Pattern Recognition Letters
journal homepage: www.elsevier.com

Augmenting a 3D Morphable Model of the Human Head with High Resolution Ears

Hang Dai[a], Nick Pears[a,**], William Smith[a]

[a]*Department of Computer Science, University of York, York, YO10 5GH, United Kingdom*

ABSTRACT

We present a parts-based 3D Morphable Model (3DMM) of the full human head, with particular emphasis on modelling the complex shape of the ear as a flexible, high-resolution separate part. The 3D ear model part undergoes an iterative process of refinement that employs data augmentation using a 2D image dataset with landmarked ears. Evaluations using several performance metrics validate the training process and the resulting model. We make the new ear model and our reconstructed training dataset publicly available. We merge the trained high-resolution 3DMM of the ear with a publicly-available 3DMM of the full head that has a much lower resolution in the ear regions. The resulting parts-based 3DMM provides more shape variation and more shape detail in the ears, and we demonstrate a higher fidelity overall model fit to raw data.

1. Introduction

The process of capturing knowledge about the shape and texture variation of an object class is termed statistical modelling. One example of this is the 3D Morphable Model (3DMM) (Blanz and Vetter, 1999), which is a vector space representation, where any convex combination of vectors of a training set generates a valid example in this vector space. Trained 3DMMs provide an encoding and prior distribution of shape and texture that can be used as a constraint in analysis problems, or generatively in synthesis problems. We are particularly concerned with modelling the human head, which has a wide variety of related applications, such as in affective computing (Garrido et al., 2016), creative media (Saragih et al., 2011), biometrics (An et al., 2018) and semantic explanation of images (Tewari et al., 2017). In addition to the face, the shape of the ear has long been recognised as a means of biometric identification (Pflug and Busch, 2012; Abaza et al., 2013; Emeršič et al., 2017b,a), and is particularly useful in certain head poses.

Recent literature shows that it is very difficult to capture the detailed 3D structure of the ear when training 3DMMs associated with the face (Booth et al., 2018) or the full head (Dai et al., 2017). Many of these approaches require morphing (Amberg et al., 2007; Myronenko and Song, 2010) a pre-defined mesh called a shape template over the whole face/head. Even if the template has a detailed ear structure, the high frequency detail of the fleshy folds is often badly fitted, due to the template morphing optimisation being dominated by the much larger facial and cranial regions. As a consequence, it is difficult to construct a powerful statistical prior in the ear regions.

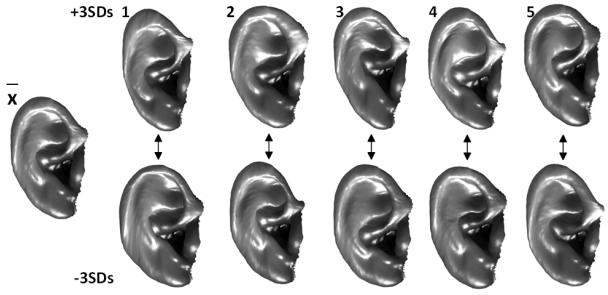

Fig. 1. 3D morphable model of ear. The mean and the first five principal components are shown at +3SD (top row) and -3SD (bottom row).

Here we are particularly concerned with how a 3DMM of the full human head can be improved by replacing its relatively low resolution and inflexible ear shape, with a high resolution and flexible one. This requires us to solve the problems of: i. constructing a high-resolution 3DMM of the human ear in the absence of a suitable large 3D dataset; ii. replacing the ears on an existing 3DMM of the full head with the new ears; iii. treating

**Corresponding author: Tel.: +44 1904 325658;
e-mail: nick.pears@york.ac.uk (Nick Pears)

the upgraded 3DMM of the full head as a parts-based model in terms of fitting the model to raw data. Our model is a general tool that can address a range of analysis and synthesis problems, although these applications are outside the scope of this paper.

Our contributions are: i. a 2D-data-augmented 3DMM training pipeline; ii. the first publicly-available 3DMM of the ear (shape channel only, not texture), as shown in Fig. 1, and public- availability of the 2D-augmented 3D ear training data (Dai et al., 2018); iii. a parts-based 3DMM merging process.

2. Training Data for the Ear Model

To train a 3DMM of the human ear, we would like to have a large dataset of high quality, high resolution 3D ear images. However, such data is very limited in its public availability and, indeed, is very difficult to collect because the folded structure of the ear creates significant self-occlusion. To our knowledge, Zolfaghari et al. (2016) described the only construction of a morphable model for external ear shapes based on a deformation framework that uses diffeomorphic metric mapping. They release high quality 3D meshes of the ear for 10 subjects. Obviously, this is insufficient to construct a 3DMM that is a good population representation of shape. However, using this dataset, with the left ear reflected to be compatible with the right ear shape, we construct an initial approximate model of the ear. The model has over 7K vertices (7111) and we employ a modified version of our morphing technique (Dai et al., 2017) to build the model, which is an extension of Coherent Point Drift (CPD) (Myronenko and Song, 2010). Note that we use a high resolution ear mesh template for morphing that is made to be compatible with the LYHM full head model (Dai et al., 2017) at the joining boundary. This is detailed in Section 4.

