A 3D Morphable Model of Craniofacial Shape and Texture Variation

Hang Dai, Nick Pears and William Smith
Department of Computer Science, University of York, UK
{hd816,nick.pears,william.smith}@york.ac.uk

Christian Duncan
Alder Hey Hospital, Liverpool, UK
Christian.Duncan@alderhey.nhs.uk

Abstract

We present a fully automatic pipeline to train 3D Morphable Models (3DMMs), with contributions in pose normalisation, dense correspondence using both shape and texture information, and high quality, high resolution texture mapping. We propose a dense correspondence system, combining a hierarchical parts-based template morphing framework in the shape channel and a refining optical flow in the texture channel. The texture map is generated using raw texture images from five views. We employ a pixel-embedding method to maintain the texture map at the same high resolution as the raw texture images, rather than using per-vertex color maps. The high quality texture map is then used for statistical texture modelling. The Headspace dataset used for training includes demographic information about each subject, allowing for the construction of both global 3DMMs and models tailored for specific gender and age groups. We build both global craniofacial 3DMMs and demographic sub-population 3DMMs from more than 1200 distinct identities. To our knowledge, we present the first public 3DMM of the full human head in both shape and texture: the Liverpool-York Head Model. Furthermore, we analyse the 3DMMs in terms of a range of performance metrics. Our evaluations reveal that the training pipeline constructs state-of-the-art models.

1. Introduction

Morphable models of 3D shape have many applications in creative media, medical image analysis and biometrics. Traditional model building pipelines have used manual landmarking to initialise surface alignment. However, this is extremely time-consuming and laborious for large-scale datasets. Here we present a fully automatic approach and apply it to a large dataset of 3D images of the human head, thus generating the first public shape-and-texture 3D morphable model (3DMM) of the full human head: the Liverpool-York Head Model (LYHM).

A full head model opens up new opportunities not afforded by existing 3D face models [21]. For example: (i) we can reconstruct the most likely full head shape from partial data. This allows the modelling of different hair styles onto a correctly proportioned cranium (cranial shape usually cannot be directly observed). (ii) any part of the head shape can be matched against population norms thus facilitating tools for craniofacial diagnoses and surgical intervention evaluation, and (iii) the ability to complete partial views of the head may be exploited in biometric applications.

Contributions: (i) We propose a fully automatic pipeline to build 3DMMs, with contributions in pose normalisation and dense correspondence - in particular we propose a *hierarchical parts-based* template morphing framework with optical flow correspondence refinement. (ii) We propose a high quality texture map from several views of cameras for the morphed template and use the texture map for texture modelling. (iii) We build both global craniofacial 3DMMs and demographic sub-population 3DMMs from more than 1200 distinct identities in the *Headspace* dataset [1, 16] and we will make them publicly available [1]. Fig. 1 illustrates the principal components of the global and child model. To our best knowledge, our models are the first public shape-and-texture craniofacial 3DMMs of the full human head. (iv) We demonstrate *flexibility modes* of our model such that, when given a fixed face shape, we compute the range of possible cranial shapes and vice-versa.

After discussing literature, we describe our dataset and 3DMM training pipeline, while the following sections present details of the proposed methods. Section 8 provides an extensive evaluation of the 3DMMs, and a final section is used for conclusions.

2. Related work

In the late 1990s, Blanz and Vetter built a 3DMM from 3D face scans [7] and employed it in 2D face recognition [8]. Two hundred scans were used to build the model (young adults, 100 males and 100 females). Dense correspondences were computed using optical flow with an energy dependent on both shape and texture.

There are very few publicly available morphable models

[1] https://www-users.cs.york.ac.uk/~nep/research/LYHM/

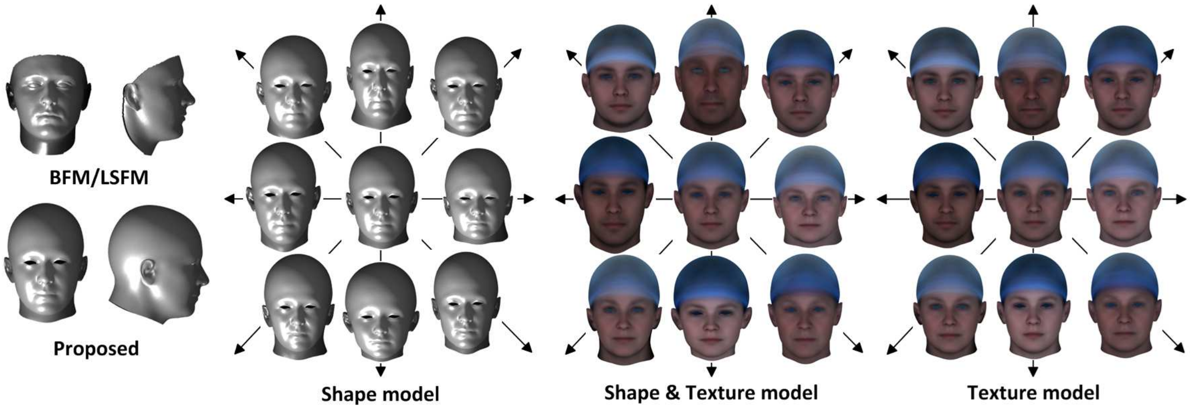


Figure 1. The proposed Liverpool-York Head Model. 1st block - shape compared to the Basel Face Model (BFM) [25] and Large Scale Face Model (LSFM) [10]; 2nd-4th blocks: the central head is the mean and the first four principal modes of variation are shown.

of the human face and, to our knowledge, none that include the full cranium. The Basel Face Model (BFM) is the most well-known and widely used and was developed by Paysan et al. [25]. Again 200 scans were used, but the method of determining corresponding points was improved. Instead of optical flow, a set of hand-labelled feature points is marked on each of the 200 training scans. The corresponding points on a template mesh are known, which is then morphed onto the training scan using under-constrained per-vertex affine transformations, which are constrained by regularisation across neighbouring points [3]. The technique is known as optimal-step Non-Rigid Iterative Closest Points (NICP). In the method of De Smet et al. [15], they found the segmentation automatically by clustering the vertices, which is based on features derived from their displacements. In order to address the potential discontinuities at the boundaries of the segments, they smoothly weight the segments to obtain regionalised basis functions for the training data. Recently Booth et al. [10] built a Large Scale Facial Model (LSFM), using the same NICP template morphing approach, followed by Generalised Procrustes Analysis (GPA) for alignment, and Principal Component Analysis (PCA) for the model construction.

