3D Morphable Face Models Revisited

Ankur Patel William A. P. Smith
Department of Computer Science, The University of York
{ankur,wsmith}@cs.york.ac.uk

Abstract

In this paper we revisit the process of constructing a high resolution 3D morphable model of face shape variation. We demonstrate how the statistical tools of thin-plate splines and Procrustes analysis can be used to construct a morphable model that is both more efficient and generalises to novel face surfaces more accurately than previous models. We also reformulate the probabilistic prior that the model provides on the distribution of parameter vector lengths. This distribution is determined solely by the number of model dimensions and can be used as a regularisation constraint in fitting the model to data without the need to empirically choose a parameter controlling the trade off between plausibility and quality of fit. As an example application of this improved model, we show how it may be fitted to a sparse set of 2D feature points (approximately 100). This provides a rapid means to estimate high resolution 3D face shape for a face in any pose given only a single face image. We present experimental results using ground truth data and hence provide absolute reconstruction errors. On average, the per vertex error of the reconstructed faces is less than 3.6*mm.*

1. Introduction

The problem of estimating 3-dimensional face shape from one or more images has attracted considerable attention in recent years [4, 8, 10, 11, 18, 19]. The primary motivation for this work is that 3D shape information provides a pose and illumination invariant description of a face, which can either be used for recognition directly [4, 7, 18], or to produce illumination and pose normalised images for input to a 2D recognition system [2, 18, 19]. The benefits of such an approach are improved robustness to changes in pose, illumination and expression, while still only requiring a single intensity image as input.

Although shape-from-shading provides a possible route to estimating facial shape, the most promising results have been obtained using a statistical model of 3D face shape. The best known work in this area is the 3D morphable model of Blanz and Vetter [3]. This is a low-dimensional parametric model of 3D face shape and texture. To solve the problem of face shape recovery, the challenge is to fit the model to images of previously unseen subjects. This amounts to solving a highly complex nonlinear minimization problem which requires estimation of: 1. shape and texture parameters, 2. pose, scale and position of the subject, 3. camera and surface reflectance parameters and 4. illumination conditions present in the scene. This approach has been shown to be robust and provides high accuracy on real world data. Indeed, the estimated appearance parameters contain useful identity information and provide a route to state-of-the-art performance in face recognition across pose and illumination variation from a single gallery image [16].

The most recent work in this area has focused on developing more sophisticated morphable model fitting algorithms. At the expense of simplifying the reflectance assumptions by using a Lambertian model, Zhang and Samaras [18] showed that a morphable model could be fitted under unknown and arbitrarily complex illumination conditions using a spherical harmonic basis. On the other hand, both Romdhani and Vetter [17] and Moghaddam *et al*. [15] focus on improving the accuracy and efficiency of the fitting process respectively. In both cases, they avoid the problems of local minima in the optimisation function by using features derived from the input images rather than the intensity data itself. Romdhani and Vetter [17] used edges and specular highlights to obtain a smooth cost function, while Moghaddam *et al*. [15] used silhouettes computed from a large number of input images. Knothe *et al*. [13] have begun to consider the problem of model dominance and used local feature analysis to locally improve the fit of the model to a set of sparse feature points.

All of these methods are based on explicitly modelling the underlying physical processes that give rise to an observed image, by rendering each hypothesised appearance. It is not clear that this is either necessary, nor the most practical approach. As already mentioned, the optimisation is complex and prone to becoming trapped in local minima. An alternative approach proposed by Blanz *et al*. [1] instead uses only the position of a sparse set of 2D feature

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

points (perhaps as few as 17). In this case, shape parameters are found by minimising the error between the observed and predicted positions of the feature points in the image plane. The model is also used as a regularisation constraint to balance quality of fit against plausibility.

In this paper we make a number of contributions to the use of morphable models for face shape recovery. First, we provide a new framework for constructing a 3D morphable model from a training set of facial meshes. By making use of techniques from the statistical shape analysis literature, we show how to construct a morphable model whose captured variance is of greater utility in the sense that the generalisation error (i.e. average error when representing out of sample surfaces) is lower for both a fixed number of model dimensions or fixed percentage of total variance captured. Second, we show that the distribution of parameter vector lengths follows a chi-square distribution and discuss how the parameters of this distribution can be used as a regularisation constraint on the length of parameter vectors. Finally, we use our improved model and statistical prior in the setting of fitting a dense 3D morphable model to sparse 2D feature points. We verify empirically that our analytical prediction of the parameter vector length constraint coincides with the optimum operating point of our algorithm.