Given the lack of 3D data, we aim to leverage a significantly larger annotated 2D ear dataset by reconstructing it into a 3D ear dataset, thereby boosting the initial approximate morphable model in terms of its ability to represent larger populations. Helpfully, Zhou and Zaferiou (2017) made a 2D ear image dataset available with 55 ground-truth landmarks over 600 images, partitioned into 500 training and 100 test images.

3. A 3DMM of the Ear

The process of 2D data-augmented 3DMM training is shown in Fig. 2. There are three stages within the main iterative loop: 1) Initial 3DMM fit using landmarks; 2) Smoothing stage; 3) Landmark position refinement. These are described in the following three subsections respectively. In Section 3.4 we describe alignment and statistical modelling. Finally, in Section 3.5, the iterative loop of 3DMM bootstrapping is described.

3.1. Initial 3DMM Fit

The Scaled Orthographic Projection (SOP) model assumes that variation in depth over the object is small relative to the mean distance from camera to object. Under this assumption, the projected 2D position of a 3D point $\mathbf{X}_i = [x_i, y_i, z_i]^T \in \mathbb{R}^3$, given by $\text{SOP}(\mathbf{X}_i; \mathbf{R}, \mathbf{t}, s) \in \mathbb{R}^2$ does not depend on the distance of the point from the camera, but only on a uniform scale s

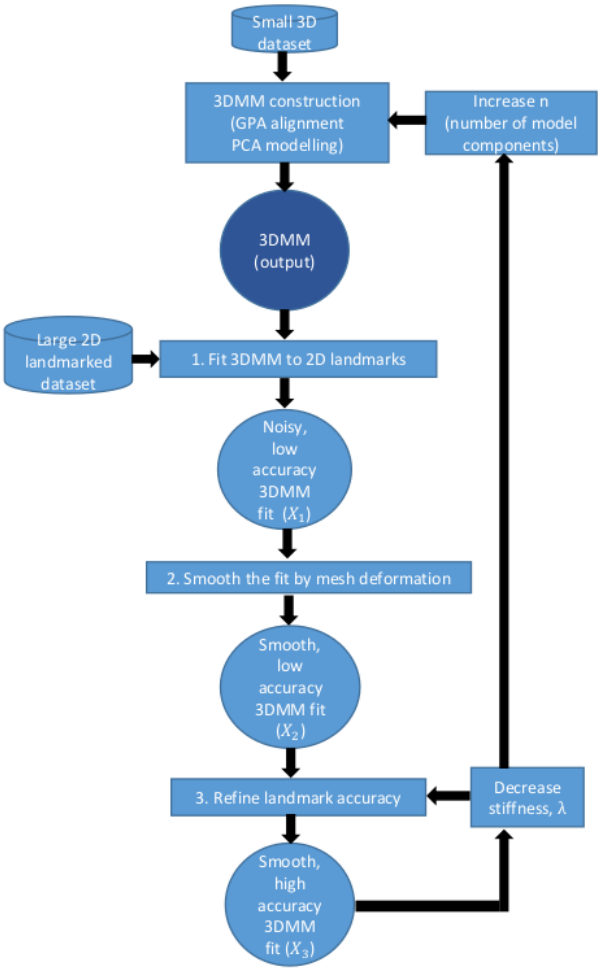

Fig. 2. Iterative model construction process.

given by the ratio of the focal length of the camera and the mean distance from camera to object:

$$\text{SOP}(\mathbf{X}_i; \mathbf{R}, \mathbf{T}, s) = s\mathbf{P}_o\,(\mathbf{R}\mathbf{X}_i + \mathbf{T}) \tag{1}$$

where the 3D pose parameters are given by a rotation matrix $\mathbf{R} \in \text{SO}(3)$ and 3D translation $\mathbf{T} \in \mathbb{R}^3$ and $\mathbf{P}_o$ is the orthographic projection from 3D to 2D:

$$\mathbf{P}_o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2}$$

Defining the 2D translation, $\mathbf{t} \in \mathbb{R}^2$ in the image plane we have

$$\text{SOP}(\mathbf{X}_i; \mathbf{R}, \mathbf{T}, s) = s\mathbf{P}_o\mathbf{R}\mathbf{X}_i + \mathbf{t}, \quad \mathbf{t} = s\mathbf{P}_o\mathbf{T}. \tag{3}$$

Initially, we fit a morphable model to M observed 2D positions $\mathbf{x}_i = [u_i, v_i]^T$, $(i = 1 \ldots M)$ arising from the SOP projection of corresponding vertices in the morphable model. (This is known from a single manual mark up of the mean mesh of the initial approximate model). Without loss of generality, we assume that the i-th 2D position corresponds to the i-th vertex in the morphable model. The objective of fitting a morphable model to these observations is to obtain the size (s), shape ($\mathbf{b}$)

and pose parameters $(\mathbf{R}, \mathbf{t})$ that minimise the reprojection error, E_l, between observed and predicted 2D landmark positions:

$$E_l(s, \mathbf{b}, \mathbf{R}, \mathbf{t}) = \frac{1}{M} \sum_{i=1}^{M} \|\mathbf{x}_i - s\mathbf{P}_o\mathbf{R}(\bar{\mathbf{X}}_i + \mathbf{P}_i\mathbf{b}) - \mathbf{t}\|^2, \quad (4)$$

where $\mathbf{P}_i \in \mathbb{R}^{3\times n}$ are the eigenvector elements associated with the i-th landmark for a n-component model. (Note that the number of model components n associated with $\mathbf{P}_i$ and $\mathbf{b}$ starts small and is gradually increased in our training process, as described later). The problem in (4) is nonlinear least squares that can be solved by various means. Here we use the trust region approach (Coleman and Li, 1996) encapsulated in Matlab's `lsqnonlin` function. The initial landmark-based ear template fit is then recovered as:

$$\mathbf{X}_1 = \bar{\mathbf{X}} + \mathbf{P}\mathbf{b} \quad (5)$$

where $\mathrm{X}_1 = [x_1, \ldots, x_N, y_1, \ldots, y_N, z_1, \ldots, z_N]^T \in \mathbb{R}^{3N}$ represents the vertices of the noisy template fit from stage 1 over $N = 7,111$ vertices of the ear template.