Existing 3DMMs employ either deformable template methods [25, 15, 10, 28, 9, 27] or optical flow [7, 13, 24] to establish dense correspondence. The former has the advantage of good correspondence over most of the mesh. However, it often ends up with bad correspondences in local regions. In contrast, optical flow requires a good initialisation and the weighting between shape and texture costs is ad hoc. However, it is successful when employed in a refinement step to make small adjustments to local region correspondence. The small displacement version of optical flow is well suited to this.

Other deformable template methods could be used to build 3DMMs and include the work of Li et al. [22]. Their global correspondence optimization method solves simultaneously for both the deformation parameters as well as the correspondence positions. Myronenko et al. [23] consider the alignment of two point sets as a probability density estimation using Expectation-Maximisation (EM) and they call the method Coherent Point Drift (CPD). This remains a highly competitive template morphing algorithm.

Template morphing methods need an automatic initialisation to bring them within the convergence basin of the global minimum of alignment and morphing. To this end, Active Appearance Models (AAMs) [12] and elastic graph matching [35] are the classic approaches of facial landmark and pose estimation. Many improvements over AAM have been proposed [32, 17]. Recent work has focused on global spatial models built on top of local part detectors, sometimes known as Constrained Local Models (CLMs) [30, 36]. Zhu and Ramanan [37] use a tree structured part model of the face, which both detects faces and locates facial landmarks. One of the major advantages of their approach is that it can handle extreme head poses even at relatively low image resolutions, and we exploit these qualities directly in our model building pipeline.

For optical flow, the variational framework [19], together with coarse-to-fine refinement [4], is widely used in optical flow estimation [11]. On the Middlebury optical flow evaluation website [5], almost all top-ranked methods adopt this scheme. Wedel et al. [34] proposed a structure-texture decomposition method to reduce the discrepancy between two frames caused by illumination change. Lempitsky et al. [20] computed the matching cost only using high frequency components.

3. Database Overview

This work uses the *Headspace* dataset of approximately 1.5K subjects, all of whom are wearing tight fitting latex caps [1, 16]. We excluded some subjects due to either ex-

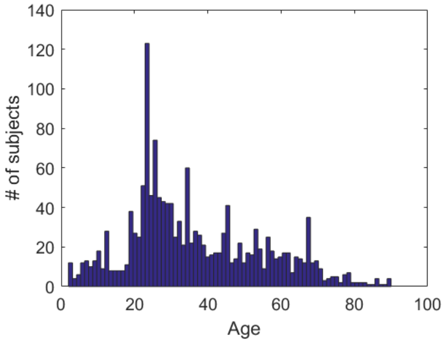


Figure 2. Age distribution of subjects

cessive hair bulge or declared craniofacial surgery/trauma condition and employed around 1.2K subjects in the global model build. A *3dMD* five-camera system was employed, creating a 3D triangular surface for each subject composed of typically 180K vertices joined into typically 360K triangles, along with a high-resolution texture image.

FRGC [26], MeIn3D [10] and FaceBase [29] are large 3D face datasets, but the full cranium is not included. In contrast, the Size China project [6] does include the cranium, but it is a commercial dataset and 3DMMs have not been built. Our Headspace dataset is the only full head dataset for academic use and will be made publicly available (both the trained models and raw data).

Subjects were imaged at Alder Hey Hospital, Liverpool, and provided metadata including gender, age, ethnicity, eye color, and any craniofacial surgery/trauma condition. This information allows for the construction of models for targeted populations, such as within a defined age range or gender. The dataset covers a wide variety of age (see Fig.2), and is balanced in gender.

4. Model construction

Our 3DMM training pipeline, illustrated in Fig.3, operates in three main functional blocks. These are outlined below and detailed in the following three sections.

Data preprocessing: We use automatic 2D landmarking and map to 3D using the known 2D-to-3D registration supplied by the 3D camera system. These 3D landmarks can then be used for pose normalisation.

Dense correspondence: A collection of 3D scans are reparametrised into a form where each scan has the same number of points joined into a triangulation that is shared across all scans. Furthermore, the semantic or anatomical meaning of each point is shared across the collection, as defined by a template mesh.

Alignment and statistical modelling: The collection of scans in dense correspondence are subjected to Generalised Procrustes Analysis (GPA) to remove similarity effects (rotation, translation, scale), leaving only shape information. The processed meshes are statistically analysed, typically with Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes. This allows for the generation of novel shape instances.

5. Data preprocessing

Data preprocessing includes 2D landmarking, projection to 3D landmarks and pose normalisation.

5.1. Automatic 2D/3D facial landmarking

We use the method of Zhu and Ramanan [37] to localise facial landmarks on the texture channel of each 3D image. This 2D image contains all 5 viewpoints of the capture system and usually two face detections are found, 15-45 degrees yaw from frontal pose, corresponding to the left and right side of the face. Detected 2D points are in a tree structure and are projected to 3D using OBJ texture coordinates.

