A 3D Morphable Model learnt from 10,000 faces

James Booth* Anastasios Roussos* Stefanos Zafeiriou*,* Allan Ponniah†
David Dunaway†
*Imperial College London, UK †Great Ormond Street Hospital, UK
∗ Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland
*{james.booth,troussos,s.zafeiriou}@imperial.ac.uk, †{allan.ponniah,david.dunaway}@gosh.nhs.uk

Abstract

We present Large Scale Facial Model (LSFM) — a 3D Morphable Model (3DMM) automatically constructed from 9,663 distinct facial identities. To the best of our knowledge LSFM is the largest-scale Morphable Model ever constructed, containing statistical information from a huge variety of the human population. To build such a large model we introduce a novel fully automated and robust Morphable Model construction pipeline. The dataset that LSFM is trained on includes rich demographic information about each subject, allowing for the construction of not only a global 3DMM but also models tailored for specific age, gender or ethnicity groups. As an application example, we utilise the proposed model to perform age classification from 3D shape alone. Furthermore, we perform a systematic analysis of the constructed 3DMMs that showcases their quality and descriptive power. The presented extensive qualitative and quantitative evaluations reveal that the proposed 3DMM achieves state-of-the-art results, outperforming existing models by a large margin. Finally, for the benefit of the research community, we make publicly available the source code of the proposed automatic 3DMM construction pipeline. In addition, the constructed global 3DMM and a variety of bespoke models tailored by age, gender and ethnicity are available on application to researchers involved in medically oriented research.

1. Introduction

3D Morphable Models (3DMMs) are powerful 3D statistical models of human face shape and texture. In the original formulation, as presented by the seminal work of Blanz and Vetter [6], a 3DMM used in an analysis-by-synthesis framework was shown to be capable of inferring a full 3D facial surface from a single image of a person. 3DMMs have since been widely applied in numerous areas, such as computer vision, human behavioral analysis, computer graphics and craniofacial surgery [7, 3, 2, 28].

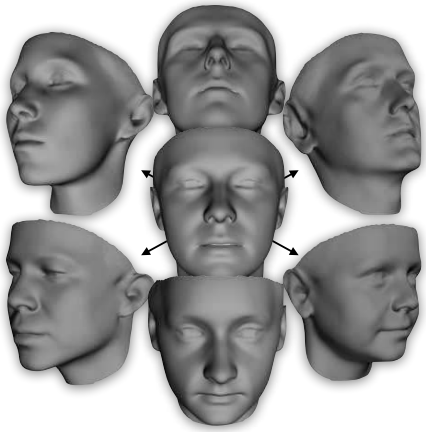


Figure 1: The sheer number of facial meshes used in training LSFM produces a 3D Morphable Model with an unprecedented range of human identity in a compact model.

A 3DMM is constructed by performing some form of dimensionality reduction, typically Principal Component Analysis (PCA), on a training set of facial meshes. This is feasible if and only if each mesh is first re-parametrised into a consistent form where the number of vertices, the triangulation, and the anatomical meaning of each vertex are made consistent across all meshes. Meshes satisfying the above properties are said to be in dense correspondence with one another. Whilst this correspondence problem is easy to state, it is challenging to solve accurately and robustly between highly variable facial meshes.

Once built, 3DMMs provide two functions. Firstly, 3DMMs are powerful priors on 3D face shape that can be leveraged in fitting algorithms to reconstruct accurate and complete 3D representations of faces from data-deficient sources like in-the-wild 2D images or noisy 3D depth scan data. Secondly, 3DMMs provide a mechanism to encode any 3D face in a low dimensional feature space, a compact

representation that makes tractable many 3D facial analysis problems.

In this paper we revisit 3DMMs under a new context — that we have access to a database of around 10,000 high-quality 3D facial scans, and for each subject we have detailed demographics including the subject's age, gender, and ethnic background. We show clear evidence that the manifold of plausible faces is clustered by demographics like age and ethnicity, and use this insight to devise new approaches to 3DMM construction and fitting that advance on the state of art. We further demonstrate for the first time that a large-scale model coupled with accurate demographics enables accurate age classification from 3D shape data alone.

2. Previous Work

The construction of a 3DMM usually consists of two main steps — establishing group-wise dense correspondence between a training set of facial meshes, and then performing some kind of statistical analysis on the registered data to produce a low-dimensional model.

In the original formulation [6], Blanz and Vetter solved the dense correspondence problem by representing each facial mesh in a cylindrical 'UV' map, flattening each 3D surface down into a 2D space. This reduced establishing correspondence to a well-understood image registration problem, which was solved with a regularised form of optical flow. Blanz and Vetter employed PCA to construct their model, and showed that in their framework, model performance was improved by segmenting the facial surface into regions (eyes, nose, mouth, other), building individual models per-segment, before blending resulting segments back together. Amberg et al. [3] extended this approach to emotive facial shapes by adopting an additional PCA modeling of the offsets from the neutral pose. This resulted to a single linear model of both identity and expression variation of 3D facial shape.