3.2. Smoothing Stage

The 3DMM fitting to a 2D image with landmarks is over-fitted, appearing as surface noise, see Fig. 3 (2). To overcome this we employ a smoothing stage that is composed of two sub-stages. Firstly, we employ the mean of the initial 3DMM, see Fig. 3 (1), as a template, and we deform it using the Coherent Point Drift (CPD) algorithm (Myronenko and Song, 2010) applied with a non-rigid deformation model. The motivation is that the deformed template is able to preserve the same shape, the same number of vertices and also the same triangulation relationship as the over-fitted data, while it can overcome the noise due to over-fitting. Here CPD works well because there is a known one-to-one correspondence between the N vertices on the template and the N vertices on the target.

Secondly, to improve the fit, we implement a projection towards corresponding points that is regularised by the template shape-preserving Laplace-Beltrami (LB) operator (Sorkine and Alexa, 2007). This is achieved by solving the linear system:

$$\begin{bmatrix} \lambda\mathbf{I}_3 \otimes \mathbf{L}_{\mathbf{X}_{\mathrm{CPD}}} \\ \mathbf{I}_{3N} \end{bmatrix} \mathbf{X}_2 = \begin{bmatrix} \lambda\mathbf{L}_{\mathbf{X}_{\mathrm{CPD}}}\mathbf{X}_{\mathrm{CPD}} \\ \mathbf{X}_1 \end{bmatrix} \quad (6)$$

for improved vertex positions $\mathbf{X}_2$, where $\mathbf{X}_{\mathrm{CPD}} = \mathrm{CPD}(\mathbf{X}_{template}, \mathbf{X}_1)$ is the CPD deformed template. We use $\mathbf{I}_N$ to denote the $N \times N$ identity matrix, $\otimes$ is the Kronecker product and, for an N vertex mesh $\mathbf{X}$, $\mathbf{L}_{\mathbf{X}} \in \mathbf{R}^{N\times N}$ denotes the cotangent Laplacian approximation to the LB operator computed from a mesh with vertices $\mathbf{X}$. The parameter λ weights the relative influence of the position and shape regularisation constraints, effectively determining the template shape 'stiffness' of the process. As $\lambda \to 0$ (reducing shape stiffness) the projected shape in $\mathbf{X}_2$ tends towards the original shape $\mathbf{X}_1$. The much smoother outcome $\mathbf{X}_2$ is shown in Fig. 3 (3).

3.3. Landmark Position Refinement

The locations of the fitted landmarks after template deformation are not precise and so we invoke an additional mesh manipulation stage to improve this. Given the 2D landmarks stored in the matrix $\mathbf{x} = [u_1, v_1, \ldots, u_M, v_M]^T \in \mathbb{R}^{2M}$, we define the selection matrices $\mathbf{S} \in [0, 1]^{3M\times 3N}$ that select the M vertices which are the correspondences of the 2D landmarks. We the solve the following linear system for the stage 3 output $\mathbf{X}_3 \in \mathbb{R}^{3N}$:

$$\begin{bmatrix} \lambda\mathbf{I}_3 \otimes \mathbf{L}_{\mathbf{X}_2} \\ \mathbf{G}(M)\mathbf{S} \end{bmatrix} \mathbf{X}_3 = \begin{bmatrix} \lambda\mathbf{I}_3 \otimes \mathbf{L}_{\mathbf{X}_2}\mathbf{X}_2 \\ \mathbf{x} \end{bmatrix} \quad (7)$$

where $\mathbf{G}(M) \in \mathbb{R}^{2M\times 3M}$ projects the 3D landmarks to 2D:

$$\mathbf{G}(M) = \begin{bmatrix} \mathbf{I}_M \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \mathbf{I}_M \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \mathbf{0}_{2M\times M} \end{bmatrix}. \quad (8)$$

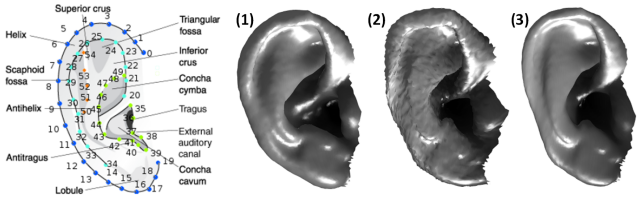

Fig. 3. Left, the 55 landmarks on the ear and their semantic annotations (Zhou and Zaferiou, 2017). Deformation: (1) mean ear template, (2) noisy fit ($\mathbf{X}_1$), (3) smoothed fit ($\mathbf{X}_2$).