2. Morphable model construction

The process of constructing a morphable model is divided into three stages: 1. establishing a dense correspondence, 2. shape alignment and 3. statistical modelling. In each case, we outline previous methods before describing our approach. We begin by describing how our approach allows us to construct a morphable model as a shape space.

2.1. Morphable Models as Shape Spaces

The 3D morphable face model of Blanz and Vetter [3] captures the class-specific properties of faces by finding a low dimensional parameterisation of 3D face shape and texture. The model is learnt from a sample of high resolution 3D face scans. A common interpretation of the meaning of the *shape* of an object is Kendall's [9] notion that shape is the geometrical information that remains after the effects of location, scale and rotation have been removed. There is a comprehensive toolbox of techniques available for the statistical analysis of shape using data which is provided in terms of coordinates of named point locations or landmarks. Namely, these are Kendall's shape space and the application of linear multivariate statistics in the tangent space at the Procrustes average. This group of techniques has become to be known as geometric morphometrics. A *landmark* is a hypothesis of equivalence under a particular measure of similarity, e.g. anatomy, topology or function. In effect, the implied meaning of a landmark point is, in some sense, the same across the whole population.

However, the statistical analysis of continuous curves or surfaces (which contain only relatively sparse salient points) is not so well developed. For example, only a relatively small proportion of the face surface contains salient points which may be identified with good repeatability across all faces. The remainder of the face comprises large areas of smoothly shaded, textureless surface. It is therefore not obvious how a landmark-based statistical approach can be applied to model the variations in the face surface.

The morphable model of Blanz and Vetter [3] described above is based on transforming a set of face surfaces into a vector space such that any convex combination of members of the training set results in a viable new face. However, their model is not a shape space. They only coarsely remove the effects of rotation, translation and scale *before* the dense correspondence between samples is known. In other words, they ultimately treat every vertex in the model as a landmark but do not remove pose effects with respect to these landmarks.

We propose an alternative approach for constructing a 3D morphable model using Kendall's notion of shape space. Our work closely follows the semilandmark approach of Bookstein [6]. The key idea is to compute correspondences for 'deficient' regions (i.e. those lacking landmark points) using the part of the data that is not deficient. This is done in a principled manner by minimising a physically motivated bending energy of the data about its Procrustes average. We use Procrustes analysis to obtain pose free shape vectors. This combination of techniques allows us to construct a dense 3D morphable model as a shape space.

2.2. Finding Dense Correspondences

Blanz and Vetter's [3] approach is to effectively treat every vertex of a face mesh as a distinct landmark point. This is possible because a modified optical flow algorithm is used to find dense correspondences between all samples in the training set. These correspondences are based on matching regions with similar colour and topography to a reference face and subsequently resampling every face in a consistent manner.

The advantage of their approach is that a model may be constructed automatically with little manual intervention. However, the similarity measure used to find corresponding points between faces relies on an *ad hoc* formulation of local surface features, such as 3D position, texture, local curvature and the surface normal. It is unclear which features should be chosen and how their relative importance should be weighted. Moreover, the utility of different features will vary spatially and between samples. For example, when registering a sample with a beard to one without, texture is an unreliable feature to use. The second problem is that large areas of the face contain no salient structures,

either in the texture or shape domain. For example, the forehead and cheeks. In these regions the calculated flow field is noisy and unreliable. Blanz and Vetter [3] overcome this problem by smoothing and interpolating the flow fields. Finally, the choice of reference face will affect the quality of detected correspondences and ultimately the final model.

At the expense of introducing some manual intervention, we suggest an alternative approach which offers potentially more stable performance. Because our method does not require the selection of a reference face, only one possible model can be constructed from a given set of training data.

We commence with a set of face surfaces obtained by a Cyberware 3030PS laser range scanner. These surfaces are parameterised in cylindrical coordinates. This provides a convenient representation of the facial manifold in 2 dimensions, (u, v). A set of sparse 2D landmark points are manually identified on the parameterisation of each face surface. The landmark points are chosen such that they can be reliably located on all training samples. With these sparse, but reliable, correspondences in hand, the mean coordinates of each landmark point are found. The x, y and z coordinates of each vertex can be expressed as a function in (u, v) space, e.g. $x(u, v)$. Similarly for each colour channel in the texture map. We warp the landmark points of each sample to the mean landmarks. We interpolate this warp using a physically motivated bending energy, through the application of a thin-plate spline warp [5]. Finally, we resample the vertex coordinate functions in a consistent manner across all faces. The result is that a point (u, v) corresponds to the same point on each face in the training set, i.e. we have established a dense correspondence. This process is demonstrated in Figure 1.