Blanz and Vetter's correspondence technique was only used to align the facial meshes of 200 subjects of a similar ethnicity and age [6]. This approach was effective in such a constrained setting, but it is fragile to large variance in facial identity. To overcome this limitation, Patel and Smith [23] proposed to manually annotate the cylindrical face projections with a set of sparse annotations, employing a Thin Plate Splines (TPS) warp [11] to register the UV images of the meshes into a common reference frame. Cosker et al. [16] automated the procedure of landmark annotations required for the TPS warp, for the special case of temporal sequences of a single identity displaying emotions. Several facial landmarks on a handful of meshes for a given temporal sequence were manually annotated and used to build a person-specific Active Appearance Model (AAM) [15] that was then used to automatically find sparse annotations for each frame in the data set.

As an alternative to performing alignment in a UV space, Paysan et al. [24] built the Basel Face Model (BFM) by using an optimal step Nonrigid Iterative Closest Point (NICP) algorithm [4] to directly align scans of 200 subjects with a template. This native 3D approach was guided by manually-placed landmarks to ensure good convergence.

Brunton et al. [13] adopt wavelet bases to model independent prior distributions at multiple scales for the 3D facial shape. This offers a natural way to represent and combine localised shape variations in different facial areas.

Vlasic et al. [30] modeled the combined effect of identity and expression variation on the facial shape by using a multilinear model. More recently, Bolkart and Wuhrer [10] show how such a multilinear model can be estimated directly from the training 3D scans by a joint optimisation over the model parameters and the groupwise registration of the 3D scans.

For the case where a temporal sequence of meshes is available, Bolkart and Wuhrer [9] fit a multilinear model and estimate a 4D sequence parametrisation. This can be used to animate a single 3D scan with a specific facial expression. Another alternative to modeling emotive faces is the blendshape model, which was used by Salazar et al. [27] to place into correspondence emotive faces in a fully automated way. For more details on 3D facial shape modeling, we refer the interested reader to the recent extensive review article of Brunton et al. [14].

Due to the costly manual effort currently required to construct 3DMMs from 3D data, recent efforts in the field have also focused on trying to build models from other data sources. Kemelmacher recently presented a technique that attempts to learn a full 3D facial model automatically from thousands of images [21]. Whilst impressive given the input data, such techniques cannot currently hope to produce models comparable in resolution and detail to techniques that natively process 3D input data.

All the aforementioned works do not use more than 300 training facial scans. In this paper we show that such a size of training set is far from adequate to describe the full variability of human faces. On top of that, all existing works use training sets with a very limited diversity in the ethnic origin (mostly Caucasian) as well as in the age (mostly young and middle adulthood) of the subjects. Due to this kind of limitations of the training sets adopted, no existing work so far, to the best of our knowledge, has developed models tailored for specific age, gender or ethnicity groups. The above issues pose severe limitations in the descriptive power of the resultant morphable models.

Regarding public availability of 3DMMs of human faces, there exist only two available resources: First, a University of Basel website [25] that provides the BFM model [24]. Second, a website of Bolkart, Brunton, Salazar and Wuhrer [8] that provides the 3DMMs constructed by

their recent works, modeling 3D face shapes of different subjects in neutral expression [14] as well as 3D shapes of different subjects in different expressions [12, 9].

3. Contributions

In this paper, we introduce a robust pipeline for 3DMM construction that is completely automated. More precisely, we develop a novel and robust approach to 3D landmark localisation, followed by dense correspondence estimation using the NICP algorithm. Then, we propose an approach to automatically detect and exclude the relatively few cases of failures of dense correspondence, followed by PCA to construct the deformation basis. We pay particular attention to the efficiency and scalability of all the aforementioned steps. We make the source code of this pipeline publicly available, for the benefit of the community[1].

We then use our pipeline on a 3D facial database of 9,663 subjects to construct LSFM, the largest and most information-rich 3DMM of face shapes in neutral expression produced to date. LSFM is built from two orders of magnitude more identity variation than current state-of-the-art models. We conduct extensive experimental evaluations that show that this additional training data leads to significant improvements in the characteristics of our 3DMM, and demonstrate that LSFM outperforms existing models by a wide margin. We also present experiments that study the effect of using larger datasets in model construction. These experiments provide for the first time a comprehensive answer to the question of how much training data is needed for 3DMMs before effects of diminishing returns set in.

Apart from building LSFM using the commonly-used global PCA, we also build a collection of PCA models tailored by age, gender and ethnicity, capitalising on the rich demographic information of the used database. We present quantitative experimental evidence of why and when such tailored models should be preferred over the global PCA.

Using the demographic information, we are also able to analyse for the first time the distribution of faces on the low-dimensional manifold produced by the global PCA. We visualise the manifold of faces using t-distributed stochastic neighbor embedding (t-SNE) [29], and report on clear age and ethnic clustering that can be observed. As an application example, we utilise the proposed model to perform age classification, achieving particularly accurate results.

