Multilevel Models of Age-Related Changes in Facial Shape in Adolescents

DJJ Farnell[†], J Galloway, AI Zhurov, and S Richmond

School of Dentistry, Cardiff University, Heath Park, Cardiff CF14 4XY

FarnellD@cardiff.ac.uk ([†]corresponding author)

GallowayJL@cardiff.ac.uk

ZhurovAI@cardiff.ac.uk

RichmondS@cardiff.ac.uk

Abstract. Here we study the effects of age on facial shape in adolescents by using a method called multilevel principal components analysis (mPCA). An associated multilevel multivariate probability distribution is derived and expressions for the (conditional) probability of age-group membership are presented. This formalism is explored via Monte Carlo (MC) simulated data in the first dataset; where age is taken to increase the overall scale of a three-dimensional facial shape represented by 21 landmark points and all other "subjective" variations are related to the width of the face. Eigenvalue plots make sense and modes of variation correctly identify these two main factors at appropriate levels of the mPCA model. Component scores for both single-level PCA and mPCA show a strong trend with age. Conditional probabilities are shown to predict membership by age group and the Pearson correlation coefficient between actual and predicted group membership is $r = 0.99$. The effects of outliers added to the MC training data are reduced by the use of robust covariance matrix estimation and robust averaging of matrices. These methods are applied to another dataset containing 12 GPA-scaled (3D) landmark points for 195 shapes from 27 white, male schoolchildren aged 11 to 16 years old. 21% of variation in the shapes for this dataset was accounted for by age. Mode 1 at level 1 (age) via mPCA appears to capture an elongation in facial shape with age, which is consistent with age-related shape changes in children. Component scores for both single-level PCA and mPCA again show a distinct trend with age. Conditional probabilities are again shown to reflect membership by age group and the Pearson correlation coefficient is given by $r = 0.63$ in this case. These analyses are an excellent first test of the ability of multilevel statistical methods to model age-related changes in facial shape in adolescents.

Keywords: multilevel principal components analysis; multivariate probability distributions; facial shape; age-related changes in adolescents.

1 Introduction

The importance of modeling the effects of groupings or covariates in shape or image data is becoming increasingly recognized, e.g., a bootstrapped response-based imputation modeling (BRIM) of facial shape [1], a linear mixed model of optic disk shape [2], or variational auto-encoders more generally (see, e.g., [3–5]). Multilevel principal components analysis (mPCA) has also been shown [6–10] to provide an efficient method of modeling shape and image texture in such cases. Previous calculations using the mPCA approach have focused on: facial shape for a population of subjects that demonstrated groupings by ethnicity and sex [7–8], image texture for two expressions (neutral and smiling) [9,10], and time-series shape data tracked through all phases of a smile [10]. Here we consider how age-related changes in facial shape can be modelled by multilevel statistical approaches for Monte Carlo (MC) simulated data and for real data by using a model that is illustrated schematically in Figure 1.

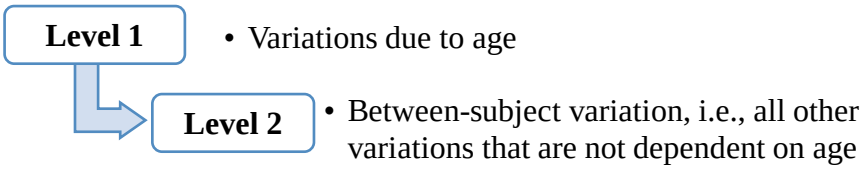

Figure 1. Multilevel model of the effects of age on facial shape.

2 Methods

2.1 Mathematical Formalism

3D landmark points for each shape are represented by a vector z. Single-level PCA is carried out by finding the mean shape vector μ over all shapes and a covariance matrix

$$\Sigma_{k_1,k_2} = \frac{1}{N-1}\sum_{i=1}^{N}(z_{ik_1} - \mu_{ik_1})(z_{ik_2} - \mu_{ik_2}) \cdot \tag{1}$$

k_1 and k_2 indicate elements of this covariance matrix and i refers to a given subject. The eigenvalues λ_l and (orthonormal) eigenvectors u_l of this matrix are found readily. For PCA, one ranks all of the eigenvalues into descending order and one retains the first l_1 components in the model. The shape z is modeled by

$$z^{model} = \mu + \sum_{l=1}^{l_1} a_l u_l \,, \tag{2}$$

The coefficients $\{a_l\}$ (also referred to as "component scores" here) are found readily by using a scalar product with respect to the set of orthonormal eigenvectors, i.e., $a_l = u_l \cdot (z - \bar{z})$, for a fit of the model to a new shape vector z. The component score a_l is standardized by dividing by the square root of the eigenvalue λ_l.

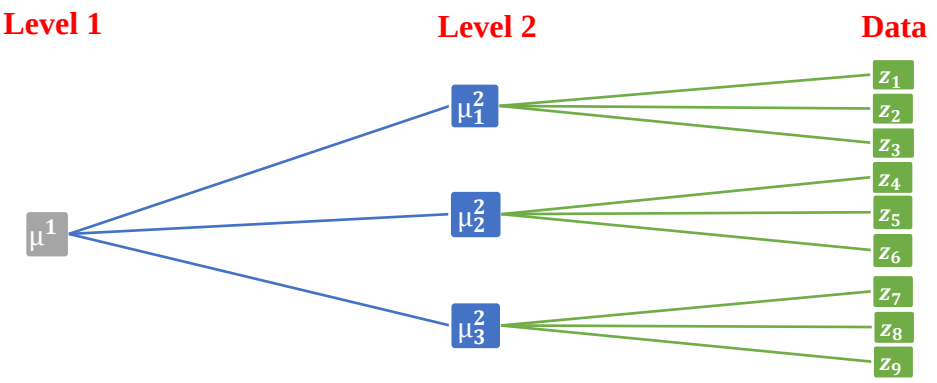

Figure 2. Multilevel model represented as a tree. Shapes μ_l^2 at level 2 are average shapes over all shape data z in a given group l (e.g., 3 shapes per group are shown above). The shape μ^1 at level 1 is the average shape over all of the shape data μ_l^2 at level 2 (e.g., 3 groups at level 2 are shown above).