In drawing a comparison between our approach and Blanz and Vetter's [3] optical flow algorithm: both methods begin with a sparse set of correspondences (ours manually landmarked, theirs the set of automatically detected correspondences that are considered reliable) and both interpolate the remainder. Construction of the morphable model is done offline prior to its use in an application such as face shape recovery. Hence the manual processing required by our methods is an acceptable burden if it results in more accurate correspondences.

2.3. Shape Alignment

In Blanz and Vetter's [3] morphable model, shape alignment is treated as a preprocessing step. The raw face meshes are marked with a small number of feature points and a 3D-3D transform is used to align each face to a reference face. In other words, when this alignment takes place, the dense correspondence between faces is unknown and the scale, translation and rotation necessary to register each face to the reference is only a very coarse approximation. Further, as with computing the dense correspondences, the

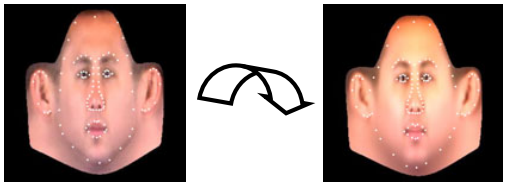

Figure 1. Shows the correspondence of scans based on the principle of thin-plate splines. Using two sets of 2D points (white dots), a novel scan (left) is warped to the mean scan (right) using the thin-plate spline function $-U(r) = -r^2 \log r^2$.

choice of reference face will affect the final model. We propose instead to use Procrustes analysis as a rigorous means to remove pose effects without having to choose a reference face (the reference face is instead the Procrustes mean which is iteratively updated).

With our sample of faces in dense correspondence (forming a vector space) we can proceed with shape alignment using the standard tools of statistical shape analysis. The idea here is to remove any effects of scale, rotation and translation to obtain a pure shape model that captures only variation in identity. The ith face is represented by the shape vector $\mathbf{x}_i = (x_1, y_1, z_1, \ldots, x_p, y_p, z_p)^T \in \mathbb{R}^{3p}$, that contains the x,y,z coordinates of its p vertices.

Our aim is to transform the shape vectors into a shape space. We do this by aligning the shape vectors to a common coordinate frame using generalised Procrustes analysis. This is an iterative procedure which alternates between aligning all samples to the current estimate of the mean shape and then re-estimating the mean from the aligned vectors. These two steps are iterated until convergence.

Our mean shape estimate (for m face scans) is the Procrustes mean:

$$\mathbf{x}_0 = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i. \tag{1}$$

In order to maintain a constant scale for the model, we fix the length of the mean shape at each iteration:

$$\bar{\mathbf{x}} = \frac{\mathbf{x}_0}{\|\mathbf{x}_0\|}. \tag{2}$$

All samples are aligned to the current estimate of the mean shape using a 3D similarity transform, $T_r(\mathbf{x}_i, \gamma_i) = (x'_1, y'_1, z'_1, \ldots, x'_p, y'_p, z'_p)^T$, where

$$\begin{pmatrix} x'_k \\ y'_k \\ z'_k \end{pmatrix} = s\mathbf{R} \begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} + \mathbf{t}. \tag{3}$$

Here, $\mathbf{R} \in \mathrm{SO}(3)$ is a rotation matrix, $s \in \mathbb{R}$ is a scaling and $\mathbf{t} \in \mathbb{R}^3$ is a translation. The optimal pose parameters $\gamma_i = (\mathbf{R}, s, \mathbf{t})$ which map a sample onto the mean shape

vector are found by using Horn's method [12] to solve:

$$\gamma_i = \arg\min_{\gamma} \|T_r(\mathbf{x}_i, \gamma) - \bar{\mathbf{x}}\|^2. \quad (4)$$

A summary of the steps involved in generalised Procrustes analysis is as follows:

1. Find the Euclidian mean of the face shape vectors (1).
2. Rescale the the mean shape vector to unit length (2).
3. Find the optimal pose parameters, γ_i, to align each shape vector, $\mathbf{x}_i$, to the mean.
4. Set $\mathbf{x}_i = T_r(\mathbf{x}_i, \gamma_i)$.
5. If change in estimate of mean indicates convergence, stop. Otherwise iterate to step 1.