5.2. Pose normalisation

Each face detection employs one of thirteen tree models [37] and we automatically learn how to orientate each of these to frontal pose, based on their 3D structure. To do this, we apply GPA to each collection of 3D trees (11 of the 13 models are used by the dataset) and find the nearest-to-mean tree shape in a scale-normalised setting. We then apply a 3D face landmarker [14] to the 3D data of the nearest-to-mean tree shape (11 of these), which generates a set of 14 landmarks with clear semantic meaning. Finally, we find the alignment that moves the symmetry plane of these 14 landmarks to the Y-Z plane with the nasion above the subnasale (larger Y coordinate) and at the same Z-coordinate, in order to normalise the tilt (X rotation). To complete the training phase, the mean 3D tree points for each of the 13 trees are then carried into this canonical frontal pose using the same rotation, and are used as reference points for the frontal pose normalisation of the 3D trees.

In around 1% of the dataset, only one tree is detected and that is used for pose normalisation, and in the rest 2-3 images are detected. In the cases where 3 trees are detected, the lowest scoring tree is always false positive and can be discarded. For the remaining two trees, a weighted combination of the two rotations is computed using quaternions, where the weighting is based on the mean Euclidean error to the mean tree, in the appropriate tree component.

6. Dense correspondence

In this section, we propose a new template morphing framework along with optical flow correspondence refinement. Both shape information and texture information are used for correspondence establishment.

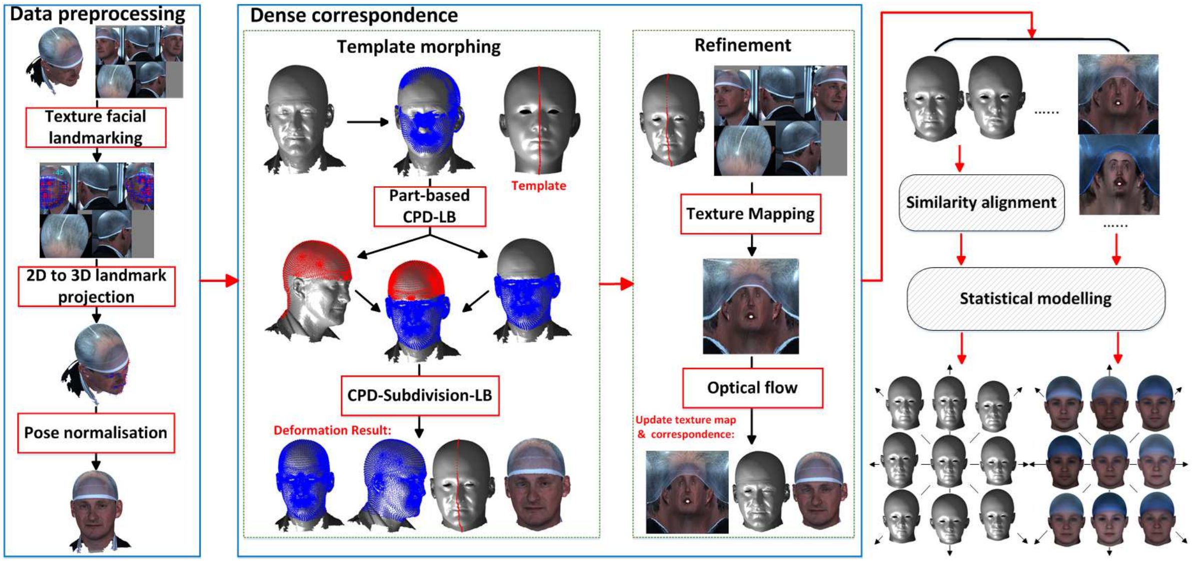


Figure 3. The fully-automatic 3DMM training pipeline used to construct the Liverpool-York Head Model. Left block shows landmarking and pose normalisation, central block shows dense correpondence establishment, right block shows statistical modelling.

6.1. Template morphing

Compared to NICP, CPD can avoid tangential point sliding and over-fitting, resulting in a good fit and good symmetry preservation. But standard CPD fails to resolve the under-fitting problem, where convergence terminates before good correspondence is established over the ears. This is because there are relatively few points on the ears and the facial/cranial points dominate. To avoid under-fitting, we propose a two-stage approach that consists of CPD followed by mutual nearest neighbour projection regularised by the Laplace-Beltrami (LB) operator. This two-stage approach is then applied within a parts-based framework.

Two-stage CPD-LB approach: When CPD has deformed the template close to the scan, point projection is required to eliminate any (normal) shape distance error. Point projection is a potentially fragile process. If the scan data is incomplete or noisy then projecting vertices from the deformed template to their nearest vertex or surface position on the scan may cause large artefacts. We overcome this by treating the projection operation as a mesh editing problem with two ingredients. First, position constraints are provided by those vertices with mutual nearest neighbours between the deformed template and raw scan. Using mutual nearest neighbours reduces sensitivity to missing data. Second, regularisation constraints are provided by the LB operator which acts to retain the local structure of the mesh.

Thus we optimise two cost functions in a sequential manner. The Expectation-Maximization (EM) algorithm is used to solve the CPD cost function $\mathbf{E}_{CPD}$. For this cost function, we refer to [23]. For the second stage, we write the LB mesh editing problem as a linear system of equations. Given the vertices of a scan stored in the matrix $\mathbf{X}_{\text{scan}} \in \mathbb{R}^{n\times 3}$ and the deformed template obtained by CPD whose vertices are stored in the matrix $\mathbf{X}_{\text{deformed}} \in \mathbb{R}^{p\times 3}$, we define the selection matrices $\mathbf{S}_1 \in [0,1]^{m\times p}$ and $\mathbf{S}_2 \in [0,1]^{m\times n}$ as those that select the m vertices with mutual nearest neighbours from deformed template and scan respectively. This linear system can be written as:

$$\begin{pmatrix} \lambda\mathbf{L} \\ \mathbf{S}_1 \end{pmatrix} \mathbf{X}_{\text{proj}} = \begin{pmatrix} \lambda\mathbf{L}\mathbf{X}_{\text{deformed}} \\ \mathbf{S}_2\mathbf{X}_{\text{scan}} \end{pmatrix} \quad (1)$$

where $\mathbf{L} \in \mathbb{R}^{p\times p}$ is the cotangent Laplacian approximation to the LB operator and $\mathbf{X}_{\text{proj}} \in \mathbb{R}^{p\times 3}$ are the projected vertex positions that we wish to solve for. The parameter λ weights the relative influence of the position and regularisation constraints, effectively determining the 'stiffness' of the projection. As $\lambda \to 0$, the projection tends towards nearest neighbour projection. As $\lambda \to \infty$, the deformed template will only be allowed to rigidly transform.