Finally, a selection of models is made available from this work, including a global statistical model, and models broken down by demographics[1]. These models were built using data collected for medical research applications, and the models are provided under a similar license.

It is worth mentioning that current progress in computer vision would not be possible without the collection of large and comprehensive datasets e.g. [19, 26, 20, 18], and we believe that our publicly available models contributes towards this effort.

[1] http://www.ibug.doc.ic.ac.uk/resources/lsfm

4. Background

The geometry of a 3D facial mesh is defined by the vector $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T]^T \in \mathbb{R}^{3n}$, where n is the number of vertices and $\mathbf{x}_i = [x_x^i, x_y^i, x_z^i]^T \in \mathbb{R}^3$ describes the X,Y and Z coordinates of the i-th vertex.

4.1. 3DMM construction

The construction of a 3DMM happens in two main stages:

Dense correspondence: A collection of meshes are reparametrised into a form where each mesh has the same number of vertices joined into a triangulation that is shared across all meshes. Furthermore, the semantic or anatomical meaning of each vertex is shared across the collection.

Similarity alignment & statistical modelling: The collection of meshes in dense correspondence are subjected to Procrustes Analysis to remove similarity effects, leaving only shape information. The processed meshes are statistically analysed, typically with PCA [17], generating a 3D deformable model as a linear basis of shapes. This allows for the generation of novel shape instances:

$$\mathbf{X}^* = \boldsymbol{M} + \sum_{i=1}^{d} \alpha_i \boldsymbol{U}_i = \boldsymbol{M} + \mathbf{U}\boldsymbol{\alpha} \tag{1}$$

where $\boldsymbol{M} \in \mathbb{R}^{3n}$ is the mean shape and $\mathbf{U} = [\boldsymbol{U}_1 \cdots \boldsymbol{U}_d] \in \mathbb{R}^{3n \times d}$ is the orthonormal basis matrix whose columns contain the shape eigenvectors $\boldsymbol{U}_i$. Also, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_d] \in \mathbb{R}^d$ is the shape vector that contains the parameters (coefficients) that define a specific shape instance under the given deformable model. The degrees of freedom of this model are given by the number of principal components d, which is much smaller than the dimensionality $3n$ of the original space of 3D shapes.

Any input 3D mesh $\mathbf{X}$ can be projected on the model subspace by finding the shape vector $\boldsymbol{\alpha}$ that generates a shape instance (1) that is as close as possible to $\mathbf{X}$. The optimum shape vector and the corresponding projection $P(\mathbf{X})$ on the model subspace are given by [17]:

$$\boldsymbol{\alpha} = \mathbf{U}^T(\mathbf{X} - \boldsymbol{M}), \; P(\mathbf{X}) = \boldsymbol{M} + \mathbf{U}\mathbf{U}^T(\mathbf{X} - \boldsymbol{M}) \tag{2}$$

4.2. `MeIn3D` face database overview

The collected `MeIn3D` database contains approximately 12,000 3D facial scans captured during a special exhibition in the Science Museum, London, over a period of 4 months. The data was collected with the goal of constructing large-scale statistical models that could ultimately advance facial

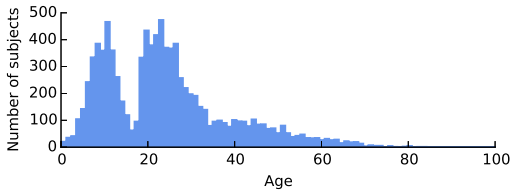


Figure 2: Age distribution of subjects in MeIn3D dataset.

reconstruction and craniofacial surgery. A 3dMD™ face capture system was utilised, creating a 3D triangular surface for each subject composed of approximately 60,000 vertices joined into approximately 120,000 triangles, along with a high-resolution texture map. Furthermore, 9,663 subjects also provided metadata about themselves, including their gender, age and ethnicity. This information allows for the construction of models for targeted populations, such as within a defined age range or from a particular ethnic background. The dataset covers a wide variety of age (see Figure 2), gender (48% male, 52% female), and ethnicity (82% White, 9% Asian, 5% Mixed Heritage, 3% Black and 1% other).

5. Methodology

Let us consider the scenario that, as with MeIn3D database, one has a large-scale database of 3D facial scans and wants to apply a technique to construct a high-quality 3DMM. Such a large database raises some unique scalability challenges. We believe that it is highly beneficial to have a fully automated technique that would not require any kind of manual annotation. It is also very important that this technique is efficient in terms of both runtimes and memory requirements.

We introduce a 3DMM construction pipeline that meets all the aforementioned specifications, see Fig. 3. It starts with a novel and robust approach to 3D landmark localisation. The 3D landmarks are then employed as soft constraints in NICP to place all meshes in correspondence with a template facial surface. With such a large cohort of data, there will be some convergence failures from either landmarking error or NICP. We propose a refinement post-processing step that weeds out problematic subjects automatically, guaranteeing that the LFM models are only constructed from training data for which we have a high confidence of successful processing.

