A Data-augmented 3D Morphable Model of the Ear

Hang Dai, Nick Pears and William Smith
Department of Computer Science, University of York, UK

***Abstract*— Morphable models are useful shape priors for biometric recognition tasks. Here we present an iterative process of refinement for a 3D Morphable Model (3DMM) of the human ear that employs data augmentation. The process employs the following stages 1) landmark-based 3DMM fitting; 2) 3D template deformation to overcome noisy over-fitting; 3) 3D mesh editing, to improve the fit to manual 2D landmarks. These processes are wrapped in an iterative procedure that is able to bootstrap a weak, approximate model into a significantly better model. Evaluations using several performance metrics verify the improvement of our model using the proposed algorithm. We use this new 3DMM model-booting algorithm to generate a refined 3D morphable model of the human ear, and we make this new model and our augmented training dataset public.**

I. INTRODUCTION

The shape of the ear has long been recognised as a means of biometric identification [1], [2], [3], [4]. There are many existing ear recognition systems in the literature [5], [6], [7], [8], [9], [10], with a recent survey by Emeršič et al. [11] and ear biometrics continues to be an active research area [12]. Morphable models provide powerful statistical priors on shape and so can be used in biometric ear analysis. We present a pipeline capable of building a 3D Morphable Model (3DMM) of the human ear from a very limited training sample of 3D ears, using data augmentation.

We have 20 high quality 3D meshes of the ear [13], taken from 10 subjects, with the left ear reflected to be compatible with the right ear shape. This is insufficient to construct a 3D morphable model that is a good representation of the mean ear shape and the variance and covariances of size-and-shape (form), over a large population. However, with such a limited dataset, we construct an initial approximate model of the form:

$$\mathbf{X}(\alpha) = \bar{\mathbf{X}} + \mathbf{P}\alpha \tag{1}$$

where the 3DMM parameters are the mean shape $\bar{\mathbf{X}}$, the shape variation components $\mathbf{P}$, and shape parameters α. The model has over 7K vertices (7111) and we employ a modified version of the the morphing technique in Dai et al [14] to build the model, which is an extension of Coherent Point Drift (CPD) [15]. Subsequently, 3D data augmentation is able to generate new samples for the 3DMM construction, thereby boosting the initial morphable model in terms of its accuracy in representing larger populations. Recently, Zhou et al. made a 2D ear image dataset available with 55 ground-truth landmarks [16] over 600 images, partitioned into 500 training images and 100 test images. Fig.1 (left) shows the 55 landmarks and their semantic annotations.

This work was supported by a Google Faculty Award

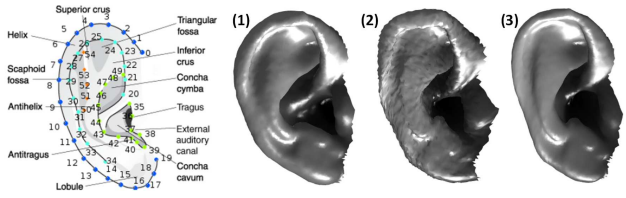

Fig. 1. 55 landmarks on ear and their semantic annotations [16]. Deformation: (1) template, (2) over-fitting, (3) deformed template.

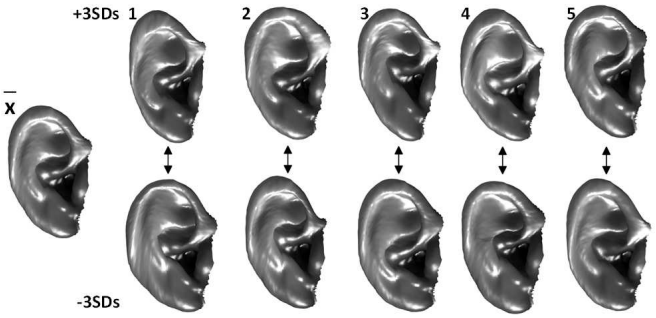

Fig. 2. 3D morphable model of ear. The mean and the first five principal components are shown at +3SD (top row) and -3SD (bottom row).