Hierarchical parts-based CPD-LB framework: Our template fitting algorithm, using temporary splitting of the template into parts and multiple applications of the two-stage CPD-LB deformation process is illustrated in the central panel of Fig.3 and consists of the following six steps:

1. CPD-affine deforms the full-head template to fit the pose-normalised input until it converges to a global minimum according to the convergence criterion.
2. The template is then divided into two parts: cranial (red part in Fig.3) and facial (blue part in Fig.3) using a predetermined manual mesh segmentation. The input mesh is also divided into two corresponding parts using nearest neighbours.

3. The segmentation in step 2 releases the two individual parts from the global minimum in step 1 and CPD affine continues on the two parts separately to re-establish convergence. In practice, the cranial part is already very close to its global minimum, but the facial part continues to deform over many more iterations.
4. After convergence, the two parts of the template are updated using LB-regularised projection; however, there may exist a gap or an overlap between the two parts, due to the separate deformation processes.
5. The deformed full template from step 1 is now used to deform towards the two deformed parts-based templates from the previous step. Note that we now have a known one-to-one correspondence between the full template and the two template parts. Under these circumstances CPD-nonrigid performs well and is used.
6. After CPD-nonrigid converges, a final stage of LB-regularised projection is used to give the final morphed template of the full head, which is devoid of any overlaps or discontinuities that appear between the two separate parts in step 4 and is superior to the initial deformation from step 1 (standard CPD-affine).
7. Morphed template mesh subdivision followed by LB-regularised projection is used to upscale resolution.

6.2. Texture map

It is preferable to store texture information in a UV space texture map where resolution is unconstrained rather than store only per-vertex colours where resolution is limited by mesh resolution. To do so requires the texture information from each scan to be transformed into a standard UV texture space for which the embedding of the template is known.

After template morphing, the deformed template has the same number of points joined into a triangulation that is shared across all scans. Thus in UV coordinates, UV faces of the morphed template are shared with the template. Given the morphed vertex positions $\mathbf{X}_{\text{morphed}} \in \mathbb{R}^{p\times 3}$ from template morphing stage, we can first compute the UV coordinates for each point of the morphed template in original texture image:

$$[\mathbf{u},\mathbf{v}] = g(\mathbf{S}_3\mathbf{X}_{\text{scan}}) \tag{2}$$

where $\mathbf{S}_3 \in \mathbb{R}^{p\times n}$ is the selection matrix that select the p vertices with nearest neighbours from morphed template $\mathbf{X}_{\text{morphed}}$ to the scan $\mathbf{X}_{\text{scan}}$ and g is the UV coordinates mapping from the raw mesh to texture image. The UV coordinates mapping from texture image to raw mesh is a surjection but not an injection. Thus the points from the raw mesh may have several sets of UV coordinates $[\mathbf{u},\mathbf{v}]_{cand}$ in the texture image, depending on the number of capture viewpoints. To overcome this, we minimise the face area of UV face $\mathbf{UV}^k_{\text{faces}}$ to find the exact UV coordinates:

$$[\mathbf{u}_i^*,\mathbf{v}_i^*] = \underset{\mathbf{u}_i,\mathbf{v}_i}{\arg\min}\; h([\mathbf{u}_i,\mathbf{v}_i],[\mathbf{u}_{j1},\mathbf{v}_{j1}],[\mathbf{u}_{j2},\mathbf{v}_{j2}]) \tag{3}$$

where $[\mathbf{u}_i,\mathbf{v}_i] \in [\mathbf{u},\mathbf{v}]_{cand}$, $[\mathbf{u}_{j1},\mathbf{v}_{j1}]$ and $[\mathbf{u}_{j2},\mathbf{v}_{j2}] \in \mathbf{UV}^k_{face}$, and h is the face area computation.

Due to multiple capture viewpoints, the minimization of face area is not enough to guarantee that all the UV coordinates are in the image of same view (see Fig.4 (2)). To overcome this, we employ affine transformations to refine the UV coordinates. If the UV coordinates in one mesh face are placed in different views, we compute the affine transformation $\mathbf{T}$ from its adjacent face to the corresponding face in the template UV faces. Then its corresponding face is inverted by $\mathbf{T}^{-1}$ to find the correct position (see blue point in Fig.4 (1)). The outcome of affine transformation refinement is shown in Fig.4 (2), (3).