This process converges very rapidly (typically within 3 iterations).

Since scale has been removed from the shape vectors, they all lie on the surface of a curved manifold in shape space (since we set the model scale to 1, the shape vectors will all lie on a unit hypersphere). This invalidates the application of linear statistical analysis using tools such as PCA. A standard technique to overcome this problem is to apply a stereographic projection to the shape vectors in order to transform them to points on the tangent space to the Procrustes average. This is simply a case of rescaling the aligned shape vectors as follows:

$$\mathbf{x}'_i = \frac{1}{\bar{\mathbf{x}} \cdot \mathbf{x}_i} \mathbf{x}_i. \quad (5)$$

It is to these rescaled vectors that we apply further analysis. In practice, this rescaling slightly improves the efficiency of the model (typically reducing the dimensions required to capture 95% variance by one).

In our experimental results, we demonstrate that our shape alignment procedure results in a superior model to that of Blanz and Vetter [3].

2.4. Statistical Modeling

We apply PCA to the set of pose free shape vectors $\mathbf{x}'_i$. This performs a basis transformation to an orthogonal coordinate system spanned by the m eigenvectors P_i. Any face surface $\mathbf{x}$ which has been aligned to the mean and projected to the tangent space may now be represented as a linear combination of the average surface and the model eigenvectors:

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i=1}^{m} b_i P_i, \quad (6)$$

where $\mathbf{b} = (b_1, \ldots, b_m)^T$ is a vector of parameters. We stack the eigenvectors to form a matrix $\mathbf{P}$, such that we may write: $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$. The PCA eigenvalues λ_i provide a measure of how much of the variance of the training data is captured by each eigenvector. We may choose to retain $n < m$ model dimensions, such that a certain percentage of the cumulative variance is captured. We discuss the effect of the number of model dimensions in our experimental results.

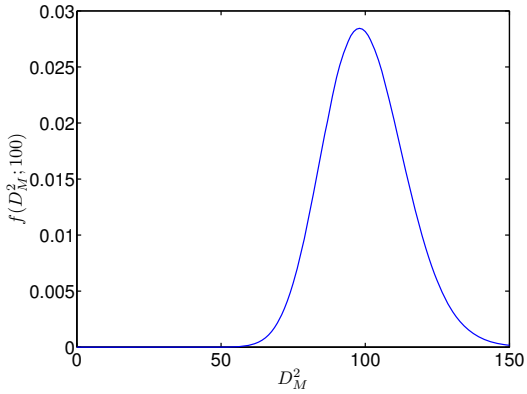

Figure 2. Probability density function for parameter vector lengths (measured in terms of the squared Mahalanobis distance) for a 100 parameter model.

Our statistical model also provides an estimate of the probability distribution of the shape vectors. We begin by defining the distance of a sample from the mean in terms of the square of the Mahalanobis distance:

$$D_M^2(\mathbf{b}) = \sum_{i=1}^{n} \left(\frac{b_i}{\sqrt{\lambda_i}} \right)^2. \quad (7)$$

Since we assume each parameter follows a Gaussian distribution, the parenthesised terms are independent, normally distributed random variables with zero mean and unit variance. This is exactly the definition of the chi-square distribution.

In other words, the lengths of the parameter vectors (as measured by the square of the Mahalanobis distance from the mean) follow a chi-square distribution with n degrees of freedom, i.e. $D_M^2 \sim \chi_n^2$. Such a distribution has a mean value of n and variance $2n$. The probability density function over the parameter vector length x for an n parameter model is:

$$f(x; n) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}. \quad (8)$$

The interesting observation here is that the expected length of the parameter vector of a n-dimensional model is n. The likelihood of a sample having a length close to zero (i.e. approximately the mean sample) is extremely small. For example, a model with 100 dimensions would have a mean vector length of 100 and over 99% of parameter vectors would have lengths greater than 70. The probability of a vector length less than 50 is negligibly small. In Figure 2 we show the probability density function for parameter vector lengths for a model with 100 parameters. Note that as n

increases, the shape of the chi-square distribution tends to a Gaussian.

This prior on the parameter vector lengths is starkly different to Blanz and Vetter's [4, 1] assertion that the parameter vector lengths are normally distributed with zero mean. Their assumption is that the most probable parameter vector is that with length zero and that the probability decreases as the parameter vector length increases. Empirical experiments on out of sample data have confirmed that our model fits real data well. The predicted mean parameter vector length matches observed values almost exactly. However, the variance tends to be underestimated for models with more than 10 parameters. This is likely to be a result of trying to estimate model eigenvectors in a very high dimensional space from a relatively very sparse sample.