Our process includes the following stages: 1) landmark-based 3DMMs fitting; 2) use of 3D deformation to overcome the over-fitting (caused by an insufficient number of training subjects); 3) 3D mesh editing, regularized by 2D image information. Our contributions are (i) the data-augmented model building pipeline and (ii) the first publicly-available 3DMM of the ear. This model is shown in Fig. 2. In addition to the model, the augmented training data will be made publicly available. After reviewing related work, we detail our proposed data augmentation method in Sec. III. Section IV then describes our iterative model construction process, whilst Sec. V presents our evaluations.

II. RELATED WORK

To our knowledge, Zolfaghari et al. [13] published the only work on a 3D morphable model of ear, but the morphable model is not publicly available. Zhou et al. built a 2D morphable model of the ear and made a 2D ear image dataset publicly available with 55 manually-labelled ground-truth landmarks [16]. Our work is based on this dataset and requires a non-rigid shape template deformation method. Several extensions of Iterative Closest Points (ICP) for the nonrigid case were proposed [17], [18], [19], [20]. The extensions of ICP have good performance in shape difference elimination but have problems in over-fitting and point sliding. Amberg et al. [17] defined the optimal-step *Nonrigid*

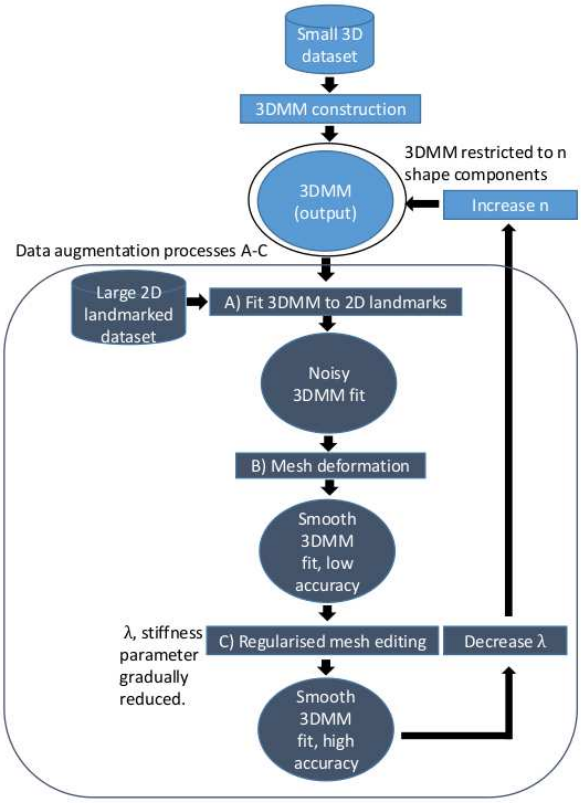

Fig. 3. Iterative model construction process.

Iterative Closest Points (NICP) framework, which extended ICP methods to nonrigid deformations while retaining its convergence properties. Li et al. [21] show that using proximity heuristics to determine correspondences is less reliable when large deformations are present. Global correspondence optimization solves simultaneously for both the deformation parameters as well as the correspondence positions [21]. Myronenko et al. consider the alignment of two point sets as a probability density estimation [15] and they call the method Coherent Point Drift (CPD). The CPD approach is extended in [22], [23], [24], [25]. Dai et al. proposed a hierarchical parts-based CPD-LB morphing framework to avoid under-fitting and over-fitting [14]. In this paper, we employ this method to overcome the over-fitting.

III. 3D Ear Data Augmentation

The process of data-augmented 3DMM construction is shown in Fig. III. Here, data augmentation has three stages: A) 3DMM fitting with 2D ear landmarks; B) 3D deformation to overcome the over-fitting of the initial approximate 3DMM; C) 3D mesh editing to manipulate the projection of the landmarks in the augmented 3D mesh towards the manually-labelled 2D landmark positions. These three stages are described in the following subsections, with the iterative loop for model construction described in Sec. IV.

