Article:

Patel, Ankur and Smith, William Alfred Peter orcid.org/0000-0002-6047-0413 (2016) Manifold-based Constraints for Operations in Face Space. Pattern recognition. pp. 206-217. ISSN 0031-3203

https://doi.org/10.1016/j.patcog.2015.10.003

Manifold-based Constraints for Operations in Face Space

Ankur Patel

Conversant Inc.

William A. P. Smith

University of York, York, UK

Abstract

In this paper, we constrain faces to points on a manifold within the parameter space of a linear statistical model. The manifold is the subspace of faces which have maximally likely distinctiveness and different points correspond to unique identities. We provide a detailed empirical validation for the chosen manifold. We show how the Log and Exponential maps for a hyperspherical manifold can be used to replace linear operations such as warping and averaging with operations on this manifold. Finally, we use the manifold to develop a new method for fitting a statistical face shape model to data, which is both robust (avoids overfitting) and overcomes model dominance (is not susceptible to local minima close to the mean face). We provide experimental results for fitting a dense 3D morphable face model to data using two different objective functions (one underconstrained and one with many local minima). Our method outperforms generic nonlinear optimisers based on the BFGS Quasi-Newton method and the Levenberg-Marquardt algorithm when fitting using the Basel Face Model.

Keywords: Optimisation on Manifolds, Face Space, 3D Morphable Models, Constrained Optimisation, Statistical Modelling

---

Email addresses: `ankurrp@gmail.com` (Ankur Patel), `william.smith@york.ac.uk` (William A. P. Smith)

1. Introduction

Modelling "face space" (the manifold on which valid faces lie) is a longstanding goal in statistical shape analysis and computer vision and has been performed in various domains including 2D [1] and 3D [2] shape, appearance [3] and texture [4]. These approaches can be viewed as manifold learning where the faces are assumed to lie on an unknown manifold, the structure of which is learnt from data. Most commonly, the manifold is assumed to be a hyperplane (linear subspace) and the principal axes of the plane are estimated from training data using Principal Components Analysis (PCA). Applying these models to face analysis tasks requires a means to fit the model to observed data. Often this fitting process is underconstrained, prone to converge on local minima and computationally expensive. For these reasons, there is strong motivation for developing more constrained face space models in order to reduce the search space of the fitting process.

An alternative to manifold learning is to assume that the structure of the face space manifold is known. For example, the Grassmannian manifold of subspaces of a vector space has been used in face recognition [5] and the Kendall manifold of shapes has been used to model face shape [6].

The model we propose in this paper can be viewed as a hybrid of these two approaches in the sense that we assume the shape of the manifold is known (hyper-ellipsoidal) but we use manifold learning (PCA) to discover its principal axes from data. The motivation for this choice of model is as follows.

Psychological results [7, 8] have shown that the parameter space of a PCA-based model has an interesting perceptually motivated interpretation: *identity* relates to direction in parameter space while *distinctiveness* is related to vector length (or equivalently distance from the mean). The reason for this is that increasing the length of a parameter vector simply exaggerates its differences from the average linearly, in other words its *features*, whereas rotating a parameter vector changes the *mix* of features present in the face. This is the justification for using angular difference in face space as a measure of dissimilarity for face recognition [4].

This decomposition also allows a useful probabilistic interpretation. Under the assumption that the original data forms a Gaussian cloud in a high dimensional space, each model parameter is independent and distributed according to a Gaussian distribution. This means that all faces lie on or near the surface of a hyperellipsoid in parameter space, with the probability density over the parameter vector lengths following a chi-square distribution. In other words, distinctiveness is subject to a statistical prior with the distinctiveness of most samples clustered around the expected length.

In this paper, we use these observations to motivate a representation for faces which decomposes face appearance into identity and distinctiveness subspaces. We focus on statistical models of 3D face shape. However, any class of objects amenable to linear statistical modelling using PCA could make the same identity/distinctiveness decomposition. We use ideas from differential geometry to develop tools which operate in the identity subspace, i.e. which retain constant distinctiveness. We provide empirical justification for constraining samples to have fixed distinctiveness, determined by the expected vector length.

We propose a new algorithm for fitting a statistical face model to data. Many such methods have been proposed previously, the details being dependent on the precise nature of the model and data. This inevitably involves a nonlinear optimisation over the model parameters. Our approach is more general and can be applied to any objective function. It operates via gradient descent on the manifold of equal distinctiveness. In other words, we solve for identity and assume distinctiveness takes its expected value. We show how the method naturally lends itself to a coarse-to-fine optimisation strategy and how the result avoids local minima or overfitting without having to select a regularisation weight parameter. We show that this offers improved performance over two generic nonlinear optimisation algorithms.

1.1. Related Work

Perhaps the best known statistical face model is the Active Appearance Model (AAM) [3] which combines a linear model of 2D shape and 2D appearance. Rather than model appearance, the 3D Morphable Model of Blanz and Vetter [4] models the shape and texture which give rise to appearance via a model of image formation. Xiao et al. [9] have used a 3D model in conjunction with a 2D appearance model to enforce geometric constraints on the 2D shape generated.

Construction or training of a statistical face model involves a number of steps: 1. data collection, 2. registration (e.g. transforming the face data to a vector space) and 3. statistical analysis. When represented in a vector space, face-like samples can be synthesised by taking convex combinations of training faces. However, it is the statistical analysis which allows us to study how the face samples distribute themselves in high dimensional space and which regions of this space correspond to plausible faces, i.e. face space.

Although statistical face models have useful applications when used in a purely generative manner (e.g. for the synthesis of faces), the most compelling applications necessitate face analysis through fitting the model to observed data. This data may take many forms, such as the appearance of a face in one [4, 3, 9] or more [10, 11] images, a noisy and incomplete 3D scan [12] or the location of a sparse set of feature points in an image [2].

When the objective function is underconstrained or ill-posed, the classical approach is to use Tikhonov regularisation (for a linear objective) or more generally to augment the objective function with a regularising term using a Lagrange multiplier. Typically, the regularisation term encourages smaller norms or equivalently, solutions closer to the mean face. With a suitable choice of the regularisation weight, this prevents overfitting and ensures that the resulting face is plausible. However, the optimal choice of regularisation weight may be different for different data samples. By choosing a conservative value, fitting results are likely to be too close to the mean face to capture features of the input face.

Much prior work uses such regularised optimisation approaches for face model fitting. For example linear regression [3], the inverse compositional algorithm [13], global optimisation [4], hybrid objective functions to encourage convexity [14] and alternating least squares for solving a multilinear system [15, 16]. All of these approaches trade off satisfaction of a model-based prior against quality of fit. To ensure robust performance, these approaches must favour the prior, resulting in model dominance.

Recently, Brunton et al. [17] proposed a method to fit a statistical shape model to 3D data. They used a hard hyper box constraint, whereby each shape parameter was constrained to lie within $\pm k$ standard deviations of the mean. In other words, they assumed a uniform distribution over the hyper box as their prior. This has the advantage of being expressed as a linear inequality constraint on the parameters, enabling it to be incorporated into standard optimisation methods. Their hyper box is more conservative than the hyper-ellipsoid constraint that we propose here, with the two only intersecting at the corners of the hyper box. This is done so as to prevent extreme values of a single parameter being allowed by the constraint. We have not found this to be a problem in our experimental results and our manifold is motivated directly by the properties of assumed distribution over the parameters. Moreover, by assuming a uniform prior they do not discourage solutions close to the mean when the objective is over constrained.

There has been a recent interest in shape modelling on manifolds. Berkels et al. [18] show how to perform discrete geodesic regression on shape manifolds. This allows them to perform nonlinear regression in shape space according to a specified discrete path energy. For the specific case of the space of thin shells (including faces), Heeren et al. [19] provide a computational framework for calculating geodesics, allowing for plausible interpolations, averaging, and even shape extrapolation applications. In an altogether difference approach, Boscaini et al. [20] formulate shape interpolation and averaging in the space of Laplacians, from which shapes are subsequently reconstructed. Shapira and Ben-Chen [21] shows how to align two face spaces (each corresponding to a different identity) by a non-rigid ICP between the corresponding manifold samples. This allows for shape analogies to be computed, providing a kind of expression transfer.