5.1. Automatic annotation

Our proposed technique allows us to bring to bear the huge expertise developed in image landmark localisation to 3D landmark localisation, allowing us to leverage the extensive datasets and state-of-the-art techniques that are now readily available in this domain [1]. This approach is similar to the work of [16] which was shown to be successful for temporal person-specific sequences, but here we pay particular attention to mesh sets with highly variable identity.

We do this by rendering each mesh from one or several virtual cameras positioned around the subject, Fig. 3a. Using the texture information of the 3D mesh, each virtual camera, which has a known perspective projection matrix, records a realistic synthetic face image with a fixed pose. Therefore, we are able to apply an AAM-based state-of-the-art image landmark localisation technique [5], trained for this specific pose and initialised from a state-of-the-art face detector [22, 1]. In this way, a set of 68 sparse annotations in the corresponding synthetic view is robustly located and then back-projected on the 3D facial mesh. In the experiments reported here, we found that a single frontal virtual camera was adequate for accurate results. However, additional virtual cameras can be supported by our pipeline.

5.2. Automatic correspondences & error pruning

After automatic annotation, each mesh is individually placed in correspondence with a template mesh, Fig. 3b. Firstly, the automatic landmarks are used to perform an optimal similarity alignment between the mesh in question and the (annotated) template. NICP is then used to deform the template so that it takes the shape of the input mesh, with the landmarks acting as a soft constraint. The resulting deformed templates are re-parameterised versions of each subject that are in correspondence with one another.

With such a large number of subjects there will be some failure cases at this stage. This is an unavoidable byproduct of the fact that both landmark localisation and NICP are non-convex optimisation problems that are sensitive to initialisation. Our approach embraces this, seeking to weed out the small number of failure cases given the huge amount of data available for processing.

To remove outliers we construct an initial global PCA from all fittings, and then project each subject onto the subspace of this model to derive the corresponding shape vector α (see Eq.(2)). The weighted squared norm of α (using the inverse of the corresponding PCA eigenvalues as weightings) yields an inverse measure of the likelihood of the fitted mesh, under the given PCA model. Therefore, inaccurate fittings exhibit a particularly high value of this norm, with a separation from the values that correspond to the accurate fittings. We thus classify as outliers the meshes whose weighted norm of α belongs at the top 1.5% of the dataset. For more details, please refer to the Supplementary Material.

6. Experiments

In this section, we present and analyse the morphable face models that we constructed. We also perform detailed evaluations and comparisons with other publicly available

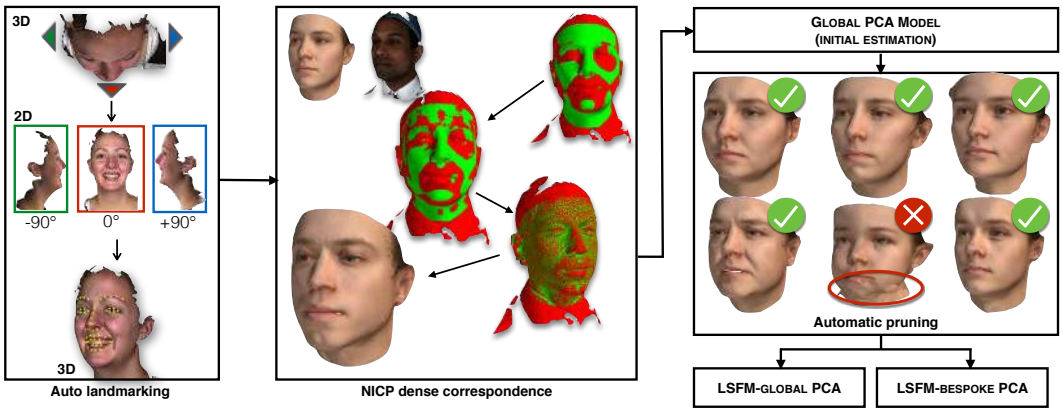


Figure 3: Our fully automated pipeline for constructing large scale 3DMMs. From left to right and top to bottom: (a) Automatic landmarking based on synthetically rendered views. (b) Guided by the automatic landmarks, the 3D template is iteratively deformed to exactly match every 3D facial mesh of the dataset. (c) An initial global PCA is constructed, and (d) erroneous correspondences are automatically removed. (e) LSFM models are constructed from the remaining clean data.

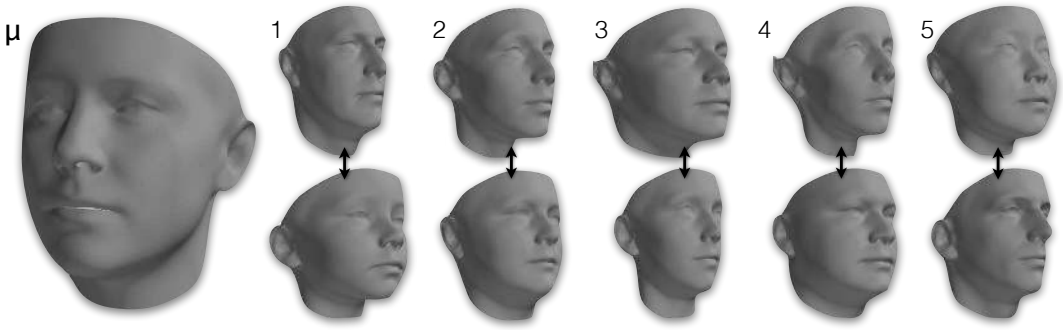


Figure 4: Visualisations of the first five principal components of shape for LSFM, each visualised as additions and subtractions away from the mean (also shown, *left*).

models. For additional results, please see the Supplementary Material.

