2 with R12000 processor. This includes forming $\mathbf{Q}$ from the much larger matrix $\mathbf{S}$, SVD of $\mathbf{Q}$, and computation of the full face model with 75972 vertices. Computation time depends mainly on the dimensions of $\mathbf{S}$.

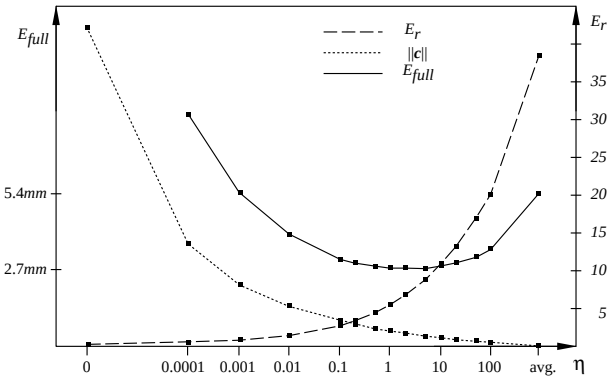

Figure 3: The effect of η on average reconstruction results for 100 novel faces, given 50 feature points, and using 99 principal components. As η increases, the feature points are matched less precisely, so E_r grows. In contrast, $\|\mathbf{c}\|$ decreases, as the results become more plausible. The overall 3D shape error E_{full} is lowest for a tradeoff between both criteria.

E_{full}	$f = 17$	$f = 50$	$f = 1000$
40 principal comp.	3.21	2.81	2.38
99 principal comp.	3.16	2.72	2.24

Table 1: The average 3D shape error E_{full} for reconstruction of 100 novel faces at optimal η depends on the number of principal components, and the number of feature point positions available.

6.1 Reconstruction of Novel Faces

In this Section, we examine how the technique performs on the test faces that are not included in the set of examples. The image coordinates of feature points provided to the algorithm are computed from the 3D vertex positions of the 3D faces. For $f = 50$, $m' = 99$, and different values of η, errors are plotted in Figure 3, and results are shown in Figure 4. Since the feature point coordinates of the novel faces may be difficult to recover exactly by the model, low values of η lead to overfitting: For $\eta = 0$ and $\eta = 0.0001$, the facial surface is heavily distorted, and the overall error E_{full} is large. Still, the feature point coordinates are precisely recovered, as indicated by the low error E_r.

As η increases, E_r grows, while $\|\mathbf{c}\|$ decreases, indicating that the prior probability of the solution gains more weight in the optimization. As the shape becomes more smooth and more plausible, the overall reconstruction error E_{full} decreases, and reaches its minimum at $\eta = 2$.

If η is too large, the output is too close to the average to fit the data, so both E_r and E_{full} are high. The values on the right in Figure 3 are the baseline obtained with the average head $\overline{\mathbf{v}}$.

Table 1 demonstrates how the number of feature points and principal components affects matching quality. The reduced set of 40 principal components is formed by those dimensions $\mathbf{s}_i$ with maximum variance. As expected, the error E_{full} is lowest with the largest set of feature points and the full set of 99 principal components.