A. Landmark-based 3DMM Fitting

The scaled orthographic projection (SOP) [26] 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 $SOP(\mathbf{X}_i; \mathbf{R}, \mathbf{t}, s) \in \mathbb{R}^2$ does not depend on the distance of the point from the camera, but only on a uniform scale s given by the ratio of the focal length of the camera and the mean distance from camera to object:

$$SOP(\mathbf{X}_i; \mathbf{R}, \mathbf{T}, s) = s\mathbf{P}_o\left(\mathbf{R}\mathbf{X}_i + \mathbf{T}\right) \tag{2}$$

where the 3D pose parameters are given by a rotation matrix $\mathbf{R}$ $s.t.$ $\mathbf{R} \in \mathbb{R}^{3\times3}$, $\mathbf{R}^T\mathbf{R} = \mathbf{I}_3$ and 3D translation $\mathbf{T} \in \mathbb{R}^3$. $\mathbf{P}_o$ is the orthogonal projection from 3D to 3D defined by

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

and so, defining the 2D translation, $\mathbf{t}$ in the image plane we have

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

We begin by showing how to fit a morphable model to M observed 2D positions $\mathbf{x}_i = [u_i, v_i]^T (i = 1...M)$ arising from the SOP projection of corresponding vertices in the morphable 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 shape and pose parameters that minimise the reprojection error, E_{lmk}, between observed and predicted 2D landmark positions:

$$\mathbf{E}_{lmk}(\alpha, \mathbf{R}, \mathbf{t}, s) = \frac{1}{M}\sum_{i=1}^{M} ||\mathbf{x}_i - SOP(\bar{\mathbf{X}}_i + \mathbf{P}_i\alpha; \mathbf{R}, \mathbf{t}, s)||^2 \tag{4}$$

The problem is non-linear least squares that can be solved by various means. Here we use the trust region approach [27] encapsulated in Matlab's `lsqnonlin` function.

B. 3D Mesh Deformation

The number of training subjects for the initial 3DMM is insufficient, so the 3DMM fitting to a 2D image with landmarks, described in Sec. III-A is over-fitted, appearing as surface noise, see Fig.1 (2). To overcome this, we employ the mean of the initial 3DMM, see Fig.1 (1) as a template, and we deform it using the Coherent Point Drift (CPD) algorithm [15] applied with a non-rigid deformation model, followed by a projection to corresponding points that is regularised by the template shape-preserving Laplace-Beltrami (LB) operator. Such a deformation regulation process was proposed by Dai et al. [14]. The motivation for the deformation process 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. The deformation algorithm works well because there is a known one-to-one correspondence

Fig. 4. Augmentation results: 1st row - original images, 2nd row - 3D landmarks projection to 2D images, 3rd row - augmented data with texture

between the 7111 vertices on the template and the 7111 vertices on the target. The outcome is shown in Fig.1 (3).

C. Mesh Editing Regularised by 2D Landmarks

The location of the fitted landmarks after the initial 3DMM fitting and the template deformation are not accurate, relative to the manually-labelled 2D landmarks. We overcome this by treating the template mesh manipulation as a mesh editing problem with two ingredients. First, position constraints are provided by those 2D landmarks, the correspondences of which are known in 3D mesh. Second, regularisation constraints are (again) provided by the LB operator, which acts to retain the local structure of the mesh.

The LB mesh editing problem can be written as a linear system of equations. Given the vertices of a 3D mesh stored in the matrix $\mathbf{X} = [x_1, y_1, z_1, ..., x_N, y_N, z_N]^T \in \mathbb{R}^{3N}$ and the 2D landmarks stored in the matrix $\mathbf{x} = [u_1, v_1, ..., u_M, v_M]^T \in \mathbb{R}^{2M}$, 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. This linear system can be written as:

$$\begin{pmatrix} \lambda \mathbf{L_3} \\ \mathbf{G}(M)\mathbf{S} \end{pmatrix} \mathbf{X}_{\text{edit}} = \begin{pmatrix} \lambda \mathbf{L_3 X} \\ \mathbf{x} \end{pmatrix} \tag{5}$$

$$\mathbf{G}(M) = \mathbf{I}_M \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