6.3. Optical flow and correspondence refinement

Optical flow algorithms are based on the assumption that objects in image sequences $I(x;y;t)$ retain their brightnesses as they move across the image at a velocity $(v_x;v_y)^T$ [5]. A unique solution for both components of $(v_x;v_y)^T$ can be obtained if v is assumed to be constant over each neighborhood $R(x_0;y_0)$, and the following energy function is minimized in each point $(x_0;y_0)$:

$$\mathbf{E} = \sum_{x,y\in R}\left(v_x\frac{dI(x,y)}{dx} + v_y\frac{dI(x,y)}{dy} + \Delta I(x,y)\right) \tag{4}$$

We use a 10×10 pixel neighborhood $R(x_0;y_0)$. To deal with small displacements, we use the three image RGB channels, two image gradient channels and the surface normal along with a coarse-to-fine refinement framework to solve the system:

$$\mathbf{E}_{\text{sum}} = \mathbf{E}_{\text{RGB}} + \mathbf{E}^{x,y}_{\text{grad}} + \mathbf{E}_{\text{normal}} \tag{5}$$

In order to refine the dense correspondence in shape, we compute the 2D UV coordinates to 3D mesh projection for the warped image. We treat this projection as the interpolation from the original image to the warped image. Given the 2D UV coordinates to 3D points position $(\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i)$ projection for the original image, we can infer the interpolation relation f from:

$$f(u_i,v_i) \rightarrow (\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i)_{i\in UV^k_{face}} \tag{6}$$

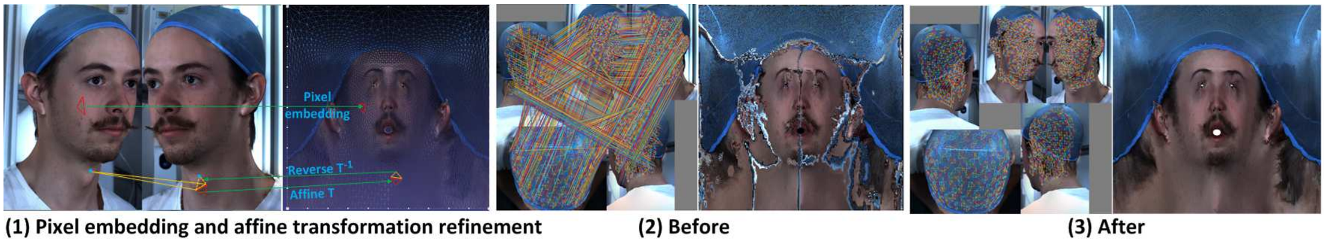
So, for the warped image, the 3D point position is:

$$(\mathbf{a},\mathbf{b},\mathbf{c}) = f_{i\in UV^k_{face}}(u_i+\Delta u, v_i+\Delta v). \tag{7}$$

When the correspondence is updated, a new texture map can be generated from the updated correspondence.

7. Alignment and statistical modelling

We use GPA to align our deformed templates before applying PCA-based statistical modelling. This generates a



(1) Pixel embedding and affine transformation refinement (2) Before (3) After

Figure 4. Texture map flow: (1) Pixel embedding, (2) UV coordinates of mesh facets before affine refinement, here we are unable to use facets that straddle viewpoints, (3) After affine refinement, all facet UV coordinates can be used.

3DMM as a linear basis of shapes, allowing for the generation of novel shape instances. Note that we may select all of the full head template vertices for this modelling, or any subset. For example, we select the cranial vertices when we build models to analyse cranial medical conditions.

In many applications, vertex resolution is not uniform across the mesh. For example, we may use more vertices to express detail around facial features of high curvature. However, standard PCA attributes the same weight to all points in its covariance analysis, making it difficult to handle shape variance modelling over such meshes. To counter this, we employ weighted PCA in our statistical modelling.

7.1. Weighted PCA

The first step of Weighted PCA (WPCA) is to construct the symmetric positive-definite matrix $\mathbf{Y^TMY}$ for $\mathbf{Y} = \mathbf{X} - \overline{\mathbf{X}}$. The first k eigenvectors $\mathbf{V_i^k}$ solving

$$\mathbf{Y^TMYV_i^k} = \lambda_\mathbf{i}\mathbf{V_i^k} \tag{8}$$

are computed. For the entries of the mass matrix $\mathbf{M}$, we refer to [33]. From [18], the relation between the left singular vector $\mathbf{U_i^k}$ and the right singular vector $\mathbf{V_i^k}$ of SVD decomposition is $\mathbf{U_i^k} = (\frac{\mathbf{1}}{\sigma_\mathbf{i}^\mathbf{k}})\mathbf{M}^{\frac{1}{2}}\mathbf{YV_i^k}$. For $1 \leq i \leq k$, we infer that the weighted PCA $\boldsymbol{\Psi}_\mathbf{i}^\mathbf{k}$ is obtained by multiplying the eigenvectors with $\mathbf{Y}$ and normalisation(w.r.t$\|.\|$) of the result:

$$\boldsymbol{\Psi}_\mathbf{i}^\mathbf{k} = \mathbf{M}^{-\frac{1}{2}}\mathbf{U_i^k} = (\frac{\mathbf{1}}{\sigma_\mathbf{i}^\mathbf{k}})\mathbf{YV_i^k} \tag{9}$$

7.2. Reconstruction

In an analogous fashion to Euclidean PCA, WPCA can also be used for reconstructing shapes.

Complete data reconstruction: Given the average $\overline{\mathbf{X}}$, the basis matrix $\mathbf{U_i^k}$, the mass matrix $\mathbf{M}$ and some (possibly unseen) shape $\mathbf{X} \in \mathbb{R}^{3n}$, we first compute the zero mean $\mathbf{U} = \mathbf{X} - \overline{\mathbf{X}}$. Then $\mathbf{U}$ is projected to WPCA space:

$$\mathbf{P^k} : \mathbf{U} \rightarrow \mathbf{P^k}\boldsymbol{\Psi}^\mathbf{k} = \mathbf{P^k}\mathbf{M}^{-\frac{1}{2}}\mathbf{U_i^k} \tag{10}$$

Finally we are able to reconstruct $\mathbf{X}$ by:

$$\mathbf{X} = \overline{\mathbf{X}} + \mathbf{P^k}\boldsymbol{\Psi}^\mathbf{k} = \overline{\mathbf{X}} + \mathbf{P^k}\mathbf{M}^{-\frac{1}{2}}\mathbf{U_i^k} \tag{11}$$