3.4. Similarity Alignment and Statistical Modelling

The collection of 500 deformed training meshes are subjected to Generalised Procrustes Analysis (GPA) (Gower, 1975) to remove similarity effects (rotation, translation, scale), leaving only shape information. Note that scale cannot be included as we have no notion of scale within the 2D image dataset, also scale-normalised shapes have a better alignment in a least-squares sense. The aligned meshes are then subject to Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes.

3.5. 3DMM Bootstrapping

Our 2D-augmented 3DMM training process is iterative, in that we rebuild the 3DMM and reapply it to the training dataset for an improved fitting, generating an improved 3DMM at each iteration. This approximate-to-accurate iterative system encapsulates each of the three stages in Sec. 3.1 to Sec. 3.3 within each iteration, see Fig. 2. We increase flexibility relative to the previous iteration, as follows: 1) we increase the number of the shape components in Sec. 3.1 to give the algorithm more variance to do the fitting; 2) we decrease λ in Sec. 3.3 to manipulate the projection of the landmarks in $\mathbf{X}_{\mathrm{edit}}$ towards the 2D landmarks position. 3DMM fitting and mesh manipulation are potentially fragile processes when the 3DMM is approximate, thus we push the algorithm carefully, step-by-step, in this iterative fashion. The resulting model is illustrated in Fig. 1.

4. Merging 3D Morphable Models

We aim to merge our new flexible, high-resolution 3D ear model with the Liverpool-York Head Model (LYHM) (Dai et al., 2017), replacing its relatively low-resolution and relatively inflexible ears. The immediate problem is compatibility of the connecting boundary between the high-resolution ear

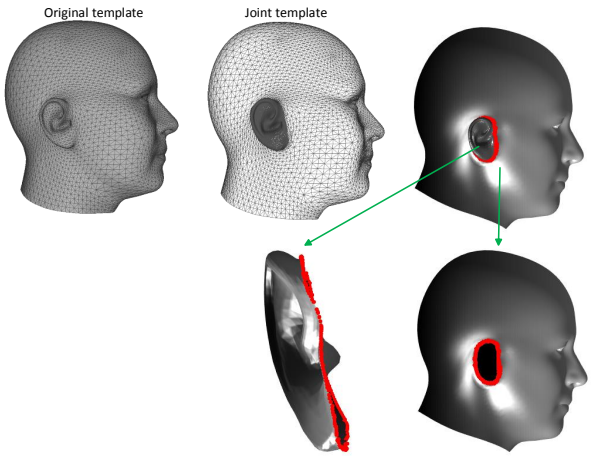

Fig. 4. Original template (left), joint template (centre), and joint template split into separate parts with a common boundary (shown in red)

mesh and the LYHM mesh with its ears removed. This is solved by generating a joint mesh template with a shared boundary, described in Section 4.1, which is then partitioned into ear and no-ear head template parts. Model-part sampling, or fitting, is followed by a three-stage alignment and merging of the constituent parts, described in Section 4.2. Finally in Section 4.3, we briefly outline the parts-based model .

4.1. Joint Template

We use the approach of Schmidt and Singh (2010) to blend the high resolution ear template with the head template to create our joint template. We cut off the high resolution ear from the joint template we use this template to construct the 3DMM of the ear. The connection relation between the two separate parts is known from the joint template, see Fig. 4, and so we can use this to merge the two separate morphable models.

In Fig. 4, the red points represent the shared vertices for the high resolution ear $\mathbf{X}'_e$ and the rest of the head $\mathbf{X}^*_{h-e}$. To merge the two models, we require:

$$\mathbf{S}'_{\mathrm{b}}\mathbf{X}'_e = \mathbf{S}^*_{\mathrm{b}}\mathbf{X}^*_{h-e} \tag{9}$$

where $\mathbf{S}'_{\mathrm{b}}$ selects the boundary vertices on the high resolution 3D ear and $\mathbf{S}^*_{\mathrm{b}}$ selects the boundary vertices on the no-ear head.

4.2. Merging model parts

The merging of the model parts requires: i. rigid alignment, ii. ARAP mesh manipulation (Sorkine and Alexa, 2007), iii. patch smoothing. The basic idea is shown in Fig. 5. Note that, if we omit the central mesh manipulation stage, we end up with an undesirable bump. Each stage is now described.

Rigid alignment. We begin by rigidly aligning the high resolution 3D ear sample $\mathbf{X}'_e$ to the low resolution 3D ear $\mathbf{X}^*_{\mathbf{e}}$ on the LYHM head mesh sample $\mathbf{X}^*$. This can be solved by normalising the scale of the high-resolution ear to that of the low-resolution ear, and then using ICP (Besl and McKay, 1992).

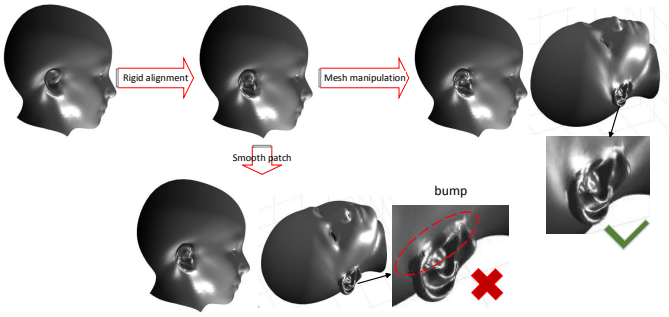

Fig. 5. The merging flowchart includes two stages: 1) rigid alignment; 2) mesh manipulation. If the output of the rigid alignment undergoes a patch smoothing operation only, it suffers from a discontinuity problem, ending up with a bump, shown in the lower three views. Our mesh manipulation overcomes this, as shown in the top-right three views.