6.1. Global LSFM model

We derive our global LSFM model (hereafter referred to as *LSFM-global*) by applying the proposed construction pipeline on the MeIn3D dataset. Figure 4 visualises LSFM-global by showing the mean shape along with the top five principal components of shape variation. We observe that the principal modes of variation capture trends of facial shape deformation due to gender, age, ethnicity and other variability in a particularly plausible way, yielding high-quality 3D facial shapes.

An additional visualisation of LSFM-global is provided by Figure 1, which shows synthetic facial shapes generated by the model. It can be seen that all synthetic faces exhibit a high degree of realism, including fine details in the facial structures. Furthermore, we observe that the statistical distribution of LSFM-global succeeds in capturing a plethora of demographic characteristics (age, gender and ethnicity).

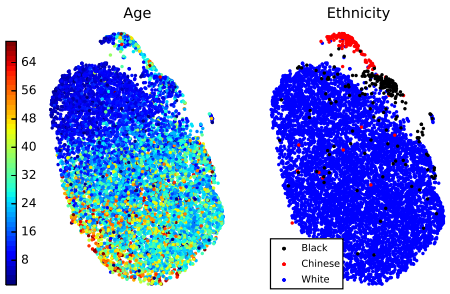


Figure 5: t-SNE embedding of the high dimensional face manifold in two dimensions. **Left:** a clear trend of increasing age can be seen. **Right:** the two smaller structures are explained largely as ethnic variations.

6.2. Facial manifold visualisation

Here, we explore the properties of the LSFM manifold. After establishing dense correspondences with our pipeline and excluding the outliers, every retained training sample

$\mathbf{X}$ is projected on the LSFM-global model and represented by the vector of shape parameters $\boldsymbol{\alpha}$ that yields the closest shape within the model subspace, see Eq. (2). We then apply t-SNE [29] to the shape vectors from all training samples to visualise the manifold of training shapes, as represented in the d-dimensional model subspace.

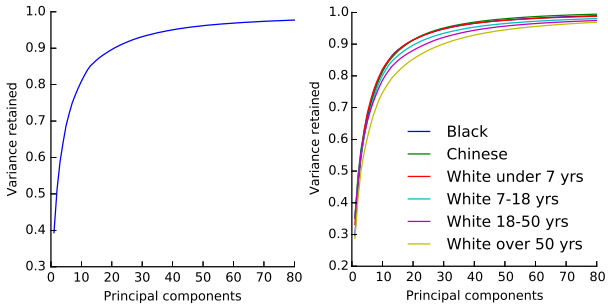
Leveraging the per-subject demographic data we have, we are able to label samples in this space by their age, see Fig. 5 (left). Interestingly, a clear trend of increasing age across the bulk of the manifold can be seen, suggesting that the facial manifold has age-related structure.

Furthermore, we visualise the space by ethnicity, Fig. 5 (right). Note that we chose three ethnic groups for which the number of samples in the used dataset was sufficient for our analysis. We observe that t-SNE has produced a nonlinear 2D embedding that dedicates the largest area for the White ethnic group, which is not surprising, given the fact that this ethnic group is over-represented in the `MeIn3D` database (82% of the samples). What is particularly interesting is the fact that the clusters that are clearly separable from the main manifold are actually specific ethnic groups.

These visualisations provide insight into how different regions of the high-dimensional space of human face shapes are naturally related to different demographic characteristics. We use this insight to define specific *bespoke models* that are trained on dedicated subsets of the full `MeIn3D` training population. Taking also into account the demographics of the training data available (see Section 4.2), we define the following groups: **Black** (all ages), **Chinese** (all ages) and White ethnic group, which due to large availability of training samples, is further clustered into four age groups: under 7 years old (**White-under7**), 7 to 18 years old (**White-7to18**), 18 to 50 years old (**White-18-50**) and over 50 years old (**White-over50**). We combine these bespoke models in a large mixture model, which we call LSFM-bespoke. The intrinsic characteristics of both LSFM-global and LSFM-bespoke will be evaluated in the next section.

6.3. Training and Test Sets

For all the subsequent experiments, `MeIn3D` dataset was split into a training set and a test set. In more detail, a set of 400 meshes of `MeIn3D` was excluded from the original training set to form a test set. This test set was randomly chosen within demographic constraints to ensure a gender, ethnic and age diversity. In particular, it contains the following number of samples from each one of the considered groups: Black: 40, Chinese: 40, White-under7: 80, White-7-to-18: 80, White-18-to-50: 80, and White-over-50: 80. In addition, each of the above amounts was drawn from 50% males and 50% females. Despite the fact that this test set does not capture the full range of diversity present in the demographics of humans, its diversity is still a huge step up from existing datasets used in testing 3DMMs.