In this paper, we propose to solve the model fitting problem within the subspace of maximally likely faces. This requires the solution of an optimisation problem on a manifold. This problem has been considered previously in the medical imaging [22], signal processing [23], computer vision [24], robotics [25] and projective geometry [26] communities. Generic methods for optimisation on arbitrary manifolds have also been proposed [27]. In particular, the recently released Manopt toolbox [28] allows local optimisation on a number of manifolds through the expression of an objective and its gradient in the Euclidean embedding space. We focus on the case of a hyperspherical manifold and develop a hyphersphericalgradient descent algorithm. In contrast to Manopt, our method operates in a coarse-to-fine manner in order to reduce susceptibility to local minima and exploits the closed nature of the manifold to reduce line searches to interval searches. We extend our previous presentation of this work [29] by demonstrating results on expression interpolation (Section 3.1) and underconstrained optimisation (Section 5.2), more thorough empirical evaluation of the manifold assumption and describing the theoretical ideas more thoroughly.

1.2. Outline

In Section 2 we begin by describing our statistical model and manifold. We first introduce tools from differential geometry which are necessary for developing our methodology and then provide empirical validation to justify our choice of manifold. In Section 3 we describe how warps and averages between two or more faces can be constrained to the manifold and compare the result with linear methods. In Section 4 we present our principal contribution: a method for fitting the model to data within the subspace defined by the manifold. In Section 5 we provide results for two contrasting objective functions (one overconstrained, the other underconstrained) and compare with generic nonlinear optimisers using a regularised objective. Finally, in Section 6 we provide conclusions and directions for future work.

2. Statistical Modelling

Consider a sample of 3-dimensional face meshes which are in dense correspondence (i.e. the same point on every face has the same vertex index). The ith shape is represented by a vector of p vertices

$$\mathbf{s}_i = (x_1, y_1, z_1, \ldots, x_p, y_p, z_p) \in \mathbb{R}^{3p}.$$

Given m such shape vectors, we use principal components analysis to obtain an orthogonal coordinate system spanned by the m eigenvectors, where $\mathbf{p}_i$ is the ith eigenvector. Any shape vector $\mathbf{s}$ may now be represented as a linear combination of the average shape and the model eigenvectors:

$$\mathbf{s} = \bar{\mathbf{s}} + \sum_{i=1}^{m} c_i \mathbf{p}_i, \tag{1}$$

where $\mathbf{c} = [c_1 \ \ldots \ c_m]^T$ is a vector of parameters. We stack the eigenvectors to form a matrix $\mathbf{P}$, such that we may write: $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Pc}$. The PCA eigenvalues, denoted λ_i for the ith eigenvalue, provide a measure of how much of the variance in 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. Psychological results show us that the dimensionality of face space is relatively small (Meytlis and Sirovich [30] suggest 100 dimensions is sufficient, even using a crude eigenface model). We discuss the effect of the number of model dimensions and empirically evaluate their stability in Section 2.4.

Our interest in this paper is to explore how shape samples drawn from a population distribute themselves in parameter space and how we can use this knowledge to constrain operations. We define the vector $\hat{\mathbf{c}} = [c_1/\sqrt{\lambda_1} \ \ldots \ c_n/\sqrt{\lambda_n}]^T$ as the variance-normalised parameter vector. This vector is distributed according to a multivariate Gaussian with zero mean and unit variance, i.e. $\hat{\mathbf{c}} \sim \mathcal{N}(0, \mathbf{I}_n)$. This is the prior constraint typically used in the model fitting process to ensure that solutions remain plausible. It is maximised by a zero vector, which corresponds to the mean sample.

However, another interpretation based on the parameter vector length is possible. The squared norm of $\hat{\mathbf{c}}$ corresponds to the square of the Mahalanobis distance of $\mathbf{c}$ from the mean:

$$\|\hat{\mathbf{c}}\|^2 = D_M^2(\mathbf{c}) = \sum_{i=1}^{n} \left(\frac{c_i}{\sqrt{\lambda_i}} \right)^2 . \tag{2}$$

Since we assume each parameter follows a Gaussian distribution, the parenthesised terms are independent, normally distributed random variables with zero mean and unit variance. The sum of the square of such variables follows a chi-square distribution with n degrees of freedom, i.e. $\|\hat{\mathbf{c}}\|^2 \sim \chi_n^2$. This distribution has expected value n and variance $2n$. Hence, the standard deviation grows as the square root of the mean and the vector lengths become relatively more tightly concentrated about the mean length as the number of dimensions grows.

These two apparently contradictory distributions suggest that the mean face is the most probable sample but has a highly improbable vector length (this has been reported in the psychology literature as *The Face-Space Typicality Paradox* [31]). For example, a model with 100 dimensions would have an expected vector length of 100 and over 99% of parameter vectors would have lengths between 70 and 130. The probability of a vector length less than 50 is negligibly small.

A note of caution is required to accompany this analysis. Under the assumption that each parameter vector follows a Gaussian distribution, the chi-square analysis holds. However, since PCA eigenvectors capture the maximum possible variance, the eigenvalues decay rapidly with increasing dimension number. The effect of this is to amplify less significant parameters, i.e. the denominator in Equation 2 becomes small. The reason that this causes a problem is that eigenvectors associated with smaller eigenvalues are more susceptible to the influence of noise, particularly when the training set size is small. The empirical analysis in Section 2.4 confirms this prediction.

2.1. Identity as Direction

Our argument is that valid members of the class will occupy a subspace of parameter space. These points will lie close to the surface of a hyperellipsoid,

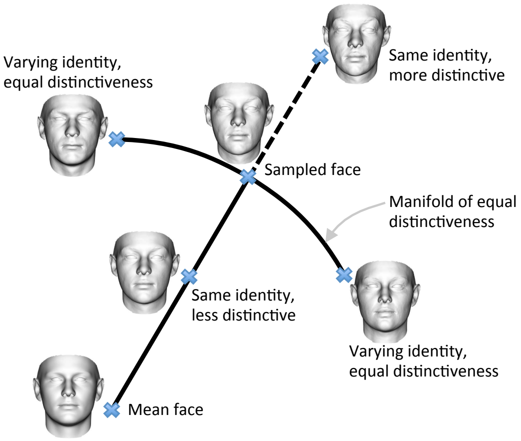

Figure 1: The manifold of equally distinctive faces. A face is randomly sampled from the manifold. Linearly scaling the parameter vector varies distinctiveness while keeping identity fixed. Moving the sample over the manifold varies identity while keeping distinctiveness fixed.

the diameters of which are determined by the eigenvalues of the data.

To negate the need for regularisation, we choose to force all samples to lie on the surface of the hyperellipse, i.e. we fix distinctiveness (vector length) to its expected value as a hard constraint. With distinctiveness fixed, points on the manifold correspond to unique identities. Using this representation, face processing and analysis are transformed to operations on a manifold. This manifold is visualised in Figure 1.

The analysis of data on a hyperellipsoidal manifold is extremely complex. Therefore, without loss of generality, we transform the manifold to a hypersphere by scaling each dimension by its corresponding standard deviation. For the remainder of this paper, we therefore represent parameter vectors with squared Mahalanobis length n as unit vectors in $\mathbb{R}^n$: $\mathbf{x} = \frac{1}{\sqrt{n}} \left[\frac{c_1}{\sqrt{\lambda_1}} \ \dots \ \frac{c_n}{\sqrt{\lambda_n}} \right]^T$, where $\|\mathbf{x}\| = 1$.

2.2. Log and Exponential Maps

Linear operations in Euclidean space such as averaging, warping and computing partial derivatives must be reformulated for data which lies on a curved

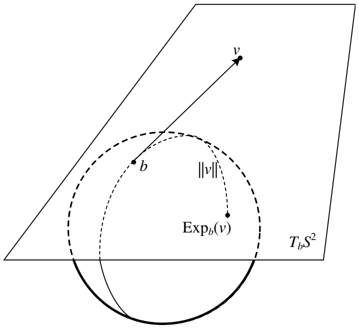

Figure 2: Exponential map for the S^2 manifold.

manifold. This is conveniently done in tangent space, where geodesic curves through the point of tangency correspond to straight lines. Transforming points from the manifold to the tangent space and back again is done using operations from differential geometry, namely the log and exponential map.

A unit vector in n-dimensional space $\mathbf{x} \in \mathbb{R}^n$, may be considered as a point lying on the hyperspherical manifold $x \in S^{n-1}$. The two are related by $\mathbf{x} = \Phi(x)$ where $\Phi : S^{n-1} \mapsto \mathbb{R}^n$ is an embedding. If $v \in T_b S^{n-1}$ is a vector in the tangent space to S^{n-1} at a base point $b \in S^{n-1}$, the exponential map, denoted Exp_b of v is the point on S^{n-1} along the geodesic in the direction of v at distance $\|v\|$ from b. Figure 2 provides a visual illustration of the operation for the S^2 manifold. The inverse of the exponential map is the log map, denoted Log_b.