Mesh Boundary Manipulation. We manipulate the boundary on $\mathbf{X}'_e$ towards the boundary on $\mathbf{X}^*_e$ and the rest of $\mathbf{X}'_e$ is moved *As Rigid As Possible* (ARAP) (Sorkine and Alexa, 2007). Given a refined high resolution 3D ear mesh, whose vertices are stored in the matrix $\mathbf{X}'_{\text{refined}} \in \mathbb{R}^{p\times 3}$. This can be written as:

$$\begin{bmatrix} \lambda\mathbf{L}_{\mathbf{X}'_e} \\ \mathbf{S}'_b \end{bmatrix} \mathbf{X}'_{\text{refined}} = \begin{bmatrix} \lambda\mathbf{L}_{\mathbf{X}'_e}\mathbf{X}'_e \\ \mathbf{S}^*_b\mathbf{X}^*_{\mathbf{h-e}} \end{bmatrix} \tag{10}$$

where $\mathbf{X}'_{\text{refined}}$ is the refined ear position 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 mesh manipulation. As $\lambda \to \infty$, the ear part stays in its original position. As $\lambda \to 0$, the boundary on the ear part is moved onto its target positions.

Patch smoothing. After mesh manipulation, small artefacts can be removed by a patch smoothing technique proposed by Desbrun et al. (1999), which employed an implicit integration method along with a scale-dependent Laplacian operator and a robust curvature flow operator to portray a smooth surface.

4.3. Parts-based morphable model

Given a selection matrix $\mathbf{S}^*_{\text{h-e}}$ that selects the no-ear head part $\mathbf{X}^*_{\mathbf{h-e}}$ on the a head sample $\mathbf{X}^*$, which is generated by the head model, a new instance $\mathbf{X}'$ generated by the merged model can also be represented as $\mathbf{X}' = [\mathbf{X}'_{\text{refined}}; \mathbf{S}^*_{\text{h-e}}\mathbf{X}^*]$. $\mathbf{X}'_{\text{refined}}$ can be solved from a linear system and $\mathbf{S}^*_{\text{h-e}}\mathbf{X}^*$ can obtained from the head model linearly. So the parts-based morphable model is still a linear model, which facilities its application in 3DMM fitting to 2D and 3D images. Note that our high resolution ear model is a right ear and we use a reflection to fit a left ear.

5. Ear Modelling and Fitting Evaluation

We used the training method described in Section 3 to build a 3DMM of the ear using 500 training images (Zhou and Zafeiriou, 2017). Section 5.1 qualitatively illustrates model fitting performance over a wide range of head poses. In Section 5.2, we validate the complexity of our training process in an ablation

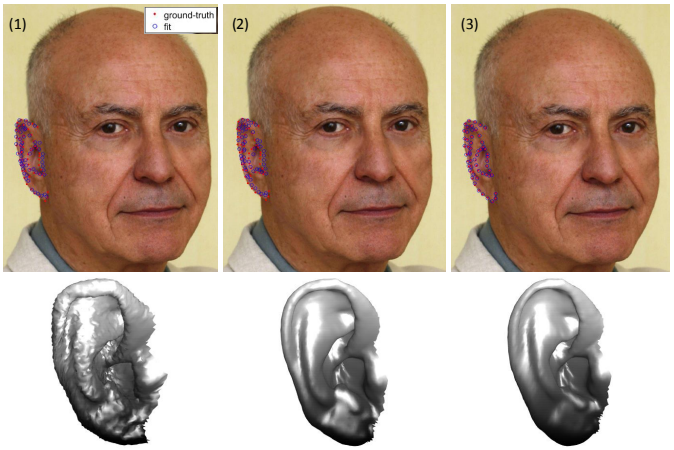

Fig. 6. Results of each training step, with ground truth landmarks shown as red dots and fitted landmarks as blue circles (best viewed on high zoom): 1) 3DMM overfitted to 2D landmarks; 2) smoothing process using mesh deformation; 3) LB-regularised mesh manipulation to refine the landmark positions.

Fig. 7. Model fitting for augmentation: First row - original raw images, with the model's 3D landmarks (blue circles) projected towards their 2D counterparts (red dots). Second row - augmented 3D data with per-vertex texture mapped over the model's surface, and rendered in a canonical view. Note that the first two columns are the same person.

study i.e. how does omitting some module, eg data augmentation, affect performance. The relative performance of the final models in terms of compactness, generalisation and specificity (Styner et al., 2003) is presented in Section 5.3. (Obviously, we would like to test our model against other ear models but there is no public 3DMM of the ear available for direct comparison.) We compare the proposed 3DMM merging method with other methods in Section 5.4. Finally, visualisations of the merged 3DMMs are presented in Section 5.5.

5.1. Qualitative Evaluation of Model Fitting

We illustrate the outcome of each training step in Fig. 6. The landmark positions of the fitted results get closer and closer to the ground truth manual landmarks. There is obvious overfitting to the 2D landmarks in the first stage. The smoothing step via template deformation removes the noise, but it still keeps the same landmark positions as the first step. The outcome of final step (LB mesh manipulation) has refined landmark positions that are almost the same as the ground-truth, yet with a smooth ear shape.