Incomplete data reconstruction: Given partial data, we can divide the shape components into two parts: one for the partial data $\boldsymbol{\Psi}_a$ and the other for missing data $\boldsymbol{\Psi}_b$ [2]. The zero mean for the partial data is $\mathbf{U_a} = \mathbf{X_a} - \overline{\mathbf{X}}_\mathbf{a}$:

$$\mathbf{P} : \mathbf{U_a} \rightarrow \mathbf{P}\boldsymbol{\Psi}_\mathbf{a} = \mathbf{P_a}\mathbf{M}^{-\frac{1}{2}}\mathbf{U_a} \tag{12}$$

Then $\mathbf{P}$ can be used to reconstruct the complete data. The aim is to model the remaining flexibility of the variable points $\mathbf{X}_a$ when the fixed points $\mathbf{X}_b$. The deformation is give as $\mathbf{P}\boldsymbol{\Psi}_a$ and $\mathbf{P}\boldsymbol{\Psi}_b$. We formulate the aim as a constrained maximization problem:

$$\max \mathbf{P}\boldsymbol{\Psi}_a\boldsymbol{\Psi}_a^T\mathbf{P}^T, \text{ subject to } \mathbf{P}\boldsymbol{\Psi}_b\boldsymbol{\Psi}_b^T\mathbf{P}^T = c \tag{13}$$

where c quantifies the amount of change allowed in the fixed shape principle components. Introducing a Lagrangian multiplier μ and differentiating with respect to $\mathbf{P}$ leads to the eigenvalue decomposition of:

$$\mathbf{P}\boldsymbol{\Psi}_a\boldsymbol{\Psi}_a^T = \mu\mathbf{P}\boldsymbol{\Psi}_b\boldsymbol{\Psi}_b^T \tag{14}$$

where for each parameter in $\mathbf{P}$, $\|\mathbf{P}_i\boldsymbol{\Psi}_b\|^2 = \mathbf{P}_i\boldsymbol{\Psi}_b\boldsymbol{\Psi}_b^T\mathbf{P}_i^T$.

8. Results

We select 1212 individuals (606 males and 606 females) to derive our global 3DMM by applying our fully-automatic 3DMM construction pipeline. Subpopulations are employed to build gender-specific models, LYHM-male, LYHM-female, and six age-specific models (LYHM-age-X), focusing on demographic-specific shape variation.

8.1. Correspondence comparison

Qualitative evaluation: The proposed method is compared with NICP [3], LSFM pipeline [10], Li's method [22] and standard CPD [23]. Expression rendering with texture is used to validate correspondence accuracy. Fig. 5 shows that the proposed method is qualitatively superior to other methods with respect to accuracy of the symmetry line, correspondence accuracy, and shape preservation relative to the input scan. Note that, the symmetry line slides to one side on the cranium for several methods due to over-fitting.

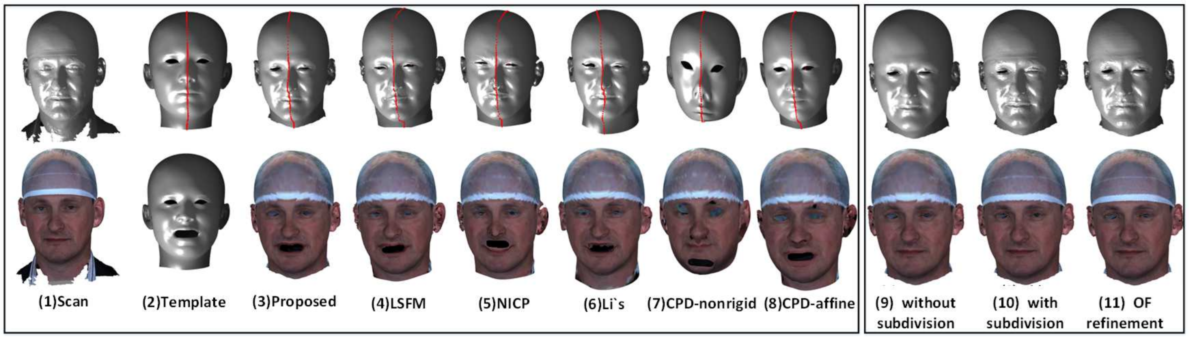


Figure 5. Left box - deformation results and their expression rendering with texture; Right box - subdivision and optical flow (OF).

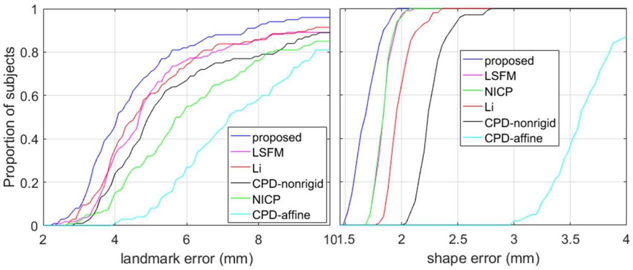


Figure 6. Proportion of subjects with $<$ Euclidean distance error: left-landmark error, right-shape error. The higher, the better.

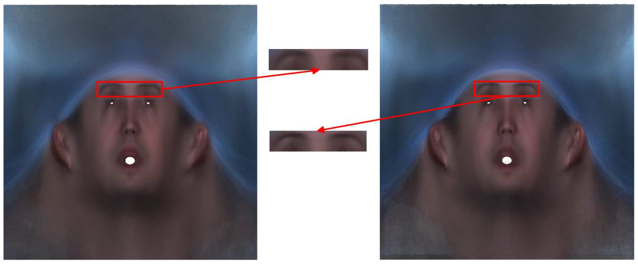


Figure 7. Average gradient magnitude of pre-OF (left, 4.663) and post-OF (right, **5.617**). The higher, the better (sharper).