(a) LSFM-global (b) LSFM-bespoke

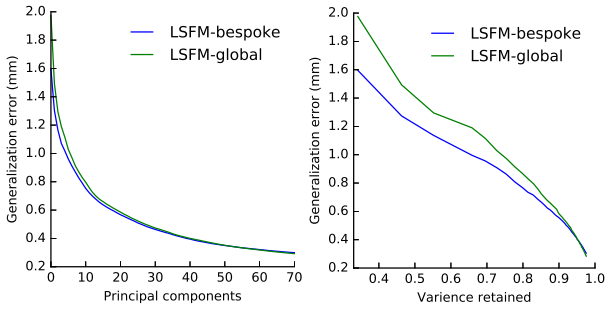
Figure 6: Compactness of the LSFM models.

6.4. Intrinsic Evaluation of LSFM models

Following common practice in the literature of statistical shape models, we evaluate the intrinsic characteristics of LSFM-global and LSFM-bespoke using *compactness*, *generalisation* and *specificity*, see e.g. [17, 14, 10]. We consider the 3D shapes of `MeIn3D` dataset after establishing dense correspondences, using our pipeline.

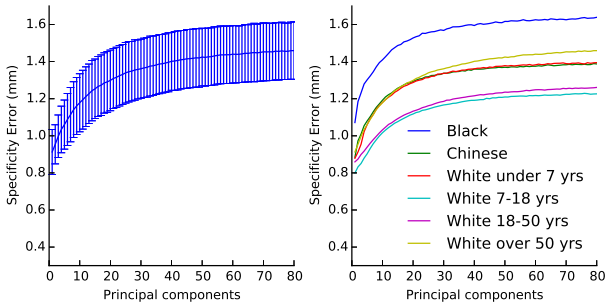
Figure 6 shows the **compactness** plots for the LSFM models. Compactness measures the percentage of variance of the training data that is explained by a model when certain number of principal components are retained. Note that in the case of the bespoke models, the training samples of the corresponding demographic group are only considered, which means that the total variance is different for every model. We observe that all trained models exhibit similar traits in the variance captured, although this naturally varies with the size of the training set in each case of the tailored models. Both global and bespoke LSFM models can be considered sufficiently compact; for example for all the models, as few as 40 principal components are able to explain more than 90% of the variance in the training set.

Figure 7 presents plots of model **generalisation**, which measures the ability of a model to represent novel instances of face shapes that are unseen during training. To compute the generalisation error of a model for a given number of principal components retained, we compute the per-vertex Euclidean distance between every sample of the test set $\mathbf{X}$ and its corresponding model projection $P(\mathbf{X})$, Eq. (2), and then take the average value over all vertices and all test samples. In order to derive an overall generalisation measure for LSFM-bespoke, for every test sample we use its demographic information and project on the subspace of the corresponding bespoke model and then compute an overall average error. We plot the generalisation errors with respect to both the number of principal components (Fig. 7a) and percentage of total variance retained (Fig. 7b). We observe that both LSFM-global and LSFM-bespoke are able



(a) (b)

Figure 7: Generalisation of the LSFM models, with respect to: (a) the number of principal components retained and (b) the portion of variance explained.



(a) LSFM-global (b) LSFM-bespoke

Figure 8: Specificity of the LSFM models.

to generalise well, since for even low number of components and total variance retained, they yield particularly low generalisation errors. Interestingly, we see in Fig. 7a that LSFM-bespoke achieves superior generalisation measures when compared to LSFM-global for an equivalent number of components for fewer than 60 components. After this stage the global model starts to outperform the specific models, which might attributed to the fact that many of the specific models are built from smaller cohorts of training data, and so run out of interesting statistical variance at an earlier stage. When changing the visualisation to one based on retained variance (Fig. 7b), we observe that the demographic-specific LSFM-bespoke model achieves better generalisation performance for the vast majority of values of retained variance.