For illustration of the accuracy of the ear shape, we rigidly align the 3D ear shape to the 2D landmarks on the 2D image.

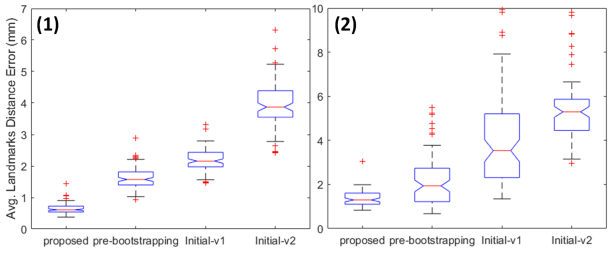

Fig. 8. Mean landmark distance error for four system variants: (1) Landmark error distribution, (2) Fitting consistency distribution.

We then sample the image intensities onto the vertices of the model, normalise to a canonical pose and render. As can be seen in Fig. 7 across widely different head poses, the textured mesh appears correct under rotation and the complete ear texture (with no background) is sampled onto the mesh. Note that we do not build a statistical texture model - we show sampled textures simply for visualisation.

5.2. Ablation Study

Here, the 3DMM training variants include: i) the proposed 3DMM training method, using several bootstrapping iterations, and 500 2D training images, ii) the proposed method without any bootstrapping iterations (i.e. one pass of the three stages in Sec. 3) and 500 2D training images, iii) the initial 20-image 3DMM passed through the three steps in Sec. 3, with no 2D landmarked data augmentation (Initial-v1 method) and iv) the initial approximate 3DMM with just 3DMM fitting, i.e. no template morphing or mesh manipulation stages, and no 2D data augmentation (Initial-v2 method). For all four methods, we use two metrics: *landmark error* and *fitting consistency* to evaluate the performance quantitatively.

Landmark error is the average landmark distance error between the projected 3D landmarks and the 2D landmarks. Fig. 8(1) shows that the proposed method has the lowest landmark error, which is below 1mm, and that both data augmentation and bootstrapping (iterative model improvement) have significant beneficial effects.

Fitting consistency can be measured as the dataset contains multiple images of the same person, as shown in the first two columns of Fig. 7. First we fit the 3D model to the first image of a pair, thus fixing the 3D model shape. Then, without changing the model shape, we project it into the second image and measure the mean landmark error relative to the manual 2D landmarks. We compensate for differences in scale between the two images in the fitting process. As shown is Fig. 8(2), the proposed method has the lowest distance error, which implies that the fitting from the proposed method is more consistent with other images of the same person.

5.3. Compactness, Generalisation and Specificity

For quantitative model evaluation, we employ the three metrics proposed by Styner et al. (2003) namely: compactness, generalisation and specificity. Such metrics require that the compared models should have the same number of model components. In this context, we compare the proposed method and the proposed method without bootstrapping.

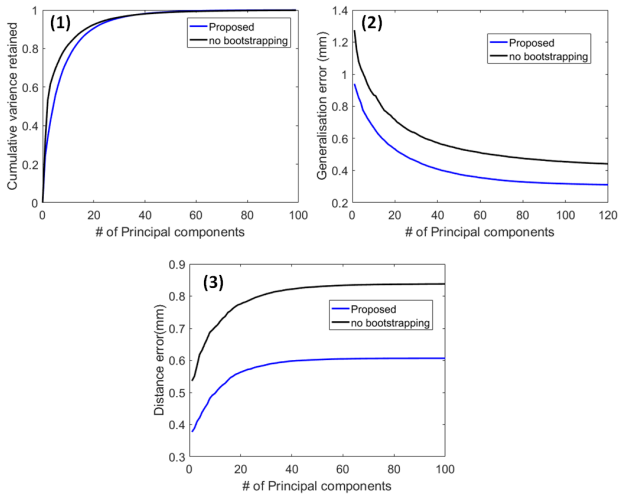

Fig. 9. Model evaluation: (1) Compactness, (2) Generalisation, (3) Specificity.

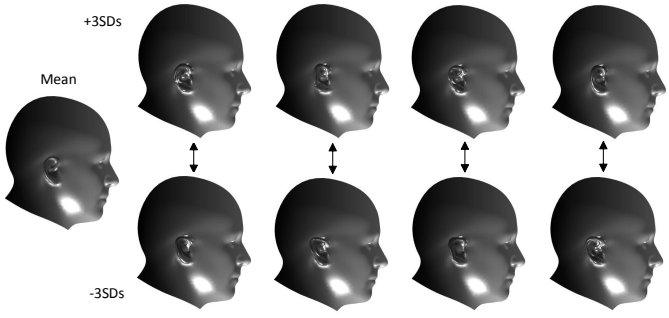

Fig. 10. The first four modes of high resolution 3DMM of ear merged into a mean head from LYHM.

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. As can be from Fig. 9, the proposed method without bootstrapping has better compactness than the proposed method when < 25 principal components are used. When >25 principal components are used, the compactness is similar. The proposed method has the lower generalisation error, which implies that proposed method has the better performance in describing unseen examples. The proposed method has the lower distance error in the specificity metric, which implies that the proposed method is better at generating instances close to real data.