where $\mathbf{I}_M$ is the $M \times M$ indentity matrix and $\mathbf{G}(M) \in \mathbb{R}^{2M \times 3M}$ project the 3D landmarks to 2D, $\mathbf{L}_3 \in \mathbb{R}^{3N \times 3N}$ is the cotangent Laplacian approximation to the LB operator and $\mathbf{X}_{\text{edit}} \in \mathbb{R}^{3N}$ are the edited vertex positions that we wish to solve for. The parameter λ weights the relative influence of the position and shape regularisation constraints, effectively determining the template shape 'stiffness' of the mesh editing process. As $\lambda \to 0$ (reducing shape stiffness) the projected 3D landmarks in $\mathbf{X}_{\text{edit}}$ tend towards the same positions as the 2D manual landmarks.

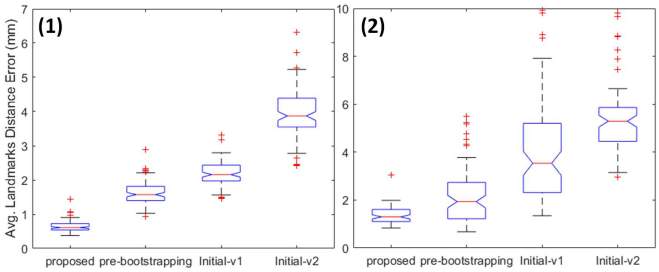

Fig. 5. Average landmarks distance error for four system variants: (1) Landmark error, (2) Fitting consistency.

IV. Morphable Model Construction

A. Similarity Alignment & Statistical Modelling

The collection of the augmented meshes are subjected to Generalised Procrustes Analysis (GPA) [28] to remove similarity effects (rotation, translation, scale), leaving only shape information. (Scale cannot be included as we have no notion of scale within the 2D image dataset.) The aligned meshes are then subject to Principal Component Analysis (PCA), generating a 3DMM as a linear basis of shapes. This allows for the generation of novel shape instances.

B. 3DMM Bootstrapping

We propose a 3DMM bootstrapping procedure where, at each bootstrap iteration, we rebuild the 3DMM and reapply it to the augmented dataset for an improved fitting to that dataset, and hence we can generate a better 3DMM in the next iteration. This approximate-to-accurate iterative system encapsulates each of the three key stages in Sec. III-A to Sec. III-C within each iteration. We push each procedure harder relative to the previous iteration, as follows: 1) we increase the number of the shape components in Sec. III-A to give the algorithm more variance to do the fitting; 2) we decrease λ in Sec. III-C to manipulate the projection of the landmarks in $\mathbf{X}_{\text{edit}}$ towards the 2D landmarks position. 3DMM fitting and mesh editing are potentially fragile processes when the

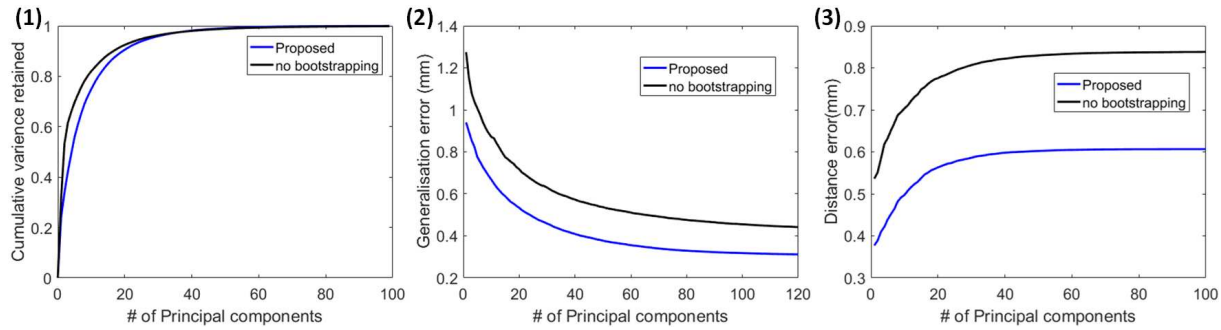

Fig. 6. Model evaluation: (1) Compactness, (2) Generalisation, (3) Specificity.

3DMM is approximate, thus we push the algorithm step-by-step in this iterative fashion.