We use this prior distribution on the parameter vector lengths to motivate imposing a hard constraint on the length during the fitting process. In effect, we hypothesise that all samples lie approximately on the shell of a hyperellipsoid in parameter space.

3. 3D Face shape from sparse feature points

In this section we lay the framework for a nonlinear, iterative fitting algorithm to estimate a high resolution 3D face surface given the positions of k 2D annotations on the subjects face ($k << p$). In contrast to the analysis-by-synthesis approach of Blanz and Vetter [4], we do not use face appearance or a model of texture variation to reconstruct the 3D surface of a face. This makes the shape recovery process approximately two orders of magnitude faster.

3.1. The need for regularisation

If the effects of pose are discounted (i.e. if it is assumed that the rotation required to align the model with a set of feature points in the image plane is known), then estimating the morphable model shape parameters that give rise to a particular configuration of landmark points can be solved using linear least squares. Such an approach is impractical as it leads to gross overfitting of the data. Clearly, there is a trade off between the quality of fit to the observed data and prior probability as measured by the model. Using our model of prior probability described in Section 2.4 as a regularisation constraint results in a nonlinear optimisation problem.

Blanz *et al*. [1] proposed a linear, single step solution to this problem based on their assumption of a Gaussian distribution over the parameter vector lengths. However, the 3D rotations on the position of 2D feature points in the image plane also introduce nonlinearities. To sidestep this problem, Blanz *et al*. [1] use small angle approximations which are only valid for very small changes in pose. Due to this approximation, the estimated pose may be inaccurate. To overcome this problem, they repeat the process using the result of the first pass as an initialisation, in effect turning their one shot method into an iterative one.

We choose instead to separate the influence of pose and shape parameters on the optimisation and solve the problem using nonlinear, iterative optimisation. As stated above, the chi-squared distribution of parameter vector lengths implies that the parameter vectors lie on the surface of hyperellipsoid in parameter space. We use this observation to motivate imposing a hard constraint on the length of the estimated parameter vectors. If the number of parameters in the model is n, we enforce the constraint $D_M^2 < n$. In practice, because of the tendency to overfit, the result is that $D_M^2 = n$. To impose this constraint, at each iteration of the minimization we scale the estimated parameter vector such that its length in terms of squared Mahalanobis distance from the mean is n:

$$\mathbf{b} = \frac{\sqrt{n}}{D_M(\mathbf{b})}\mathbf{b}. \tag{9}$$

3.2. Fitting to sparse data

Given a set of k annotations marked on the input face ($L_{2d} \in \mathbb{R}^{2k}$), we can determine the vertices corresponding to those salient points on the mean face as shown in Figure 6. Once we have the k indexed vertices we can extract their corresponding 2D projections using:

$$\hat{L_{2d}} = P_k T_r^{-1} \left(\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}, \gamma\right). \tag{10}$$

where, the term $\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$ provides the estimated shape, γ provides the 3D pose with respect to the mean shape and P_k is the projection of the k indexed vertices under an orthographic projection.

Using (10) our aim is to minimize the error between L_{2d} and $\hat{L_{2d}}$ subject to the constraint on the parameter vector length. The quality of fit to the data is measured by:

$$E\left(\mathbf{b}, \gamma\right) = \left\| L_{2d} - \hat{L_{2d}} \right\|. \tag{11}$$

The optimal parameters are therefore given by:

$$(\mathbf{b}^*, \gamma^*) = \underset{D_M^2(\mathbf{b}) \leq D_{\max}^2, \gamma}{\arg\min} E\left(\mathbf{b}, \gamma\right), \tag{12}$$

where $D_{\max}^2$ is the maximum allowable parameter vector length. We examine the effect of varying this value in our experimental results. We solve this minimization using Levenberg-Marquardt optimization [14].

4. Experimental Results

In this section we present the results of our experimental evaluation. We begin by evaluating our approach to constructing a morphable model as a shape space and compare it to that of Blanz and Vetter [3]. We then demonstrate an example application of our model by using it to reconstruct high resolution 3D face surfaces from sparse 2D landmarks.