5.4. Comparison of Ear Merging

We compare the proposed ear merging method with mesh smoothing (Desbrun et al., 1999) after ear alignment and Laplacian mesh manipulation (Sorkine et al., 2004). As shown in Fig. 12 (2), with ear alignment, the high resolution ear mesh is rigidly transformed to the right position. If we use mesh smoothing directly after ear alignment, the joint area ends up with undesirable bumps which is presented in Fig. 12 (3). Laplacian mesh manipulation is based on the Laplacian to do

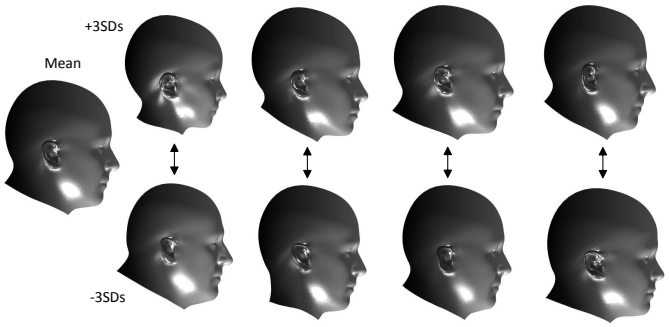

Fig. 11. The first four modes of high resolution 3DMM of ear merged into the first four modes of head model.(For illustration only, please note that the model parts are independent.)

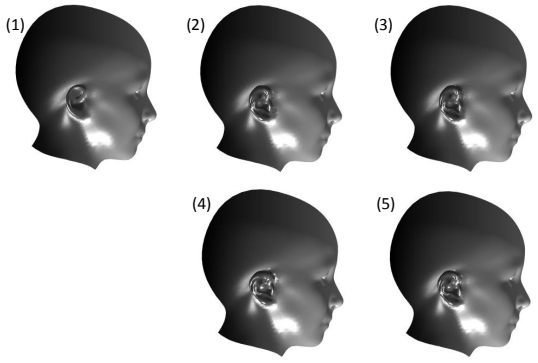

Fig. 12. Comparison of Ear Merging: (1) original head mesh; (2) ear alignment; (3) mesh smoothing after ear alignment; (4) Laplacian mesh manipulation; (5) proposed. Best viewed on zoom.

interactive free-form deformation. As can be seen from Fig. 12 (4), Laplacian mesh manipulation presents a non-rigid deformation in ear shape. However, this changes the high resolution ear shape, which is not desirable in this process. Fig. 12 (5) demonstrates the the proposed method. It shows a smoothed joint area between the ear part and face part. The ear shape is the same as that after rigid alignment, which results from the *As Rigid As Possible* property of LB mesh manipulation. Fig. 13 shows the ear merging results for different identities using the proposed method.

5.5. Visualisation of the Merged Morphable Model

The merged morphable model is derived from merging the proposed ear model with the LYHM head model (Dai et al., 2017). Fig. 10 demonstrates the first 4 modes of high resolution 3DMM of ear merged into a mean head. In this case, the head shape is fixed and the ear shape is varied. Fig. 11 presents the first 4 modes of high resolution 3DMM of ear merged into the first 4 modes of head model. Here, the head shape and ear shape are both varied (although no correlation is implied). In order to validate the improvement in 3DMM fitting to 2D images, we use the landmark fitting algorithm of Zhou and Zaferiou (2017). The ear landmarks are given, and we use the facial landmarking system of Zhu and Ramanan (2012) for full head fitting. The fitting results are shown in Fig. 14. It shows that the parts-based

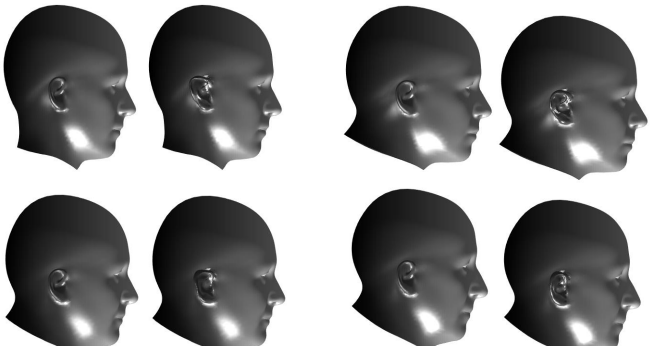

Fig. 13. Ear merging results: every pair of images includes one original head mesh followed by the merged result.

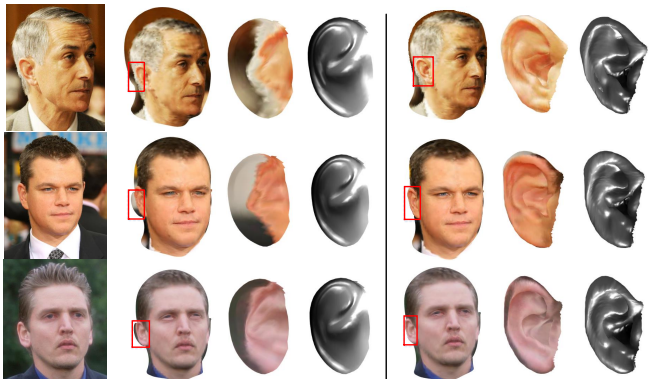

Fig. 14. Fitting results of a single 3DMM (the LYHM, left) and the proposed parts-based 3DMM, right.

morphable model improves the performance of 3DMM fitting to 2D images when compared with a head model only. One can clearly see that there is not much shape variation in the ear that is generated by the head model only.