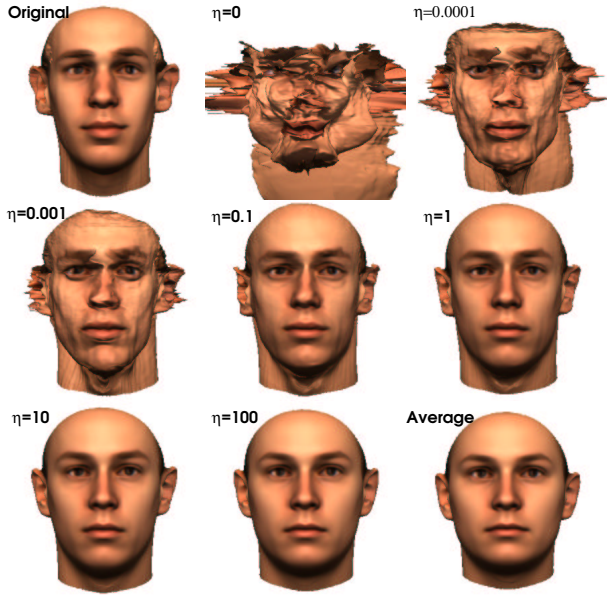

Figure 4: Given the image coordinates of 50 feature points of a novel face (top left), 3D shape was reconstructed with 99 principal components. The result depends on a tradeoff between the precision of feature point matching, and prior probability. This tradeoff is controlled by the parameter η.

6.2 Correct Reconstruction of Training Faces

In this section, we verify that the faces of the training set are exactly recovered by the system, using all $m' = 99$ principal components, if f is large enough, and if the feature point coordinates are precise.

For $f = 50$ and $f = 1000$, the dimension of $\mathbf{r}$ is $l = 2 \cdot f \geq m'$, so the problem $\mathbf{Qc} = \mathbf{r}$ has a unique solution. This solution is recovered by the system, as indicated by the low values of E_r and E_{full} in Table 2, for $\eta = 0$.

In contrast, the solution for $f = 17$ is not unique. Within the set of solutions, the method returns a vector that is closer to the average (smaller $\|\mathbf{c}\|$), yet produces an error $E_{full} > 0$. Still, this is the best guess, given the ambiguous information.

Training data	$f = 17$	$f = 50$	$f = 1000$
E_r	8.9e-5	1.4e-4	3.7e-3
$\|\mathbf{c}\|$	5.8	9.9	9.9
E_{full}	2.1	0.0017	5.6 e-05

Table 2: Average reconstruction errors for all training faces, given different numbers of feature points f. With all 99 principal components and $\eta = 0$, the problem $\mathbf{Qc} = \mathbf{r}$ is solved exactly, so E_r is low . However, for $f = 17$, the solution is not uniquely defined.

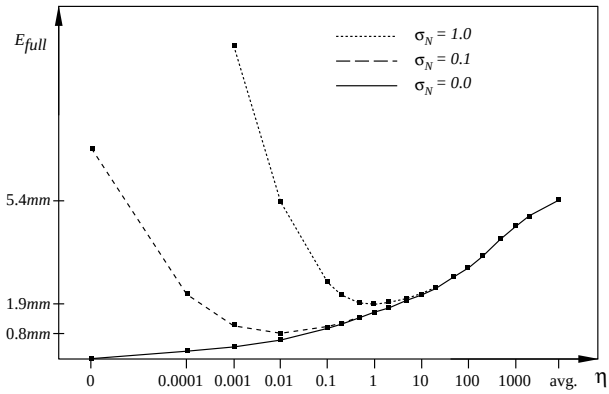

Figure 5: The average shape reconstruction errors for 100 training faces depend on the level of noise σ_N added to each feature point coordinate. While noise-free data are best analyzed with $\eta = 0$ (solid line), reconstruction quality is best at $\eta = \sigma_N^2$ for noisy data.

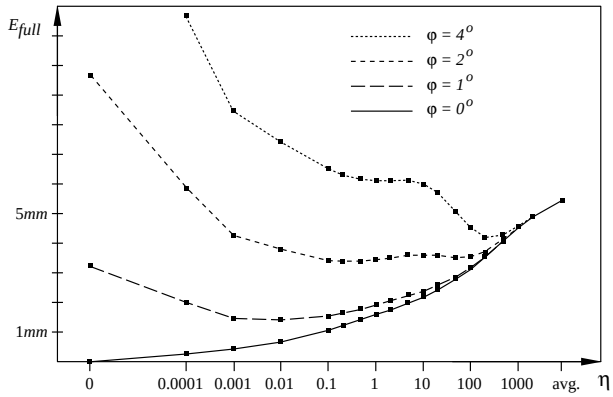

Figure 6: Reconstruction from 50 feature coordinates of the training faces at frontal orientation with an incorrect mapping $\mathbf{L}$ that includes rotations around the vertical axis. 3D shape error E_{full} is reduced by choosing appropriate weights η.