The geodesic distance (i.e. angular difference) between two points $x_1, x_2 \in S^{n-1}$ on the unit hypersphere can be expressed in terms of the log map, i.e. $d(x_1, x_2) = \|\mathrm{Log}_{x_1}(x_2)\| = \arccos\left(\Phi(x_1) \cdot \Phi(x_2)\right)$. In Section 2.3, we derive a simple and efficient means to compute the log and exponential maps for the unit hypersphere. In the remaining sections, we use the log and exponential maps to perform useful operations on the manifold.

2.3. Log and Exponential Maps for the Hypersphere

In practice, we represent points on both the hyperspherical manifold and the tangent space as vectors embedded in $\mathbb{R}^n$. The log map [32] of x at base point b is therefore computed with respect to unit vectors in $\mathbb{R}^n$: $\mathbf{b} = \Phi(b)$ and $\mathbf{x} = \Phi(x)$:

$$\Phi_T\left(\mathrm{Log}_b(x)\right) = \frac{\theta(\mathbf{x} - \pi_{\mathbf{b}}(\mathbf{x}))}{\|\mathbf{x} - \pi_{\mathbf{b}}(\mathbf{x})\|}, \tag{3}$$

where $\pi_{\mathbf{b}}(\mathbf{x}) = (\mathbf{b} \cdot \mathbf{x})\mathbf{b}$ is the projection of $\mathbf{x}$ onto $\mathbf{b}$ and $\theta = \arccos(\mathbf{b} \cdot \mathbf{x})$. The result is a vector in the tangent space $T_b S^{n-1}$ embedded in $\mathbb{R}^n$ according to an arbitrary embedding $\Phi_T : T_b S^{n-1} \mapsto \mathbb{R}^n$.

Similarly, the exponential map of a tangent vector at b embedded in $\mathbb{R}^n$, $\mathbf{v} = \Phi_T(v)$, is given by:

$$\Phi(\mathrm{Exp}_b(v)) = \cos(\theta)\mathbf{b} + \frac{\sin\theta}{\theta}\mathbf{v}, \tag{4}$$

where here, $\theta = \|\mathbf{v}\|$.

2.4. Empirical Evaluation: χ^2 Prediction

Before we consider applications of processing data on the manifold described above, we provide some empirical assessment of how well real world data adheres to the theoretical prediction made in Section 2.1. In order for all plausible data samples to lie on or near the predicted manifold, the assumption of parameter vector lengths following the chi-squared distribution must hold. In turn, the distribution of faces along each eigenvector must follow a Gaussian distribution. In practice, these eigenvectors are estimated from a sparse sample of a high dimensional space. In the case of a dense 3D face shape model, observations typically consist of tens of thousands of vertices while the training set typically comprises only hundreds of samples.

Clearly, the validity of the estimated manifold depends on the quality of the estimated eigenvectors and therefore the size and diversity of the training set. Within-sample data (i.e. that used to train the model) adheres almost exactly to the manifold assumption. Hence, we empirically evaluate whether out-of-sample data follows the theoretical prediction. For our empirical test we use the Basel Face Model (BFM) [33]. The BFM is a 3D morphable model constructed from 200 faces. An additional 10 unseen (out-of-sample) faces are provided which are in correspondence with the model.

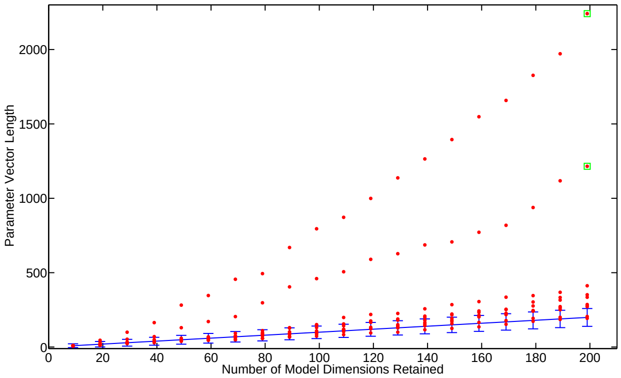

Figure 3: Parameter vector length vs Number of model dimensions retained. The red points represent the vector length for each out-of-sample face. The error bars (blue) represent three standard deviation variation around the mean of the chi-square distribution. The green squares denote the two faces which grossly overfit when projected onto the model (see column (b) in Figure 6).

Given an out-of-sample face, $\mathbf{s}$, the optimal parameter vector (in a least squares sense) is given by simply projecting the face onto the model, i.e. $\mathbf{c}^* = \mathbf{P}^T(\mathbf{s} - \bar{\mathbf{s}})$. Substituting $\mathbf{c}^*$ back into Equation 1, we obtain $\mathbf{s}^{\text{mod}}$, the shape which minimises $\|\mathbf{s}^{\text{mod}} - \mathbf{s}\|^2$. We do this for each face and measure the distance of the resulting point in parameter space from the mean (in terms of squared Mahalanobis distance). We vary the number of model dimensions and show the results in Figure 3. The blue line shows the expected vector length which grows linearly with the number of dimensions. We indicate the expected spread of vector lengths by using error bars to show three standard deviations either side of the mean. Red points represent the vector length for a sample projected onto the model.

There are a number of interesting observations to make on this plot. For a small number of model dimensions, the samples adhere to the statistical prediction. However, as the number of dimensions increases, certain samples deviate rapidly from the prediction. These are faces which are dissimilar to those in the training set and whose shape is poorly approximated by the model eigenvectors. The two extreme cases indicated by the green squares are gross overfits, visualisations of which can be seen in Figure 6. The explanation for this is that with sparse training data, the less significant eigenvectors are unstable and cannot be reliably estimated. A much larger training set may mitigate this problem and lead to a model for which out-of-sample faces adhere more closely to the statistical prediction. Nevertheless, to retain the expressiveness of the model the dimensions with smaller eigenvalues are important and cannot be discarded. What is required are constraints which prevent overfitting and ensure that model instances remain plausible. Our proposal to do this by enforcing a hard constraint on the parameter vector lengths is evaluated in the next section.

2.5. Empirical Evaluation: Manifold Approximation

Irrespective of how well out-of-sample data adheres to the manifold assumption, from a practical perspective the more important question is whether forcing samples to lie on the manifold provides a useful constraint. We attempt to answer this by measuring the effect of enforcing the manifold constraint on the "plausibility" of a face. For a face to be plausible it must appear face-like but to be a plausible representation of a specific face it must also have a low perceptual error between the original face and its model representation. There are many proposed measures for computing the perceptual error between a mesh and its reconstruction. Most are based on the surface derivatives since it is surface orientation which determines appearance. Hence, we measure perceptual error in terms of the angular difference between surface normals.

We compare the optimal model-based reconstruction described above, $\mathbf{s}^{\text{mod}}$, to that obtained by projecting $\mathbf{c}^*$ to the closest point on the hyperspherical manifold:

$$\hat{\mathbf{c}}^{\text{man}} = \frac{\sqrt{n}}{\|\hat{\mathbf{c}}^*\|}\hat{\mathbf{c}}^*. \tag{5}$$

We refer to the resulting shape as $\mathbf{s}^{\text{man}}$. It should be noted we use the variance-normalised parameter vectors in Equation 5. We begin by establishing whether the expected vector length predicted by the chi-square distribution is a good choice with which to define the manifold. To do so, in Figure 4 for the 10

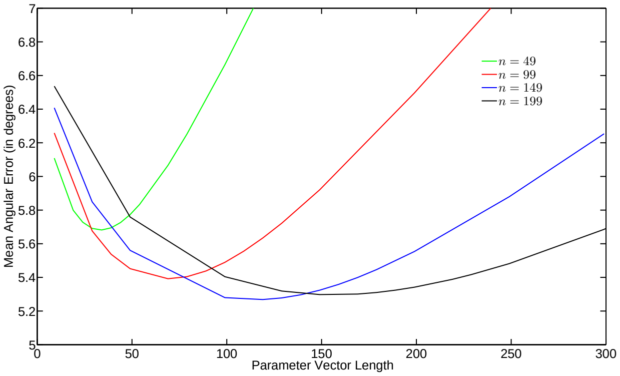

Figure 4: Mean angular error vs Parameter Vector Length (for four different values of model dimensions retained). All reported error measures are averaged over the 10 unseen faces in the BFM.

out-of-sample faces in the BFM, we show the effect when the vector given by projection onto the model is rescaled to various lengths. The x-axis shows the enforced vector length, the y-axis shows the mean angular error in the surface normals (i.e. perceptual error). We perform this test for $n = 49$, 99, 149 and 199 parameter model.