4.1. Morphable Model Construction

In this section we compare our strategy for morphable model construction described in Section 2 with the state of the art. In order to ensure a fair comparison, we use exactly the same data as was used in [3]. This comprises 100 3D face scans which have been set into correspondence using a modified optical flow algorithm. Blanz and Vetter's [3] model is obtained by applying PCA directly to these shape vectors. Our model uses generalised Procrustes analysis to obtain a stable estimate of the mean face and shape vectors that are free of scale, translations and rotations. We rescale each shape vector according to the tangent space projection given in (5).

We divide the data into a training set of 75 scans and a test set of 25 scans. In Figure 3 we plot the percentage cumulative variance captured as a function of the number of model dimensions. It is clear that the model of Blanz and Vetter captures a larger proportion of cumulative variance for a given number of model dimensions. This is often seen as evidence that a model is more efficient and hence superior. Our results show that in fact, the variance captured is spurious and is related to variations in pose rather than identity. Although our model apparently captures less cumulative variance, the variance it does capture is of more use for representing out of sample data. The modes of variation of the two models are visually distinguishable. In Figure 5 we show the effect of adding and subtracting the first two modes of variation to the mean face for both models. As can be seen, the characteristics captured by the modes are subtly different.

We compare generalisation ability by measuring the accuracy with which the two models can reconstruct out of sample face meshes. For a novel shape vector, $\mathbf{s}$, we find the optimal parameter vector $\mathbf{b}^* = \mathbf{P}^T(\mathbf{s} - \bar{\mathbf{x}})$. The reconstructed surface is given by $\mathbf{P}\mathbf{b}^* + \bar{\mathbf{x}}$. The average reconstruction error over the whole test set is given by:

$$E_{3d} = \frac{1}{tp} \sum_{i=1}^{t} \sum_{j=1}^{p} \left\| \begin{pmatrix} x_{i,j} \\ y_{i,j} \\ z_{i,j} \end{pmatrix} - \begin{pmatrix} x^r_{i,j} \\ y^r_{i,j} \\ z^r_{i,j} \end{pmatrix} \right\|, \qquad (13)$$

where t is the number of samples in the test set, $x_{i,j}$ is the x component of the jth vertex in the ith test sample, with $x^r_{i,j}$ being the corresponding reconstructed value after projection onto the model.

Figure 4 shows the absolute reconstruction errors (in mm) for both models. We vary the number of model dimensions retained and observe its effect on the generalisation error. In Figure 4a, we use the same number of dimensions for both models. In Figure 4b, we retain as many dimensions as are required to capture a fixed proportion of the cumulative variance. By either measure, our model generalises to unseen data more accurately, even when less cumulative variance has been retained. This makes our model both more efficient (fewer model dimensions required to obtain a given generalisation error) and more accurate (even with all dimensions retained, our model provides higher accuracy).

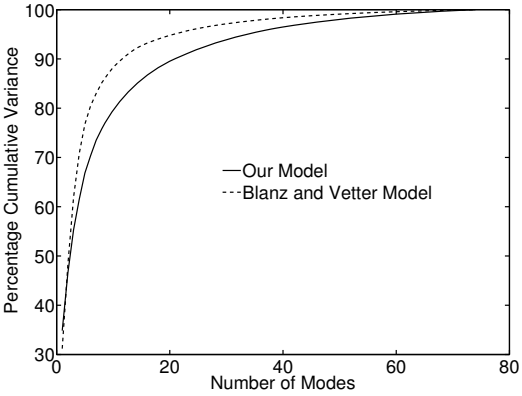

Figure 3. Cumulative Variance Curve

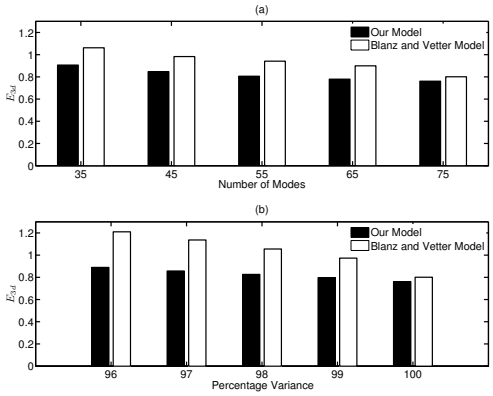

Figure 4. (a) E_{3d} (in mm) vs Number of Modes, (b) E_{3d} (in mm) vs Percentage Variance

4.2. Estimation of 3D faces from Sparse 2D Features

In this section we show the result of using the technique described in Section 3 to estimate high resolution 3D face surfaces from sparse 2D feature points. For these experiments we use a morphable model constructed from 100 face scans using the techniques described in Section 2. The scans are preprocessed to remove the hair and neck regions and are set into correspondence using a thin-plate spines based warping (Figure 1). Each face is represented by $p = 50468$ vertices. To test the effects of the resolution of the model, we also obtained lower resolution model composed of face meshes containing $p = 3147$ vertices. In both cases, we retain the 99 most significant modes.