V. Evaluation

We used the proposed method to build a 3DMM of the ear over 500 training images and used the remaining 100 images for testing the performance in the given dataset [16]. There is no public 3DMM of the ear available for direct comparison. However, we evaluate the performance of model construction, both qualitatively and quantitatively, for several variants of our method. These include: i) the proposed method, using several bootstrapping iterations, and 500 data augmentation images, ii) the proposed method without any bootstrapping iterations (i.e. one pass of the three steps in Sec. III) and 500 data augmentation images, iii) the initial 20-image 3DMM passed through the three steps in Sec. III, with no data augmentation (Initial-v1 method) and iv) the initial 20-image 3DMM with just 3DMM fitting, i.e. no template morphing or mesh editing stages, and no data augmentation (Initial-v2 method).

A. Qualitative Evaluation

As can be seen in Fig. 4, the proposed method can handle different head poses. After mesh editing, the projected positions of the model's 3D landmarks are almost the same as the ground-truth 2D landmarks.

B. Quantitative Evaluation

We use two metrics: landmark error and fitting consistency to evaluate the performance quantitatively.

1) Landmark Error: This measure is calculated by the average landmark distance error between the projected 3D landmarks and the 2D landmarks, over the test set (100 images). As shown is Fig. 5(1), the proposed method has the lowest landmark error.

2) Fitting Consistency: The dataset contains multiple images of the same person, as shown in the first two columns of Fig. 4. This allows us to do some consistency checking in the following way. 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. 5(2), the proposed method has the lowest distance error, which implies that the fitting from the proposed method is more consistent with the other images of the same person.

C. Model Evaluation

Since model evaluation requires that the models should have the same number of principal components, we compare the proposed method and the proposed method without bootstrapping. For quantitative model evaluation, Styner et al [29] 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. As can be from Fig.6, 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. Specificity measures how well a model is able to generate instances that are similar to real data. The proposed method has the lower distance error, which implies that the proposed method is better at generating instances close to real data.

VI. Conclusions

We proposed iterative 3DMM construction using 3D data augmentation to bootstrap a strong 3DMM of the human ear from a weak one. The method overcomes noisy over-fitting and manipulates the projection of 3D landmarks towards the desired 2D landmark positions. Evaluation demonstrates that the method lowers 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 a requirement for manual 2D landmarks. The next step is to do augmentation with either fewer landmarks or even none at all, requiring modelling of lighting and surface reflectance.

Acknowledgements: We thank Google Faculty Awards for their support.

REFERENCES

[1] Pflug, A. and Busch, C., 2012. Ear biometrics: a survey of detection, feature extraction and recognition methods. IET biometrics, 1(2), pp.114-129.

[2] Jain, A., Bolle, R. and Pankanti, S. eds.. Biometrics: personal identification in networked society, 2006 (Vol. 479). Springer Science & Business Media.

[3] Abaza, A., Ross, A., Hebert, C., Harrison, M. A. F., and Nixon, M. S.. A survey on ear biometrics. ACM computing surveys (CSUR), (2013) 45(2), 22.

[4] Woodard, D., Faltemier, T., Yan, P., Flynn, P., and Bowyer, K. 2006. A comparison of 3D biometric modalities. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). 5761.

[5] Islam, S.M., Davies, R., Bennamoun, M. and Mian, A.S.. Efficient detection and recognition of 3D ears. International Journal of Computer Vision, 2011, 95(1), pp.52-73.

[6] Yuan, L., Wang, Z., and Mu, Z. 2010. Ear recognition under partial occlusion based on neighborhood preserving embedding. Proc. SPIE, Biometric Technol. Hum. Identif. VII 7667.

[7] Zhang, H. and Mu, Z. 2008. Ear recognition method based on fusion features of global and local features. In Proceedings of the IEEE International Conference on Wavelet Analysis and Pattern Recognition.

[8] Yan, P. and Bowyer, K. 2006. An automatic 3D ear recognition system. In Proceedings of the 3rd IEEE International Symposium on 3D Data Processing Visualization and Transmission. pp. 326333.