The plot for each model shows a similar trend, with small and large vector lengths having a higher error (underfitting and overfitting respectively) and a minimum occurring close to the chi-square prediction (i.e. when the parameter vector lengths are forced to n). This suggests our statistically motivated choice of hard constraint is reasonable.

Finally, we wish to show that forcing samples to lie on the manifold reduces perceptual error. In Figure 5 (a) we plot the mean Euclidean error for $\mathbf{s}^{\text{mod}}$ and $\mathbf{s}^{\text{man}}$. Since it is optimal, $\mathbf{s}^{\text{mod}}$ achieves a lower Euclidean error than $\mathbf{s}^{\text{man}}$ for all n and this error decreases monotonically as the number of dimensions increases. However, the purpose of our choice of manifold is to enforce plausibility. If we repeat the same experiment but instead plot angular (perceptual) error, shown in Figure 5 (b), we see that $\mathbf{s}^{\text{man}}$ achieves a lower angular error than $\mathbf{s}^{\text{mod}}$ for all n. Increasing the number of model dimensions yields an almost

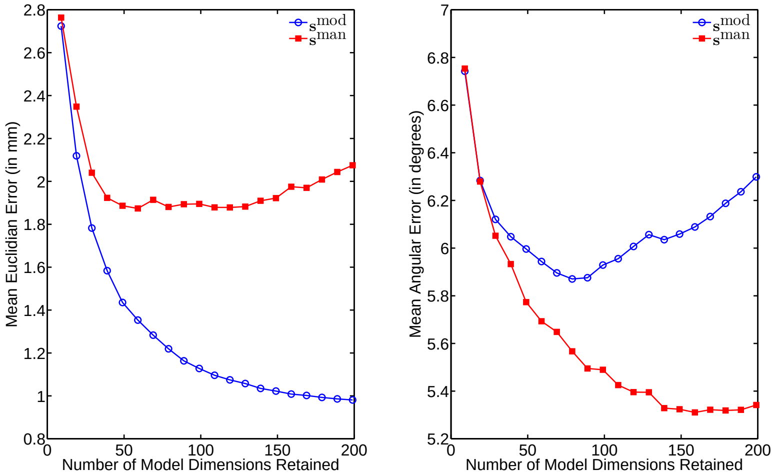

Figure 5: Left plot (a): Mean Euclidean error vs Number of model dimensions retained; Right plot (b): Mean angular error vs Number of model dimensions retained. All reported error measures are averaged over the 10 unseen faces in the BFM.

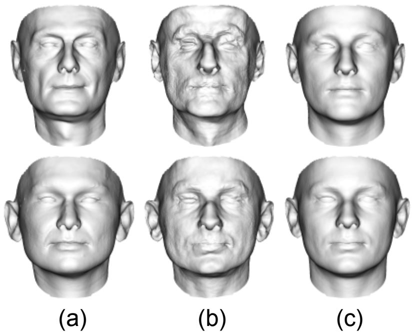

Figure 6: Manifold approximation: (a) input unseen face; (b) least squares fit to vertices; (c) parameter vector of (b) rescaled to manifold. All the results are for a $n = 199$ parameter model.

monotonic reduction in perceptual error for $\mathbf{s}^{\text{man}}$, whilst the perceptual error of the optimal least-squares surfaces ($\mathbf{s}^{\text{mod}}$) begins to increase beyond about 80 dimensions. Two visual examples are shown in Figure 6. The two out-of-sample faces in column (a) are grossly overfitted when allowed to minimise least squares

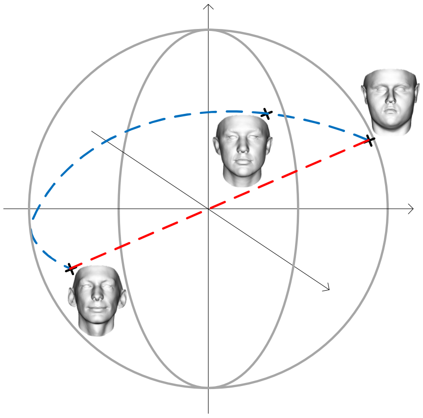

Figure 7: Warping between face and antiface on the S^2 manifold. A linear warp is shown in red and one of the possible plausibility-preserving warps is shown in blue.

error (column (b)). When rescaled to the manifold (column (c)) the perceptual error reduces and the faces are visually plausible.

3. Plausibility-preserving warps and averages

We now demonstrate a simple application of the manifold to warping and averaging of faces.

3.1. Warps

Warping between faces or, more generally, computing weighted combinations of two or more faces has applications in animation and in the production of stimuli for psychological experiments [7]. The most obvious way to warp between two shapes that are in dense correspondence is to linearly warp each vertex from its position in one shape to its position in the other. Equivalently, this can be approximated by linearly warping between the two vectors of PCA parameters. However, in either case the intermediate faces will not correspond to plausible faces. Since the manifold of maximally probable distinctiveness is curved, any linear warp will include faces that do not lie on the manifold, with the least plausible face occurring halfway along the warp.

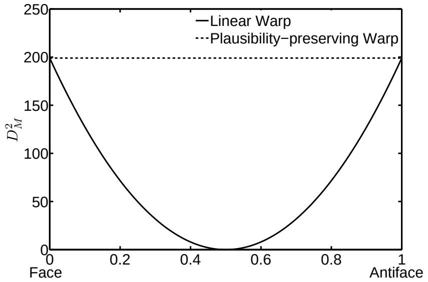

Figure 8: Vector length or 'plausibility' is plotted throughout a warp between a face and antiface (see Figure 9).

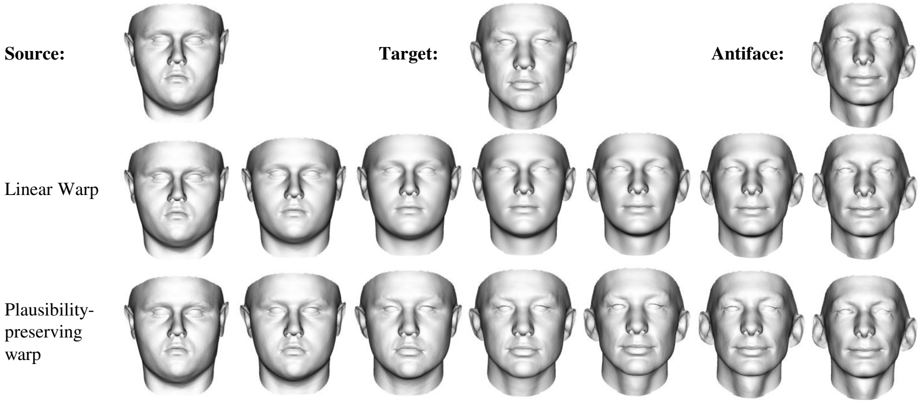

Figure 9: Linear versus plausibility-preserving warp from face to antiface. One of the infinite possible plausibility-preserving warps is specified by providing an intermediate target face.

Face-antiface warps provide a particularly interesting special case. An antiface is the antipodal point of a source face on the manifold. Perceptually, antifaces appear *opposite* in some sense to the original face. The vector connecting a face to its antiface in parameter space passes through the mean. A linear warp between a face and antiface is therefore well-defined but will include implausible faces for the duration of the warp. There is a further problem with such linear warps. Psychological studies have shown that there is a perceptual discontinuity as the face trajectory crosses the mean [7]. In other words, as identity flips from face to antiface, the perceptual effect of a small movement through face space is exaggerated.

Instead, we propose warps which take place on the manifold, following the geodesic curve between the two source faces. Another way to view these warps is as a rotation of a unit vector in $\mathbb{R}^n$. All intermediate faces in this case have equal distinctiveness and are equally plausible. In the case of antifaces, there are an infinite number of valid warps, all of length π. One way to conceptualise this is that we can set off from a point on the hyperspherical manifold in any direction and reach the antiface after travelling a distance π. An interesting result of this observation is that we can choose any intermediate face as a target which will be visited on the warp from face to antiface. This gives us a way to specify one of the infinite face-antiface warps and may also have interesting applications in generating stimuli for psychological studies. This idea is demonstrated in Figure 7 for the S^2 manifold, which shows the difference between a plausibility-preserving and linear warp.