6.3 Noisy Feature Point Coordinates

As discussed in the previous section, the shape of training faces can be recovered perfectly from 50 feature points if their coordinates are precise. In this case, $\eta > 0$ would impair the quality of the result, as shown by the solid line in Figure 5.

However, if Gaussian noise is added to the 2D point coordinates, $\mathbf{r}$ becomes more and more difficult to recover, and overfitting occurs. This is demonstrated by the large errors E_{full} observed for small η if noise with a standard deviation of $\sigma_N = 0.1$ and $\sigma_N = 1$ pixels is added to the horizontal and vertical image coordinates of each feature point.

As we observed previously with novel faces, the values of E_{full} in Figure 5 have a clear minimum for intermediate values of η. In fact, these minima occur at $\eta = \sigma_N^2$, so matching quality is best for the vectors with maximum posterior probability (Section 4.1).

6.4 Robustness with respect to L

A similar effect to noise occurs if the matrix $\mathbf{L}$ used for reconstruction is different from the mapping that produced the feature coordinates in $\mathbf{r}$. This mismatch is relevant for real images, since the geometry of the imaging setup will in general be unknown. In particular, perspective projection produces results that are slightly different from what is simulated by the orthographic projection in $\mathbf{L}$.

Figure 6 shows overall shape errors E_{full} obtained with 50 feature coordinates that were computed for a frontal view. The matrices $\mathbf{L}$ used for reconstruction include rotations of $\phi = 0°$, $1°$, $2°$, and $4°$ around the vertical axis. While the training faces are perfectly recovered with the correct mapping $\phi = 0°$ for $\eta = 0$, performance at angles $\phi > 0°$ is improved significantly with appropriate values of η.

6.5 Results on Real Images

Figure 1 shows an example of 3D shape reconstruction from a set of 26 feature points that were selected by the user. With the limited information about the face, the method cannot capture details of face shape as precisely as an optimization based on color values in the image [1]. However, the overall shape is recovered well, and if texture is extracted from the image [1], the technique provides realistic 3D head models.

7 Conclusion

We have presented a method that infers vector dimensions of data vectors from incomplete measurements. The method is based on a vector space spanned by a set of examples, and on statistical properties of the data. Derived from a Bayesian framework, the technique finds the vector with maximum posterior probability, given the measurement and the examples.

With the vector space of faces provided by a morphable face model, we estimated 3D shape of a high resolution face model from the positions of a small set of feature points in an image. We evaluated reconstruction quality in terms of 3D displacements from the veridical shape of faces, and investigated sensitivity to noise and misalignments.

Clearly, a small set of feature positions is insufficient to recover all details of a face, such as the shape of the nose. However, the technique reliably estimates the overall shape and aligns the 3D face with the image, which can be useful for many application. Since the reconstruction is calculated in a single step, the computation is performed fast enough for interactive tools.

In the future, we are planning to develop methods for choosing the optimal weight factor η by techniques such as cross validation within the training set.

Acknowledgements The database of 200 laser scans was collected by Niko Troje and Tordis Philipps at Max-Planck-Institut für Biologische Kybernetik, Tübingen.

References

[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Computer Graphics Proceedings SIGGRAPH'99*, pages 187–194, Los Angeles, 1999.

[2] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification.* John Wiley & Sons, New York, 2nd edition, 2001.

[3] S.S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, 1998.

[4] B. W. Hwang, V. Blanz, T. Vetter, and S. W. Lee. Face reconstruction from a small number of feature points. In *International Conference on Pattern Recognition, ICPR2000*, Barcelona, Spain, 2000.

[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C : the art of scientific computing.* Cambridge University Press, Cambridge, 1992.

[6] J. Stoer. *Numerische Mathematik I.* Springer, Berlin, 8 edition, 1999.

[7] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1995.