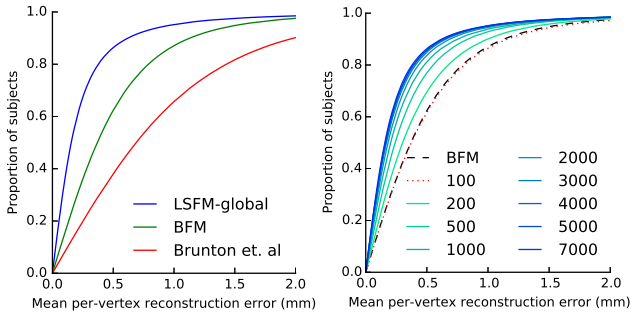
Figure 8 presents the **specificity** of the introduced models, which evaluate the validity of synthetic faces generated by a model. To measure this, we randomly synthesise 10,000 faces from each model for a fixed number of components and measure how close they are to the real faces of the test set. More precisely, for every random synthetic face, we find its nearest neighbor in the test set, in terms of minimum (over all samples of the test set) of the average (over all vertices) per-vertex distance. We record the mean of this distance over all samples as the specificity error. Figure 8a contains the specificity plot for LSFM-global (mean value as well as standard deviation bars), whereas Figure 8b contains the specificity plots for all models of LSFM-bespoke (mean values only; the standard deviation bars have been omitted for the sake of visualisation clarity). We observe that in all the cases, the specificity errors attain particularly low values, in the range of 1 to 1.6 mm, even for a relatively large number of principal components. This is a quantitative evidence that the synthetic faces generated by both global and bespoke LSFM models are realistic, which complements the qualitative observations of Section 6.1. Interestingly, Figure 8b suggests that specificity error is larger for models trained from smaller populations, as e.g. in the case of Black model. Apart from the lack of sufficient representative training data, this might also be attributed to the fact that the space of such models is more sparsely populated by training samples, so the nearest neighbor error tends to be larger, as compared to other models with more training data. Furthermore, it can be seen that the lowest specificity error comes from the White-7-18 model, which is trained on a large number of samples that lie on a smaller cluster of the space, leading to a highly specific model.

6.5. Fitting Application

In order to gauge the quality of the LSFM-global model in comparison with the state-of-the-art, we evaluate the performance of the models in a real-world fitting scenario. We compare with two publicly available morphable models of human faces in neutral expression, namely the *BFM model* [24, 25] and the PCA model of [14, 8], which will be hereafter referred to as *Brunton et al. model*. Note that for the sake of fairness towards the existing models, we do not consider the bespoke LSFM models in the fitting experiment, since these models use additional information related to demographics.

Note that for all versions of LSFM-global evaluated hereafter, we choose the number of principal components, so as to explain 99.5% of the training set variance. For BFM and Brunton et al. models, we use all the principal components, as given by the publicly available versions of these models.

To evaluate the fitting performance of every tested model, every mesh in the adopted test set is automatically annotated with facial landmarks using our technique outlined in Section 5.1. The same set of landmarks is manually placed on the mean faces of every model, and subsequently used to similarity-align them with every mesh of the test set. Similarly to [14, 32], a simple model fitting is employed, which consists of (1) searching for the nearest vertex in the test mesh to establish correspondences between that mesh and the model, (2) projecting the test mesh onto the model



(a) (b)

Figure 9: Cumulative error distributions of the per-vertex fitting error **(a)** between publicly available models **(b)** for LSFM models built from varied amounts of training data.

using Eq. (2). The per-vertex fitting error is then computed as the distance between every vertex of the test mesh and the nearest-neighbor vertex of the corresponding model-based fitting. Note that we use a simple fitting strategy to provide an appropriate framework to benchmark models against one another fairly.

Figure 9a compares the fitting performance of LSFM-global against BFM and Brunton et al. models, in terms of cumulative error distribution (CED) curves of per-vertex fitting errors. We observe that LSFM-global achieves exceptionally improved accuracy and robustness, as compared to the other two models. This is attributed to the larger training sample used. We also note that this is the first time that existing models are evaluated against a dataset containing a large variation in ethnicity and age. The significantly larger variability in the training set of LSFM-global allows it to generalise well to a much wider variety of faces than the more narrowly-focused existing models.

It is natural to question the **effect of varying the size of the training set** on 3DMM construction. To explore this, we repeat the above fitting experiment for different versions of the LSFM-global model, trained from varying numbers of samples. The results are visualised in the plots of Figure 9b. Clear improvements can be seen in model fitting performance for around one order of magnitude more data than is currently used, albeit with diminishing returns beyond a few thousand samples. We also note that even with only 100 samples, LSFM-global matches the performance of the BFM, which was trained on 200 samples. This can be attributed to the larger variability of the LSFM training set, demonstrating how crucial this is for building effective 3DMMs.

6.6. Age Classification from 3D shape

As a final evaluation, we use the unique traits of the `MeIn3D` dataset to compare the descriptive power of LSFM-global, BFM and Brunton et al. models in an age classification experiment. Following [31], we use the following age groups as classes for this experiment: 0-11, 12-21, 22-60, and over 60 years old. In more detail, we project all the face meshes of the training set onto each of the three models and use the corresponding shape vectors, $\boldsymbol{\alpha}$, to represent them, see Eq. (2). Using the demographic information of `MeIn3D` dataset, we train a Support Vector Machine classifier for each model, which maps the corresponding shape vectors to the four age groups.

	Precision	Recall	F-Score
LSFM-global	**0.80**	**0.81**	**0.80**
BFM	0.78	0.79	0.78
Brunton et al.	0.74	0.74	0.74

Table 1: Mean age classification scores.

To measure the classification accuracy, we project all samples from the test set onto the models and then use the classifier to predict the age bracket for the test subjects. This provides an application-oriented evaluation of the quality of the low-dimensional representation that each 3DMM provides for the large variety of faces present in LSFM. As can be seen in Table 1, the LSFM-global model outperformed existing models in precision and recall and f-score, correctly classifying the age of 80% of the subjects in the challenging test set.