[9] Xie, Z. AND Mu, Z. 2008. Ear recognition using lle and idlle algorithm. In Proceedings of the 19th IEEE International Conference on Pattern Recognition (ICPR).

[10] Wang, Y., Mu, Z., and Zeng, H. 2008. Block-Based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns. In Proceedings of the 19th IEEE International Conference on Pattern Recognition (ICPR). 14.

[11] Ž. Emeršič, V. Štruc, and P. Peer, Ear recognition : more than a survey, 2017, Neurocomputing, vol 255, pp26-39.

[12] Ž. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. Boult, R. Safdari, Y. Zhou, S. Zafeiriou, D. Yaman, F. Eyiokur, H. Ekenel, The Unconstrained Ear Recognition Challenge, 2017, arXiv 1798.06997v1.

[13] Zolfaghari, R., Epain, N., Jin, C.T., Glauns, J. and Tew, A., 2016, March. Generating a morphable model of ears. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 1771-1775).

[14] H. Dai, N. Pears, W. Smith, and C. Duncan. A 3d morphable model of craniofacial shape and texture variation. Proceedings ICCV, 2017.

[15] Myronenko, A. and Song, X., 2010. Point set registration: Coherent point drift. IEEE transactions on pattern analysis and machine intelligence, 32(12), pp.2262-2275.

[16] Zhou, Y. and Zaferiou, S.. Deformable Models of Ears in-the-wild for Alignment and Recognition. In Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 626-633.

[17] Amberg, B., Romdhani, S. and Vetter, T.. Optimal step nonrigid ICP algorithms for surface registration. In Proceedings of CVPR, 2007, pp. 1-8.

[18] Pears, N. and Duncan, C., 2016. Automatic 3D modelling of craniofacial form. arXiv preprint arXiv:1601.05593.

[19] Hontani H, Matsuno T, Sawada Y. Robust nonrigid ICP using outlier-sparsity regularization. In Proceedings of CVPR, 2012, pp. 174-181.

[20] Cheng, S., Marras, I., Zafeiriou, S. and Pantic, M.. Active nonrigid ICP algorithm. In Automatic Face and Gesture Recognition (FG), 2015, Vol. 1, pp. 1-8.

[21] Li, H., Sumner, R.W. and Pauly, M.. Global Correspondence Optimization for NonRigid Registration of Depth Scans. In Computer graphics forum, 2008, Vol. 27, No. 5, pp. 1421-1430.

[22] Wang, P., Wang, P., Qu, Z., Gao, Y. and Shen, Z.. A refined coherent point drift (CPD) algorithm for point set registration. Science China Information Sciences, 2011, 54(12), pp.2639-2646.

[23] Golyanik, V., Taetz, B., Reis, G. and Stricker, D., 2016, March. Extended coherent point drift algorithm with correspondence priors and optimal subsampling. In Applications of Computer Vision (WACV), 2016, pp. 1-9.

[24] Hu, Y., Rijkhorst, E.J., Manber, R., Hawkes, D. and Barratt, D., September. Deformable vessel-based registration using landmark-guided coherent point drift. In International Workshop on Medical Imaging and Virtual Reality, 2010, pp. 60-69.

[25] H. Dai, N. E. Pears, W. Smith and C. Duncan. Symmetric Shape Morphing for 3D Face and Head Modelling. In Automatic Face & Gesture Recognition (FG 2018), 2018.

[26] Bas, Anil, and William AP Smith. "What does 2D geometric information really tell us about 3D face shape?." arXiv preprint arXiv:1708.06703 (2017).

[27] Coleman, T., Li, Y.: An interior, trust region approach for nonlinear minimization subject to bounds. SIAM J. Optimiz. 6 (1996) 418445.

[28] Gower, John C. "Generalized procrustes analysis." Psychometrika 40.1 (1975): 33-51.

[29] Styner, M., Rajamani, K.T., Nolte, L.P., Zsemlye, G., Szkely, G., Taylor, C.J. and Davies, R.H.. Evaluation of 3D correspondence methods for model building. In IPMI, 2003 (Vol. 3, pp. 63-75).