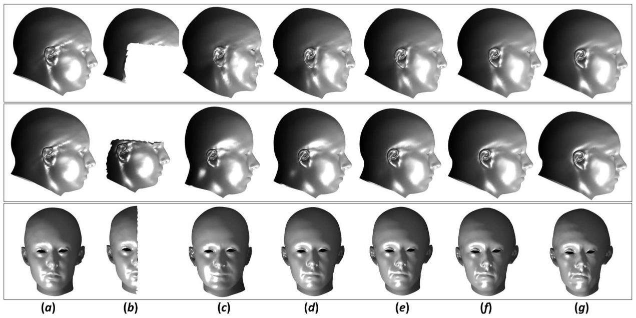


Figure 8. The flexibility of reconstruction from incomplete data: (a) example; (b) crop (incomplete data); (c)-(g) are the flexibility modes, and (e) is the mean

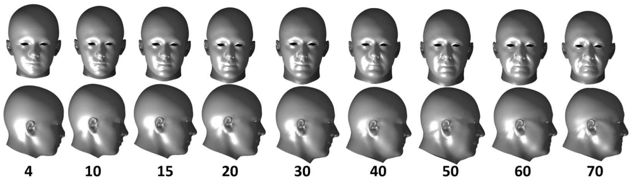


Figure 9. Age regression from 4 years to 70 years

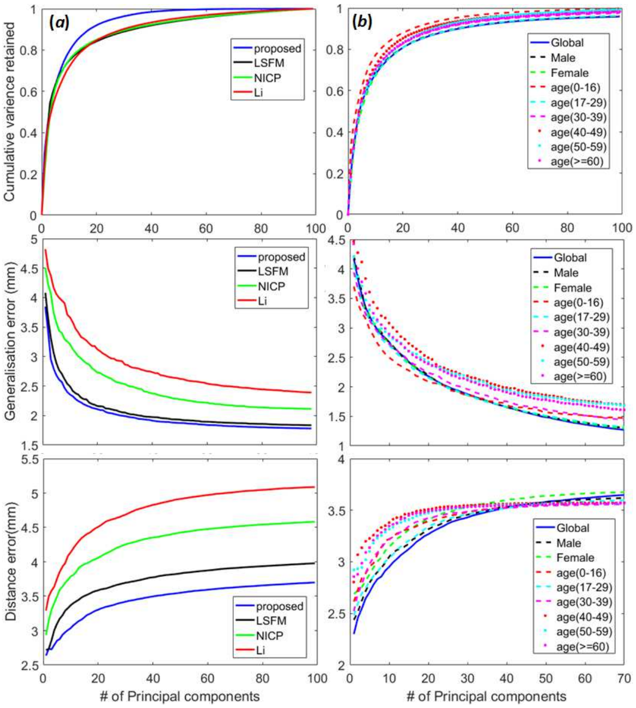


Figure 10. Top: compactness; Center: generalisation; Bottom: specificity. Left column - (a) different methods. Right column - (b) demographic-specific 3DMMs.

NICP overfits in both the eye and mouth regions. Although, the (normal) shape error is low, the registered mouth is not in the correct position. Li's method also overfits in the ear region. In contrast, CPD underfits that region in both affine and nonrigid versions. Fig. 5 (right box) shows that the proposed method with subdivision improves the quality of texture and captures more shape detail. But it also causes over-fitting in the eye region. The texture map with optical flow improves that by tuning the correspondence thus making the texture sharper, as exemplified by the edge of the cap. It also keeps more shape detail (see Fig. 5 (11)).

Quantitative evaluation: We use 14 manual facial landmarks over 100 subjects to measure landmark error of each template deformation method. As shown in Fig. 6-left, 80% of landmark errors are less than 6 mm for our method. Fig. 6-right shows that 60% of shape errors from our method are under 1.7 mm. The shape error is computed by measuring the nearest point distance from deformed template to raw scan. Overall, the proposed method outperforms all others.

8.2. Optical flow refinement

We warp all texture maps to the average using optical flow (OF). As shown in Fig. 7, the optical flow corrects small misalignments of features such as the eyebrows, chin and forehead. We computed image sharpness using average gradient magnitude averaged over 72 pre-OF and post-OF images. The 1st row of Fig. 5 (10) (11) demonstrates the outcome of the proposed correspondence refinement. The optical flow refinement clearly improves the correspondence in the eye regions, yet retains the shape-texture detail.

8.3. Reconstruction and age regression

The task of reconstruction of an unseen example with missing parts can validate the correspondence quality, model fitting and generalisation ability. In the first row of Fig. 8, we fix the shape of the cranium and reconstruct the full head from that shape, while permitting the shape of the face to vary. Here we found that most variation occurs over the chin region. The second row of Fig. 8 is the reconstruction from the face only, and we note that the principal variation in reconstructed cranium is from high/short to low/long. This offers a way to augment existing face models with a cranium. Reconstruction from one side of the sagittal symmetry plane demonstrates asymmetrical variation of the head. This application can aid shape-from-shading using 2D profile images to predict the 3D shape of the self-occluded half of the face.

Model regression can validate the model's generalisation ability on some property (eg. age, gender). Fig. 9 demonstrates a sequence of 3D images generated from age regression of the proposed model, ranging from 4 to 70 years. Note that the cranium develops before the face in children, and the shape of the cranium is not changing in adults.

8.4. Model evaluation

For quantitative model evaluation, Styner et al [31] give detailed descriptions of three metrics: compactness, generalisation and specificity. The compactness of the model describes the number of parameters required to express some fraction of the variance in the training set, fewer is better. We use the same subjects as in Sec.8.1 to build 3DMMs. As can be from Fig.10 (a), the proposed method has better compactness than other methods. The proposed method has the lowest generalisation error, which implies that proposed method has the best performance in describing unseen examples. LSFM [10] is comparable with the proposed method in generalisation ability. Specificity measures how well a model is able to generate instances that are similar to real data. The proposed method has the lowest distance error, which implies that the proposed method is best at generating instances close to real data.