For a source face x_{src} and intermediate target face x_{tar}, we can define a unit vector in the tangent space, $v \in T_{x_{src}}S^{n-1}$, from x_{src} in the direction of x_{tar}: $v = \frac{\mathrm{Log}_{x_{src}}(x_{tar})}{d(x_{src},x_{tar})}$. A geodesic warp from x_{src} to x_{tar} is therefore given by following this vector by a distance specified by the warping parameter w:

$$x_{war} = \mathrm{Exp}_{x_{src}}\left(w \frac{\mathrm{Log}_{x_{src}}(x_{tar})}{d(x_{src}, x_{tar})} \right). \tag{6}$$

When $w = 0$ we obtain the source face, i.e. $x_{war} = x_{src}$, and when $w = d(x_{src}, x_{tar})$ we obtain the target face, i.e. $x_{war} = x_{tar}$. If we set $w = \pi$ we obtain the antiface to x_{src}. Intermediate faces are obtained when $w \in (0, \pi)$.

We show an example warp from face to antiface via an intermediate target face in Figure 9 using the 199 parameter BFM [33]. Note that the effect is of smooth variation of identity, with each of the intermediate faces containing significant detail. We contrast this with a linear warp through the mean face which results in implausibly smooth intermediate faces and no transition through intermediate identities. In Figure 8 we plot the parameter vector lengths for the linear and plausibility-preserving warps.

In Figure 10 we show results on a different dataset. In this case, we built a

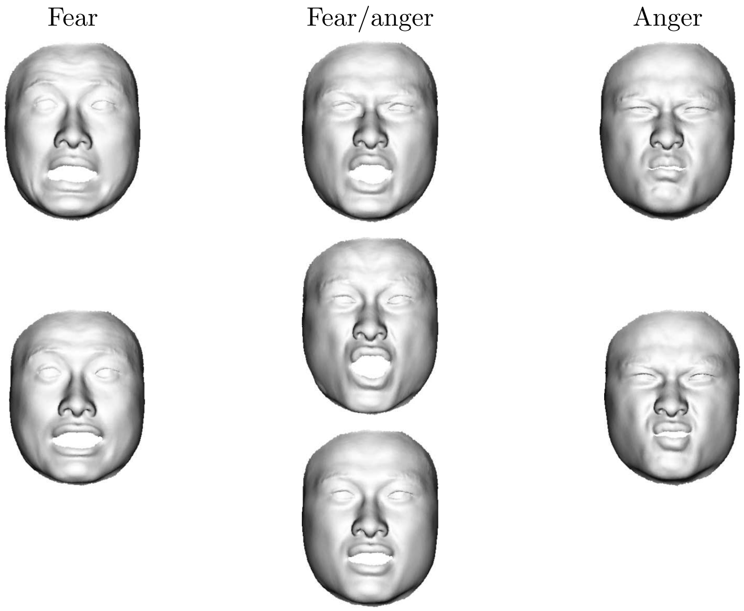

Figure 10: Linear versus plausibility-preserving expression interpolation. First row: 3 frames from captured fear-to-anger sequence. Second row: manifold interpolation between fear and anger. Third row: linear interpolation.

person-specific expression model using scans from the Spacetime faces dataset [34]. As the modal expression, we use the neutral face as the average and model expressions as displacements from neutral. In this case, "identity" is interpreted as the combination of displacements associated with a certain expression. Increasing "distinctiveness" simply exaggerates the magnitude of a particular expression. The manifold therefore spans expressions of equal magnitude and allows plausibility-preserving warping between different expressions.

In the first row, we show three frames from a captured sequence in which the face transitions from an expression of fear to anger. The middle frame contains a mix of the two expressions. All three scans are out of sample of the trained model. In the second row, we project the start and end scans to the manifold and interpolate between the two using our plausibility-preserving warp. In contrast to the linear interpolation shown in the third row, our result correctly predicts the more circular shape of the open mouth and the scrunched eyes, leading to a more detailed expression.

3.2. Averages

Given $u > 2$ source faces, $x_1, \dots, x_u \in S^{n-1}$, we wish to compute an average face which captures characteristics of each of the source faces yet remains plausible itself. The linear or Euclidean mean of the parameter vectors minimises the sum of square error in $\mathbb{R}^n$ from the average to each of the source faces. This is the *extrinsic mean* and will not lie on the manifold. The result is that the face is implausibly smooth and lacking in features. We propose the use of the *intrinsic* or Karcher mean. For $u = 2$, this can be found using the warping equation given above with $w = 0.5$. For $u > 2$, this is the point $x_\mu \in S^{n-1}$ which minimises the total squared geodesic distance to each of the source faces:

$$x_\mu = \arg\min_{x \in S^{n-1}} \sum_{i=1}^{u} d(x, x_i)^2. \tag{7}$$

This point cannot be found analytically, so we solve it as an iterative optimisation using the gradient descent method of Pennec [35]. We initialise our estimate as one of the source data points, i.e. $x_\mu^{(0)} = x_1$. The estimated intrinsic mean is then iteratively updated as follows:

$$x_\mu^{(j+1)} = \mathrm{Exp}_{x_\mu^{(j)}} \left(\frac{1}{u} \sum_{i=1}^{u} \mathrm{Log}_{x_\mu^{(j)}}(x_i) \right). \tag{8}$$

This process converges rapidly, typically within 5 iterations. In Figure 11 we compare our plausibility-preserving averages with linear averaging of the 74 dimensional parameter vectors obtained using the USF data [36]. Notice that each of the Euclidean averages appears unrealistically smooth, whereas the averages computed on the manifold clearly show the presence of distinct features present in the source faces (for example, the broader nostrils of face 1 are visible in the first three averages but not the fourth).

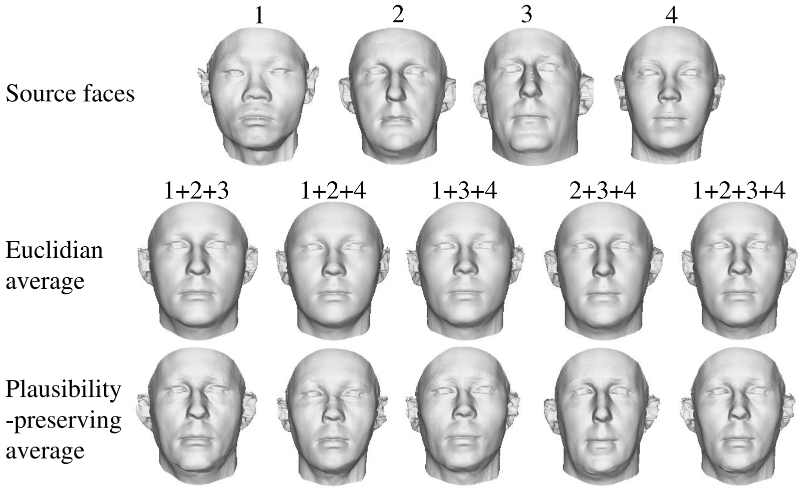

Figure 11: Linear versus plausibility-preserving averages.

4. Model fitting on the manifold of plausible faces

The most powerful application of the identity manifold is to use it for the purpose of constraining the process of fitting a model to data. Suppose the function $\varepsilon : S^{n-1} \mapsto \mathbb{R}$ is an objective function which evaluates the quality of fit of a face represented by a point on the plausibility manifold to some observed data. This function could take any form, for example the difference between predicted and observed appearance in an analysis-by-synthesis framework or the error between a sparse set of feature points. We pose model fitting as finding the point on the manifold which minimises this error, i.e.:

$$x^* = \underset{x \in S^{n-1}}{\arg\min}\, \varepsilon(x). \tag{9}$$

In doing so, we ensure that plausibility is enforced as a hard constraint. Note also that the optimisation is more heavily constrained since the dimensionality of the hypersphere is 1 less than the parameter space.