7. Conclusions & future work

We have presented LSFM, the most powerful and statistically descriptive 3DMM ever constructed. By making both the LSFM software pipeline and models available, we help to usher in an exciting new era of large scale 3DMMs. We have demonstrated that our automatically constructed model comfortably outperforms existing state-of-the-art 3DMMs thanks to the sheer variety of facial appearance it was trained on, and further reported on how the size of 3D datasets impacts model performance. We have explored for the very first time the structure of the high-dimensional facial manifold, revealing how it is clustered by age and ethnicity variations, and demonstrated accurate age prediction from 3D shape alone. In future work, we will analyse in detail the qualities of the LSFM model, exploring what it can tell us about human face variation on the large scale, as well as exploring novel statistical methods for large-scale 3DMM construction.

Acknowledgments. J. Booth is funded by an EPSRC DTA from Imperial College London, and holds a Qualcomm Innovation Fellowship. A. Roussos is funded by the Great Ormond Street Hospital Childrens Charity (Face Value: W1037). The work of S. Zafeiriou was partially funded by the EPSRC project EP/J017787/1 (4D-FAB).

References

[1] J. Alabort-i Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 679–682, New York, NY, USA, 2014. ACM. 4

[2] O. Aldrian and W. A. Smith. Inverse rendering of faces with a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1080–1093, 2013. 1

[3] B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3d face recognition with a morphable model. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008. 1, 2

[4] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 2

[5] E. Antonakos, J. Alabort-i Medina, G. Tzimiropoulos, and S. Zafeiriou. Hog active appearance models. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 224–228. IEEE, 2014. 4

[6] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999. 1, 2

[7] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003. 1

[8] T. Bolkart, A. Brunton, A. Salazar, and S. Wuhrer. Website of statistical 3d shape models of human faces, 2013. http://statistical-face-models.mmci.uni-saarland.de/. 2, 7

[9] T. Bolkart and S. Wuhrer. 3D faces in motion: Fully automatic registration and statistical analysis. *Computer Vision and Image Understanding*, 131:100–115, 2015. 2, 3

[10] T. Bolkart and S. Wuhrer. A groupwise multilinear correspondence optimization for 3d faces. In *IEEE International Conference on Computer Vision (ICCV)*, 2015. 2, 6

[11] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. 2

[12] A. Brunton, T. Bolkart, and S. Wuhrer. Multilinear wavelets: A statistical shape space for human faces. In *European Conference on Computer Vision (ECCV)*, pages 297–312. Springer, 2014. 3

[13] A. Brunton, J. Lang, E. Dubois, and C. Shu. Wavelet model-based stereo for fast, robust face reconstruction. In *Canadian Conference on Computer and Robot Vision (CRV)*, pages 347–354, 2011. 2

[14] A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer. Review of statistical shape spaces for 3d data with comparative analysis for human faces. *Computer Vision and Image Understanding*, 128:1–17, 2014. 2, 3, 6, 7

[15] T. F. Cootes, G. J. Edwards, C. J. Taylor, et al. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001. 2

[16] D. Cosker, E. Krumhuber, and A. Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2296–2303. IEEE, 2011. 2, 4

[17] R. Davies, C. Taylor, et al. *Statistical models of shape: Optimisation and evaluation*. Springer Science & Business Media, 2008. 3, 6

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 3

[19] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 3

[20] V. Jain and E. G. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report*, 2010. 3

[21] I. Kemelmacher-Shlizerman. Internet based morphable model. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3256–3263. IEEE, 2013. 2

[22] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009. 4

[23] A. Patel and W. A. Smith. 3d morphable face models revisited. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1327–1334. IEEE, 2009. 2

[24] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference On*, pages 296–301. IEEE, 2009. 2, 7

[25] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. Website of basel face model, 2009. http://faces.cs.unibas.ch/bfm/. 2, 7

[26] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 397–403. IEEE, 2013. 3

[27] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto. Fully Automatic Expression-Invariant Face Correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014. 2

[28] F. C. Staal, A. J. Ponniah, F. Angullia, C. Ruff, M. J. Koudstaal, and D. Dunaway. Describing Crouzon and Pfeiffer syndrome based on principal component analysis. *Journal of Cranio-Maxillofacial Surgery*, 43(4):528–536, 2015. 1

[29] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008. 3, 6

[30] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 426–433. ACM, 2005. 2

[31] J.-G. Wang, E. Sung, and W.-Y. Yau. Active Learning for Solving the Incomplete Data Problem in Facial Age Classification by the Furthest Nearest-Neighbor Criterion. *IEEE Trans. Image Processing (TIP)*, 20(7):2049–2062, 2011. 8

[32] S. Zulqarnain Gilani, F. Shafait, and A. Mian. Shape-based automatic detection of a large number of 3d facial landmarks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4639–4648, 2015. 7