Our reconstruction algorithm requires the user to annotate the positions of the landmarks on the input image (Figure 6). Note that in practice this could be done using a 2D

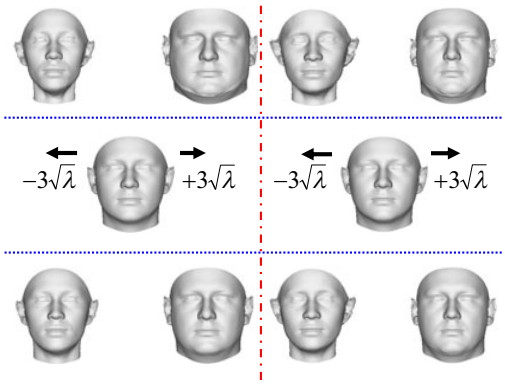

Figure 5. Shows the deviations in the two most significant parametric modes. The deviations are shown for our model (left hand side) and the Blanz and Vetter Model (right hand side). The deviations from the mean head (middle row) of the most significant (top row) and second most significant (bottom row) parameters shows the subtle differences in the morphing ability of the two models

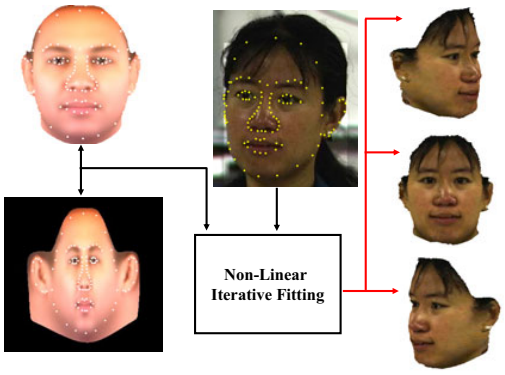

Figure 6. Shows the framework for the proposed system. The mean mesh (top left) and 2D input image with k indexed annotations (center) are the inputs to the system. The mean texture map (bottom left) is used to determine the vertices (white dots) corresponding to the k indexed points (yellow dots) in the input image. The system outputs the estimated shape and pose (right column). A simple blending function is used to estimate the occluded texture.

feature detector. We use $k = 104$ annotated points. The computation time (on a 1.78 GHz AMD Athlon processor) is approximately 550ms and 10000ms for the $p = 3147$ vertices and $p = 50468$ vertices face models respectively.

As discussed in Section 3, we impose a hard constraint, D^2_{max}, on the length of the estimated parameter vectors during the fitting process. Without this constraint, the tendency of the algorithm is to overfit the sparse data resulting in a very poor global shape estimate. We examine the effect of varying the value of this constraint on the 3D error of the reconstructed surface. In effect, this parameter controls the trade off between fitting quality and shape plausibility. Our statistical prior predicts that the average length of the parameter vectors for an n parameter model is n. For values of D^2_{max} significantly greater than n, the system clearly overfits and the faces are heavily distorted. For values of D^2_{max} close to zero, the shape estimate is always similar to the average face and the fitting quality is low. We show that the optimum operating point of our algorithm coincides with the prediction of our statistical prior on real data, i.e. optimal performance occurs when $D^2_{\text{max}} \approx n$. An example of underfitting, overfitting and optimal performance is shown in Figure 7.

To provide quantitative confirmation of this assertion, we applied our shape estimation algorithm to 50 ground truth samples. These were disjoint from the samples used to construct the morphable model. For each scan we obtained the k indexed vertices corresponding to the salient annotations. These were projected to 2D. The faces were in approximately frontal pose (variations of up to 12° from frontal occurred in practice). For different values of D^2_{max}, we fitted our morphable model to this sparse data and computed the per vertex average 3D reconstruction error over all 50 samples. This experiment was carried out for both the high and low resolution models.

Figure 8 shows the result of this experiment. For both models, the minimum reconstruction error occurs approximately when $D^2_{\text{max}} = n$. The error of the reconstructed faces is approximately $3.6mm$. Although this is only a slight quantitative improvement over using the average face, the perceptual improvement is much greater, as evidenced by Figure 7.