4.1. Local Optimisation

We can perform gradient descent on the manifold to find a local minimum in the error function. The fact that our manifold is hyperspherical has some interesting implications for such an approach. We must first compute the gradient of the objective function in terms of a vector on the tangent plane: $\nabla\varepsilon(x) \in T_x S^{n-1}$. To do so, we compute the gradient in terms of a vector in $\mathbb{R}^n$ and project the result to the tangent plane as follows:

$$\nabla\varepsilon(x) = \mathrm{Log}_x \left(\Phi^{-1} \left(\frac{\mathbf{x} - \mathbf{g}}{\|\mathbf{x} - \mathbf{g}\|} \right) \right) \tag{10}$$

where $\mathbf{x} = [x_1 \ \dots \ x_n]^T = \Phi(x)$. The gradient in $\mathbb{R}^n$:

$$\mathbf{g} = [\partial_{x_1}\varepsilon(x) \ \dots \ \partial_{x_n}\varepsilon(x)]^T , \tag{11}$$

is approximated by using finite differences to calculate the partial derivatives:

$$\partial_{x_i}\varepsilon(x) \approx \frac{\varepsilon(x_i') - \varepsilon(x)}{\epsilon}, \tag{12}$$

where $x_i' = \Phi^{-1}([x_1 \ \dots \ x_i + \epsilon \ \dots \ x_n])$.

With a means to compute the gradient, we can iteratively minimise the objective function by adapting the gradient descent algorithm to operate on the manifold:

$$x^{(t+1)} = \mathrm{Exp}_{x^{(t)}} \left(-\gamma \nabla\varepsilon(x^{(t)}) \right) , \tag{13}$$

where γ is the step size. Note that as γ varies, the point $\mathrm{Exp}_x\left(-\gamma\nabla\varepsilon(x)\right) \in S^{n-1}$ traces out a great circle about the hypersphere. This is the search space for the one-dimensional line search at each iteration of gradient descent.

4.2. Coarse-to-fine Model Fitting

The difficulty with our approach is choosing an unbiased initialisation. Existing methods for fitting statistical models to data typically commence from an initialisation of the mean (i.e. zero parameter vector), e.g. [4, 3]. However, this point lies far from the plausibility manifold and is therefore unsuitable in our case.

We tackle this problem and also reduce susceptibility to becoming trapped in local minima by proposing a coarse-to-fine algorithm which iteratively increases the number of model dimensions considered in the optimisation.

Consider in the simplest case a two dimensional model (in the one dimensional case the manifold collapses to a pair of points and is the boundary of a line segment). In two dimensions the manifold is S^1 (i.e. a unit circle) and requires the optimisation of a single angular parameter, θ. The result in two dimensions, $\mathbf{x}^{(2)} = [\cos\theta^* \;\; \sin\theta^*]^T$, is given by solving the following interval search problem:

$$\theta^* = \arg\min_{\theta} \varepsilon(\Phi^{-1}([\cos\theta \;\; \sin\theta]^T)), \; 0 \leq \theta < 2\pi, \tag{14}$$

which we solve using golden section search [37]. We use this result to initialise the solution in three dimensions, initially setting the third parameter to zero: $\mathbf{x}^{(n)}_{\text{init}} = \left[\mathbf{x}^{(n-1)} \mid 0\right]$. We then perform gradient descent. We continue this process, incrementally adding dimensions to the optimisation, each time setting the new parameter to zero and then performing gradient descent on the new manifold using this as an initialisation. Hence, the result of a local optimisation in n dimensions is used as the initialisation for optimisation in $n+1$ dimensions ensuring that the solution is already constrained to the right region of the manifold.

4.3. Constrained Line Search

The nature of the hyperspherical manifold can be used to inform the step size used in the gradient descent optimisation. Specifically, the step size is bounded and a constrained line search can be performed based on interval search.

We assume that the result in n dimensions has restricted the solution to the correct hemisphere of the hypersphere. Travelling in the direction of the negative gradient reduces the error. To travel in this direction whilst remaining in the same hemisphere means the maximum arc distance that can be moved is $\frac{\pi}{2}$. Hence, the result in n dimensions is given by $\mathbf{x}^{(n)} = h(\gamma^*)$, where

$$h(\gamma) = \text{Exp}_{\Phi^{-1}(\mathbf{x}^{(n)}_{\text{init}})}\left(-\gamma \frac{\nabla\varepsilon\left(\Phi^{-1}(\mathbf{x}^{(n)}_{\text{init}})\right)}{\left\|\nabla\varepsilon\left(\Phi^{-1}(\mathbf{x}^{(n)}_{\text{init}})\right)\right\|}\right). \tag{15}$$

The arc distance γ determines how far we travel along the great circle implied by the gradient of the objective function. Since we wish to constrain our solution to the same hemisphere, γ must lie in the interval $\left[0, \frac{\pi}{2}\right]$ and we hence find γ^* using golden section search [37] to solve: $\gamma^* = \arg\min_{\gamma} h(\gamma),\ 0 \leq \gamma \leq \frac{\pi}{2}$. Multiple iterations of gradient descent can be used each time a dimension is added to the optimisation. In our results we use four iterations per dimension.

5. Model Fitting Examples

For our experimental evaluation, we use the algorithm described above to fit our 3D morphable shape model to unseen data. We show results for two different exemplar objective functions (one overconstrained and one underconstrained) and compare our results with those obtained using two different generic optimisers.

5.1. Overconstrained Optimisation

We choose as an objective function the angular error between surface normals at each vertex of the model. This is an interesting choice of objective function for two reasons. First, the search landscape of the objective function is littered with local minima. Second, the fitted result is likely to have lower perceptual error than a least squares fit directly to the vertices. Whilst such a least squares fit gives minimal geometric error, the result is often a gross over-fit which does not resemble the input face. Minimising the surface normal error is a non-linear problem which is related to minimising appearance error, as undertaken by analysis-by-synthesis of image data [4].

From an input face shape, represented by p vertices, we compute surface normals at each vertex. If $\mathbf{N}^i$ is the surface normal at vertex i, our objective function is the sum of squared angular errors between input and model surface normals:

$$\varepsilon(x) = \sum_{i=1}^{p} \left(\arccos(\mathbf{n}^i(\Phi(x)) \cdot \mathbf{N}^i)\right)^2, \tag{16}$$

where $\mathbf{n}^i([x_1 \ \dots \ x_n])$ is the surface normal of the ith vertex of the shape given by: $\bar{\mathbf{s}} + \mathbf{P}\mathbf{c}$, where the parameter vector is computed by transforming the unit

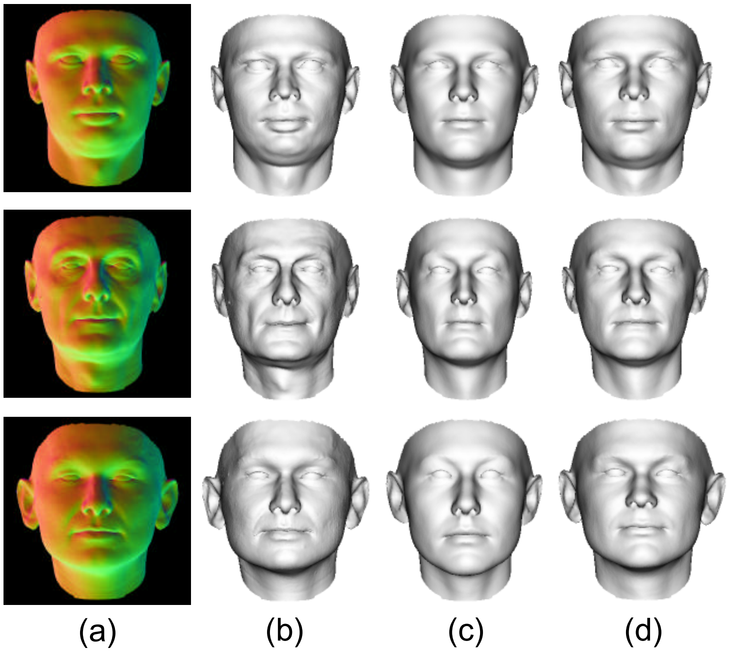

Figure 12: Overconstrained model fitting example: (a) input bump maps for 3 unseen subjects; (b) ground truth; (c) BFGS optimisation; (d) manifold optimisation. All the results are for a $n = 99$ parameter model.

vector back to the hyperellipse:

$$\mathbf{c} = \sqrt{n} \left[x_1 \sqrt{\lambda_1} \ \ldots \ x_n \sqrt{\lambda_n} \right]^T . \tag{17}$$

We compare our manifold optimisation with direct optimisation of the objective function using a generic optimiser based on the BFGS Quasi-Newton method with a cubic line search [38]:

$$\mathbf{c}^* = \arg \min_{\mathbf{c}} \sum_{i=1}^{p} \left(\arccos(\mathbf{n}^i(\mathbf{c}) \cdot \mathbf{N}^i) \right)^2 . \tag{18}$$

Note that the generic optimiser converges close to the mean if all parameters are optimised simultaneously. We therefore take the same coarse-to-fine approach as for the manifold fitting, whereby we iteratively increase the number of dimensions considered in the optimisation.