In Fig.10 (b), LYHM-global is built using 1212 subjects along with several demographic-specific 3DMMs. The LYHM-(age 0-16) is the most compact 3DMM. When up to 30 principle components are used, this also has the best generalisation ability. When more than 30 principle components are used, LYHM-global, LYHM-male and LYHM-female are superior to others in generalisation. The LYHM-male is best in specificity when up to 30 PCs are used.

9. Conclusion

We proposed a fully-automatic 3DMM training pipeline and used it to build the first shape-texture 3DMM of the full head. The correspondence framework avoids over-fitting and under-fitting in template morphing, and captures high quality texture in a refinement stage. The proposed 3DMMs have a powerful ability in reconstruction of incomplete data and model regression to observe the influence of age on craniofacial growth. The flexibility of reconstruction from incomplete craniofacial data helps in many computer vision applications.

Acknowledgements

We thank Google Faculty Awards, the Royal Academy of Engineering and the Leverhulme Trust for their support. Headspace data collection was supported by QIDIS from the National Commissioning Group. We also thank Rachel Armstrong, Headspace project coordinator.

References

[1] Alder hey headspace project. http://www.alderhey.nhs.uk/departments/craniofacial/headspace-project/. 1, 2

[2] T. Albrecht, R. Knothe, and T. Vetter. Modeling the remaining flexibility of partially fixed statistical shape models. In *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 160–169, 2008. 6

[3] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid icp algorithms for surface registration. In *Proceedings of CVPR.*, pages 1–8, 2007. 2, 6

[4] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989. 2

[5] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. 2, 5

[6] R. Ball. Sizechina: A 3d anthropometric survey of the chinese head. 2011. 3

[7] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of Computer graphics and interactive techniques*, pages 187–194, 1999. 1, 2

[8] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1063–1074, 2003. 1

[9] F. Bogo, J. Romero, M. Loper, and M. J. Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of CVPR*, pages 3794–3801, 2014. 2

[10] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of CVPR*, pages 5543–5552, 2016. 2, 3, 6, 8

[11] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proceedings of ICCV*, volume 1, pages 749–755, 2005. 2

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001. 2

[13] D. Cosker, E. Krumhuber, and A. Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling. In *Proceedings of ICCV*, pages 2296–2303. IEEE, 2011. 2

[14] C. Creusot, N. E. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *Int. Journ. Computer Vision*, (1):146–179, 2013. 3

[15] M. De Smet and L. Van Gool. Optimal regions for linear model-based 3d face reconstruction. In *Proceedings of Asian Conference on Computer Vision*, pages 276–289, 2010. 2

[16] C. Duncan, H. Dai, N. Pears, and W. Smith. A novel shape based outcomes analysis tool for craniofacial surgeons. In *17th Biennial Congress of the International Society of Craniofacial Surgery (ISCFS)*, 2017. 1, 2

[17] D. Haase, E. Rodner, and J. Denzler. Instance-weighted transfer learning of active appearance models. In *Proceedings of CVPR*, pages 1426–1433, 2014. 2

[18] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In *Dimension reduction of large-scale systems*, pages 261–306. 2005. 6

[19] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 2

[20] V. Lempitsky, S. Roth, and C. Rother. Fusionflow: Discrete-continuous optimization for optical flow estimation. In *Proceedings of CVPR*, pages 1–8, 2008. 2

[21] S. Liang, L. G. Shapiro, and I. Kemelmacher-Shlizerman. Head reconstruction from internet photos. In *Proceedings of ECCV*, pages 360–374, 2016. 1

[22] Li.H, W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *Computer graphics forum*, volume 27, pages 1421–1430, 2008. 2, 6

[23] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010. 2, 4, 6

[24] A. Patel and W. A. Smith. 3d morphable face models revisited. In *Proceedings of CVPR 2009.*, pages 1327–1334, 2009. 2

[25] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *Proceedings of Advanced video and signal based surveillance.*, pages 296–301, 2009. 2

[26] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of CVPR.*, volume 1, pages 947–954, 2005. 3

[27] J. Roth, Y. Tong, and X. Liu. Adaptive 3d face reconstruction from unconstrained photo collections. In *Proceedings of CVPR*, pages 4197–4206. 2

[28] 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

[29] M. Sharif, K. Ayub, D. Sattar, M. Raza, and S. Mohsin. Enhanced and fast face recognition by hashing algorithm. *Applied research and technology*, 10(4):607–617, 2012. 3

[30] B. M. Smith and L. Zhang. Joint face alignment with non-parametric shape models. In *Proceedings of ECCV*, pages 43–56, 2012. 2

[31] M. A. Styner, K. T. Rajamani, L.-P. Nolte, G. Zsemlye, G. Székely, C. J. Taylor, and R. H. Davies. Evaluation of 3d correspondence methods for model building. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 63–75, 2003. 8

[32] P. A. Tresadern, P. Sauer, and T. F. Cootes. Additive update predictors in active appearance models. In *Proceedings of BMVC*, volume 2, page 4, 2010. 2

[33] M. Wardetzky, M. Bergou, D. Harmon, D. Zorin, and E. Grinspun. Discrete quadratic curvature energies. *Computer Aided Geometric Design*, 24(8-9):499–518, 2007. 6

[34] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for tv-l1 optical flow, statistical and geometrical approaches to visual motion analysis, 2009. 2

[35] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997. 2

[36] F. Zhou, J. Brandt, and Z. Lin. Exemplar-based graph matching for robust facial landmark localization. In *Proceedings of ICCV*, pages 1025–1032, 2013. 2

[37] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of CVPR*, pages 2879–2886, 2012. 2, 3