Our findings show that our analytical prediction of the average parameter vector lengths coincide with the optimum operating point of our reconstruction algorithm and hence provide a non-heuristic constraint for optimal fitting.

5. Conclusions

3D morphable models are an important tool for face shape estimation, recognition and reanimation. In this paper we have revisited the process of constructing a morphable model from training data. We have described alternative approaches to finding dense correspondences, removing pose from the model (which results in a morphable model which is a shape space) and constructing a probabilistic prior over the distribution of parameter vector lengths. We applied the model and hard constraint implied by the prior to the problem of estimating high resolution 3D face surfaces from sparse 2D feature points. In future work, we intend to develop more sophisticated fitting algorithms which employ the probabilistic prior as a soft constraint. We will also investigate using the estimated shape parameter vectors for the purposes of face recognition.

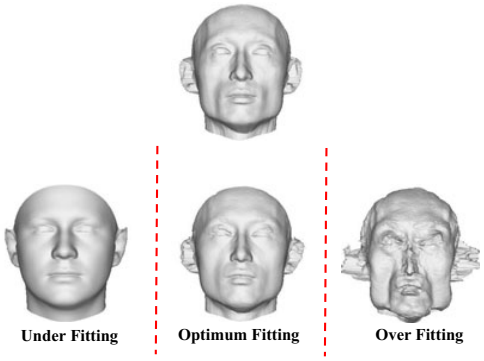

Figure 7. Shows the tradeoff between the fitting quality and shape plausibility. Given a novel face (top row), the bottom row shows the three cases of fitting: (a) Under Fitting (b) Optimum Fitting and (c) Over Fitting

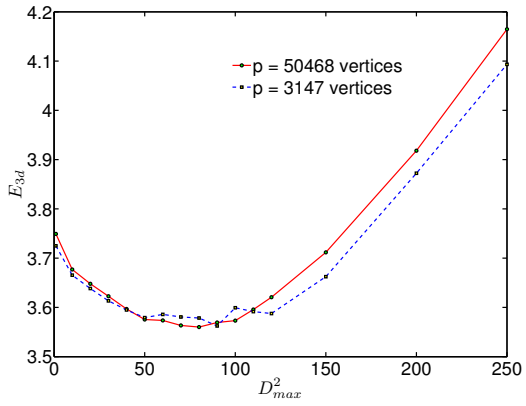

Figure 8. E_{3d} (im mm) vs $D^2_{\max}$

References

[1] V. Blanz, A.Mehl, T. Vetter, and H.-P. Seidel. A statistical method for robust 3D surface reconstruction from sparse data. In *Proc. 2nd Int. Symposium on 3D Data Processing, Visualization and Transmission*, pages 293–300, 2004.

[2] V. Blanz, P. Grother, J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non–frontal images. In *Proc. CVPR*, volume 2, pages 454–461, 2005.

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, pages 187–194, 1999.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[5] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(6):567–585, 1989.

[6] F. L. Bookstein. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal.*, 1(3):225–243, 1997.

[7] A. Bronstein, M. Bronstein, and R. Kimmel. Three–dimensional face recognition. *Int. J. Comput. Vis.*, 64(1):5–30, 2005.

[8] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3D structure recovery of faces. In *Proc. ECCV*, volume 2, pages 99–113, 2004.

[9] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons Ltd, 1998.

[10] A. Georghiades. Recovering 3–D shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering*, pages 230–240, 2003.

[11] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, 2001.

[12] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America. A*, 4(4):629–642, 1987.

[13] R. Knothe, S. Romdhani, and T. Vetter. Combining PCA and LFA for surface reconstruction from a sparse set of control points. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pages 637–644, 2006.

[14] M. I. A. Lourakis. levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++. `http://www.ics.forth.gr/~lourakis/levmar/`.

[15] B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju. Model–based 3D face capture with shape–from–silhouettes. In *Proc. IEEE Work. Analysis and Modeling of Faces and Gestures*, pages 20–27, 2003.

[16] S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman. Face recognition using 3–D models: Pose and illumination. *Proc. of the IEEE*, 94(11):1977–1999, 2006.

[17] S. Romdhani and T. Vetter. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Proc. CVPR*, volume 2, pages 986–993, 2005.

[18] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):351–363, 2006.

[19] W. Y. Zhao and R. Chellappa. Illumination–insensitive face recognition using symmetric SFS. In *Proc. CVPR*, pages 286–293, 2000.