We provide results on the BFM [33] data. The scans were obtained using the structured light scanning system provided by ABW-3D [33] and are set into correspondence using a modified version of the Optimal Step Nonrigid ICP Algorithm [39]. In Figure 12, column (a) shows input bump maps for three unseen subjects. Column (b) shows the ground truth shape estimates. Column (c) shows the result of using the BFGS non-linear optimiser to solve Equation 18. Because of local minima close to the mean, these faces are implausibly smooth. Finally, our manifold fitting result is shown in column (d). Note that this result represents a trade off between over and underfitting. Averaged over all the out-of-sample faces in the BFM, the angular error of the surface normals is 7.23° for the BFGS method and 5.33° for our method.

5.2. Underconstrained Optimisation

We now consider an objective function which is highly underconstrained. In other words, solutions which minimise the objective function lead to highly implausible faces. The problem we consider is estimation of a high resolution 3D face surface given the positions of $k = 70$ 2D annotations ($k << p$). A linear version of this problem has been considered previously [40], where it was observed that the problem leads to a trade off between the quality of fit to the observed data and prior probability as measured by the model. The parameter to control this trade off can be determined heuristically [40], although no single value will give optimal performance for all faces. In contrast, our proposed approach requires no such regularisation constraint and ensures that the fitted results have high quality shape estimates which are plausible.

Our aim is to recover face shape parameter estimates from a set of k 2D annotations. We represent the ith observed feature point by $\mathbf{L_i} \in \mathbb{R}^\mathbf{2}$. We write $\mathbf{r}^i(\Phi(x)) \in \mathbb{R}^4$ for the 3D position of the vertex corresponding of the ith feature point represented in homogeneous coordinates. This is extracted from the model shape vector given by: $\bar{\mathbf{s}} + \mathbf{Pc}$, where the parameter is vector $\mathbf{c}$ is computed by transforming the unit vector $\Phi(x)$ back to the hyperellipse using

Equation 17. The corresponding projected 2D position is given by:

$$\hat{\mathbf{L_i}} = [e_i/g_i \;\; f_i/g_i]^T \,, \tag{19}$$

where $[e_i \; f_i \; g_i]^T = \mathbf{C}\mathbf{r}^i(\Phi(x))$ and $\mathbf{C} \in \mathbb{R}^{3\times 4}$ is a camera matrix [41] which performs a perspective projection. Our objective function is taken by measuring the sum squared Euclidean distances between the observed 2D feature point positions and the projected model estimates:

$$\varepsilon(x) = \sum_{i=1}^{k} \|\mathbf{L_i} - \hat{\mathbf{L_i}}\|^{\mathbf{2}}. \tag{20}$$

The conversion from homogeneous to 2D Euclidean coordinates means that the error is a nonlinear function of the shape parameters. We assume that the camera matrix is known, since our aim here is to evaluate a simple underconstrained, nonlinear objective function. However, for a real world implementation this can be estimated using the Gold Standard algorithm [41] and the two steps of pose and shape estimation iterated to convergence.

We compare our manifold optimisation with direct optimisation of the objective function using a generic optimiser based on the Levenberg-Marquardt algorithm (LMA) [42]. Note that since the problem is underconstrained, direct optimisation of the objective function using LMA leads to gross overfitting. We therefore also provide results for the regularised version:

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \sum_{i=1}^{k} \|\mathbf{L}_i - \hat{\mathbf{L}}_i\|^2 + \eta D_M^2(\mathbf{c}), \tag{21}$$

where η is a constant which controls the influence of the regularisation term. It should be noted in this case $\hat{\mathbf{L}}_i = \mathbf{C}\mathbf{r}^i(\mathbf{c})$.

In Figure 13 we show results on the BFM [33] data. Column (a) shows the ground truth faces (unseen) with the input feature points (blue circles). Column (b) shows the result of using LMA to solve Equation 21, with $\eta = 0$. In this case there is no regularisation constraint applied and hence we obtain grossly overfitted shape estimates. Column (c) shows the result of solving Equation 21 using LMA, with η chosen experimentally to provide optimal average performance. To provide stable performance over all faces, the regularisation weight

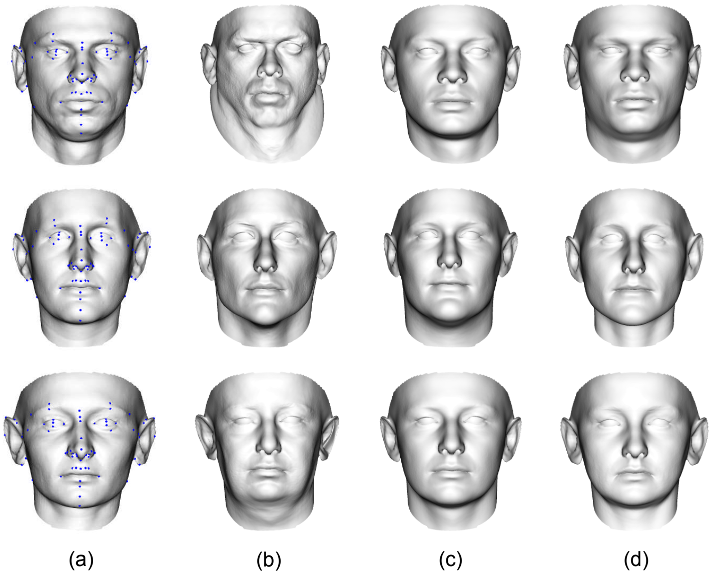

Figure 13: Underconstrained model fitting example: (a) ground truth faces (unseen) with the input feature points (blue circles); (b) LMA with no regularisation constraint; (c) LMA with regularisation constraint;(d) manifold optimisation. All the results are for a $n = 50$ parameter model.

must be set conservatively which means some solutions are restricted to lie too close to the mean. The resulting faces therefore lack salient detail. Figure 14 demonstrates the effect of varying the regularisation weight. Finally, our manifold fitting result is shown in column (d). This result represents a trade off between over and underfitting via the hard manifold constraint. Our method does not require tuning of a parameter and provides stable performance in all cases. Table 1 tabulates the mean Euclidean error over all vertices in the mesh averaged over all the out-of-sample faces in the BFM.

6. Conclusions

We have shown how a number of useful operations can be performed on the manifold of equally distinctive faces. This provides a new way to constrain operations involving the parameters of a statistical model. In particular, we have shown how to constrain the process of fitting a model to data which is

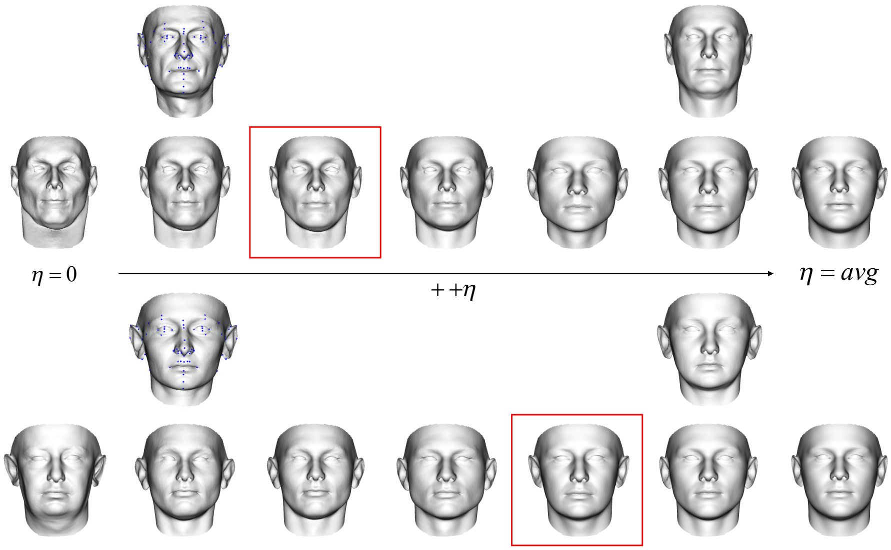

Figure 14: Demonstration of parameter selection for LMA optimisation of regularised objective. For two different subjects (shown on the left in the first and third rows), the optimal regularisation weight (corresponding to the shape estimate with the bounded box) is different. The manifold solution (shown on the right in the first and third rows) does not require parameter tuning and provides improved results.

robust but does not require the selection of a regularisation weight parameter. We avoid using a biased initialisation and improve efficiency by using a coarse-to-fine strategy. This approach outperforms the use of two generic nonlinear optimisation algorithms on two different objective functions. In this paper we provide experimental results for facial data. However, our approach could be applied to any source of data modeled using a linear statistical model. In the future we aim to consider whether nonlinear methods for deriving the statistical model could negate the need for enforcing an additional manifold constraint in parameter space, i.e. to derive a face space which, by construction, contains only plausible faces.