6. Conclusions

We proposed an iterative 3DMM training process using 2D data augmentation to bootstrap a strong 3DMM of the human ear from a weak one. Evaluation demonstrates that the method lowers the landmark error and the fitted data is more consistent within images of the same person. The bootstrapping strategy improves the model performance in both generalisation and specificity. The limitation is the requirement for manual 2D landmarks on the 2D training data. We proposed a framework of merging high resolution ear shape with a 3DMM of the head. The merged morphable models provide significantly more ear shape variation than models built by morphing a single full head template. Our modelling and merging techniques can be generalised to other shapes that require detailed modelling of specific parts.

Acknowledgments

The authors thank Google Faculty Awards 2017-18, and their Google sponsor Forrester Cole, for supporting this work.

References

Abaza, A., Ross, A., Hebert, C., Harrison, M.A.F., Nixon, M.S., 2013. A survey on ear biometrics. ACM computing surveys (CSUR) 45, 22.

Amberg, B., Romdhani, S., Vetter, T., 2007. Optimal step nonrigid ICP algorithms for surface registration. IEEE Conference on Computer Vision and Pattern Recognition , 1–7.

An, Z., Deng, W., Yuan, T., Hu, J., 2018. Deep transfer network with 3d morphable models for face recognition, in: 2018 13th IEEE Int. Conf. on Automatic Face Gesture Recognition, pp. 416–422.

Besl, P.J., McKay, N.D., 1992. Method for registration of 3-d shapes, in: Sensor Fusion IV: Control Paradigms and Data Structures, International Society for Optics and Photonics. pp. 586–607.

Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3d faces, in: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, pp. 187–194.

Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S., 2018. Large scale 3d morphable models. International Journal of Computer Vision 126, 233–254.

Coleman, T.F., Li, Y., 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on optimization 6, 418–445.

Dai, H., Pears, N., Smith, W., 2018. The York Ear Model (YEM). URL: `https://www-users.cs.york.ac.uk/~nep/research/YEM/`. accessed 6th Dec 2018.

Dai, H., Pears, N., Smith, W., Duncan, C., 2017. A 3d morphable model of craniofacial shape and texture variation, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE. pp. 3104–3112.

Desbrun, M., Meyer, M., Schröder, P., Barr, A.H., 1999. Implicit fairing of irregular meshes using diffusion and curvature flow, in: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co.. pp. 317–324.

Emeršič, Ž., Štepec, D., Štruc, V., Peer, P., George, A., Ahmad, A., Omar, E., Boult, T.E., Safdaii, R., Zhou, Y., et al., 2017a. The unconstrained ear recognition challenge, in: Biometrics (IJCB), 2017 IEEE International Joint Conference on, IEEE. pp. 715–724.

Emeršič, Ž., Štruc, V., Peer, P., 2017b. Ear recognition: More than a survey. Neurocomputing 255, 26–39.

Garrido, P., Zollhöfer, M., Casas, D., Valgaerts, L., Varanasi, K., Pérez, P., Theobalt, C., 2016. Reconstruction of personalized 3d face rigs from monocular video. ACM Trans. Graph. 35, 28:1–28:15.

Gower, J.C., 1975. Generalized procrustes analysis. Psychometrika 40, 33–51.

Myronenko, A., Song, X., 2010. Point set registration: Coherent point drift. IEEE transactions on pattern analysis and machine intelligence 32, 2262–2275.

Pflug, A., Busch, C., 2012. Ear biometrics: a survey of detection, feature extraction and recognition methods. IET biometrics 1, 114–129.

Saragih, J.M., Lucey, S., Cohn, J.F., 2011. Real-time avatar animation from a single image, in: IEEE Int Conf Automatic Face and Gesture Recognition 2011, pp. 213–220.

Schmidt, R., Singh, K., 2010. Meshmixer: an interface for rapid mesh composition, in: ACM SIGGRAPH 2010 Talks, ACM. p. 6.

Sorkine, O., Alexa, M., 2007. As-rigid-as-possible surface modeling, in: Proceedings of the Fifth Eurographics Symposium on Geometry Processing, pp. 109–116.

Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C., Seidel, H.P., 2004. Laplacian surface editing, in: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing, ACM. pp. 175–184.

Styner, M.A., Rajamani, K.T., Nolte, L.P., Zsemlye, G., Székely, G., Taylor, C.J., Davies, R.H., 2003. Evaluation of 3d correspondence methods for model building. Information processing in medical imaging , 63–75.

Tewari, A., Zollöfer, M., Kim, H., Garrido, P., Bernard, F., Perez, P., Christian, T., 2017. MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction, in: The IEEE International Conference on Computer Vision (ICCV).

Zhou, Y., Zaferiou, S., 2017. Deformable models of ears in-the-wild for alignment and recognition, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), IEEE. pp. 626–633.

Zhu, X., Ramanan, D., 2012. Face detection, pose estimation, and landmark localization in the wild, in: Proceedings of CVPR, pp. 2879–2886.

Zolfaghari, R., Epain, N., Jin, C.T., Glaunès, J., Tew, A., 2016. Generating a morphable model of ears, in: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, IEEE. pp. 1771–1775.