In addition, it would be interesting to compare our model fitting results against other methodologies that seek to preserve local, high frequency detail. For example, using a richer hierarchical model [17] which includes parameters

Method	Euclidean Error (in mm)
LMA without regularization	5.68
LMA with regularization	4.81
Our method	4.21

Table 1: Model fitting results (underconstrained objective).

to describe local deformations from the global model is an alternative modelling paradigm. We could compare whether the additional cost of many more model dimensions provides a significant benefit over our manifold constraint within a classical PCA-based model. Finally, we would like to apply our model fitting approach to more challenging datasets that include missing data (occlusions) and noise. With a suitable outlier rejection scheme, we expect that our manifold fitting approach would still improve distinctiveness of the results by avoiding overfitting/underfitting to unoccluded/clean data, whilst ensuring a plausible level of detail in missing regions. Any appropriate robust model fitting algorithm could be adapted to work with the manifold constraint for this purpose. For example, the two objective functions we consider in this paper are sums of squared residuals. These could be replaced with any robust error measure, subject to the resulting objective having a gradient that can be derived or numerically estimated.

References

[1] T. F. Cootes, C. J. Taylor, D. Cooper, J. Graham, Active shape models – their training and application, Comput. Vis. Image Underst. 61 (1995) 38–59.

[2] R. Knothe, S. Romdhani, 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, 2006, pp. 637–644.

[3] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, in: Proc. ECCV, 1998, pp. 484–498.

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

[5] Y. M. Lui, J. R. Beveridge, Grassmann registration manifolds for face recognition, in: Computer Vision–ECCV 2008, Springer, 2008, pp. 44–57.

[6] A. Patel, W. A. P. Smith, 3D morphable face models revisited, in: Proc. CVPR, 2009, pp. 1327–1334.

[7] A. J. O'Toole, T. Vetter, H. Volz, E. M. Salter, Three-dimensional caricatures of human heads: Distinctiveness and the perception of facial age, Perception 26 (6) (1997) 719–732.

[8] T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition, Quarterly Journal of Experimental Psychology A 43 (2) (1991) 161–204.

[9] J. Xiao, S. Baker, I. Matthews, T. Kanade, Real–time combined 2D+3D active appearance models, in: Proc. CVPR, 2004, pp. 535–542.

[10] B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, T. Vetter, Reconstructing high quality face-surfaces using model based stereo, in: Proc. ICCV, 2007.

[11] C. Hu, J. Xiao, I. Matthews, S. Baker, J. Cohn, T. Kanade, Fitting a single active appearance model simultaneously to multiple images, in: Proc. BMVC, 2004.

[12] V. Blanz, K. Scherbaum, H.-P. Seidel, Fitting a morphable model to 3D scans of faces, in: Proc. ICCV, 2007.

[13] I. Matthews, S. Baker, Active appearance models revisited, Int. J. Comput. Vis. 60 (2) (2004) 135–164.

[14] S. Romdhani, T. Vetter, Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior, in: Proc. CVPR, Vol. 2, 2005, pp. 986–993.

[15] S. Romdhani, V. Blanz, T. Vetter, Face identification by fitting a 3d morphable model using linear shape and texture error functions, in: Proc. ECCV, 2002, pp. 3–19.

[16] O. Aldrian, W. A. P. Smith, Inverse rendering of faces with a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2013) 1080–1093.

[17] A. Brunton, A. Salazar, T. Bolkart, S. Wuhrer, Review of statistical shape spaces for 3D data with comparative analysis for human faces, Comput. Vis. Image Underst. 128 (2014) 1–17.

[18] B. Berkels, P. T. Fletcher, B. Heeren, M. Rumpf, B. Wirth, Discrete geodesic regression in shape space, in: Proc. EMMCVPR, 2013, pp. 108–122.

[19] B. Heeren, M. Rumpf, P. Schröder, M. Wardetzky, B. Wirth, Exploring the geometry of the space of shells, Computer Graphics Forum 33 (5) (2014) 247–256.

[20] D. Boscaini, D. Eynard, D. Kourounis, M. M. Bronstein, Shape-from-operator: recovering shapes from intrinsic operators, Computer Graphics Forum (Proc. Eurographics) 34 (2) (2015) 265—274.

[21] N. Shapira, M. BenChen, Crosscollection map inference by intrinsic alignment of shape spaces, Computer Graphics Forum 33 (5) (2014) 281–290.

[22] R. Adler, J.-P. Dedieu, J. Margulies, M. Martens, M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. of Numerical Analysis 22 (2002) 359–390.

[23] J. Manton, Optimization algorithms exploiting unitary constraints, IEEE Trans. Signal Proces. 50 (3) (2002) 635–650.

[24] Y. Ma, J. Košecká, S. Sastry, Optimization criteria and geometric algorithms for motion and structure estimation, Int. J. Comput. Vis. 44 (3) (2001) 219–249.

[25] U. Helmke, K. Hüper, J. Moore, Quadratically convergent algorithms for optimal dextrous hand grasping, IEEE Trans. Robotic. Autom. 18 (2) (2002) 138–146.

[26] P. Y. Lee, J. B. Moore, Gauss-newton-on-manifold for pose estimation, J. Ind. Manag. Optim. 1 (4) (2005) 565–587.

[27] K. Huper, J. Trumpf, Newton-like methods for numerical optimization on manifolds, in: Proc. IEEE Asilomar Conference on Signals Systems and Computers, 2004, pp. 136–139.

[28] N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a Matlab toolbox for optimization on manifolds, Journal of Machine Learning Research 15 (1) (2014) 1455–1459.

[29] A. Patel, W. A. P. Smith, Exploring the identity manifold: Constrained operations in face space, in: Proc. ECCV, 2010.

[30] M. Meytlis, L. Sirovich, On the dimensionality of face space, IEEE Trans. Pattern Anal. Mach. Intell. 29 (7) (2007) 1262–1267.

[31] A. M. Burton, J. R. Vokey, The face-space typicality paradox: understanding the face-space metaphor., The Quarterly Journal of Experimental Psychology A 51 (3) (1998) 475–483.

[32] M. Zhang, P. T. Fletcher, Probabilistic principal geodesic analysis, in: Advances in Neural Information Processing Systems, 2013, pp. 1178–1186.

[33] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D face model for pose and illumination invariant face recognition, in: Proc. IEEE Intl. Conf. on Advanced Video and Signal based Surveillance, 2009.

[34] L. Zhang, N. Snavely, B. Curless, S. M. Seitz, Spacetime faces: high resolution capture for modeling and animation, ACM Trans. Graphic. (Proceedings of SIGGRAPH) 23 (3) (2004) 548–558.

[35] X. Pennec, Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements, in: Proc. IEEE Workshop on Nonlinear Signal and Image Processing, 1999.

[36] USF HumanID 3D Face Database, Courtesy of Sudeep. Sarkar, University of South Florida, Tampa, FL.

[37] J. Kiefer, Sequential minimax search for a maximum, Proceedings of the American Mathematical Society 4 (3) (1953) 502–506.

[38] C. G. Broyden, The convergence of a class of double-rank minimization algorithms, Journal of the Institute of Mathematics and Its Applications 6 (1) (1970) 76–90.

[39] B. Amberg, S. Romdhani, T. Vetter, Optimal step nonrigid ICP algorithms for surface registration, in: Proc. CVPR, 2007.

[40] V. Blanz, A. Mehl, T. Vetter, H.-P. Seidel, A statistical method for robust 3D surface reconstruction from sparse data, in: Proc. 3DPVT, 2004, pp. 293–300.

[41] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.

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