Multilevel PCA (mPCA) allows us to isolate the effects of various influences on shape at different levels of the model. This allows us to adjust for each subjects' individual facial shape in order to obtain a clearer picture of those changes due to the primary factor, i.e., age here. The covariance matrix at level 2 is formed with respect to all subjects in each age group l and then these covariance matrices are averaged over all age groups to give the level 2 covariance matrix Σ_2. The average shape for group l at level 2 is denoted μ_l^2. By contrast, the covariance matrix at level 1, Σ_1, is formed with respect to the shapes μ_l^2 at each age group at level 2. The overall "grand mean" shape at level 1 is denoted μ^1. These relationships for the multilevel model are illustrated as a tree diagram in Figure 2.

mPCA uses PCA with respect to the covariance matrices at the two levels separately. The l-th eigenvalue at level 1 is denoted by λ_l^1 with associated eigenvector u_l^1, whereas the l-th eigenvalue at level 2 is denoted by λ_l^2 with associated eigenvector u_l^2. We rank all of the eigenvalues into descending order at each level of the model separately, and then we retain the first l_1 and l_2 eigenvectors of largest magnitude at the two levels, respectively. The shape z is modeled by

$$z^{model} = \mu^1 + \sum_{l=1}^{l_1} a_l^1 u_l^1 + \sum_{l=1}^{l_2} a_l^2 u_l^2 \quad , \tag{3}$$

where μ^1 is the "grand mean" at level 1, as described above. The coefficients $\{a_l^1\}$ and $\{a_l^2\}$ (again referred to as "component scores" here) are determined for any new shape, z, by using a global optimization procedure in MATLAB R2017 with respect to an

appropriate cost function [6-10]. The mPCA component scores a_l^1 and a_l^2 may again be standardized by dividing by the square roots of λ_l^1 and λ_l^2, respectively.

From Figure 2, we define that the probability along a branch linking level 1 to group l at level 2 as $P(l)$. Furthermore, we may define the probability along a branch linking group l at level 2 to the data z as $P(z|l)$. The probability of both is therefore,

$$P(z,l) = P(l)P(z \mid l) \quad . \tag{4}$$

Assuming m groups at level 2, we see immediately also that

$$P(z) = \sum_{l=1}^{m} P(z,l) = \sum_{l=1}^{m} P(l)P(z \mid l) \quad . \tag{5}$$

These results lead on to Bayes theorem, which implies that

$$P(l \mid z) = \frac{P(l)P(z \mid l)}{\sum_{l=1}^{m} P(l)P(z \mid l)} \quad . \tag{6}$$

Here we shall use a multivariate normal distribution at level 2, which is given by

$$P(z \mid l) = N(z \mid \mu_l^2, \Sigma_2) \quad . \tag{7}$$

For small numbers of groups m at level 2, one might set $P(l)$ to be constant. In this case, the conditional probability that a given shape z belongs to group l is given by,

$$P(l \mid z) = \frac{N(z \mid \mu_l^2, \Sigma_2)}{\sum_{l=1}^{m} N(z \mid \mu_l^2, \Sigma_2)} \quad . \tag{8}$$

For larger numbers of groups m at level 2, it might be more appropriate to model $P(l)$ as a multivariate normal distribution also, by using

$$P(l) = N(\mu_l^2 \mid \mu^1, \Sigma_1) \quad . \tag{9}$$

The conditional probability that a given shape z belongs to group l is now given by,

$$P(l \mid z) = \frac{N(\mu_l^2 \mid \mu^1, \Sigma_1) N(z \mid \mu_l^2, \Sigma_2)}{\sum_{l=1}^{m} N(\mu_l^2 \mid \mu^1, \Sigma_1) N(z \mid \mu_l^2, \Sigma_2)} \quad . \tag{10}$$

The extension of this approach to three or more levels is straightforward.

2.2 Image Capture, Preprocessing, and Subject Characteristics

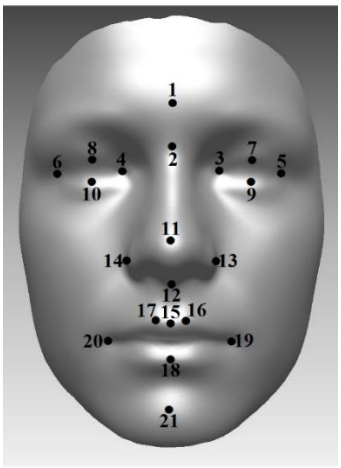

Figure 3. Illustration of the 21 landmark points for dataset 1.

MC simulations were used initially to explore age-related changes on facial shape in dataset 1. A template facial shape containing 21 landmarks points in three dimensions was constructed firstly, as shown in Figure 3. The effects of age were simulated by applying a scale factor that grew linearly with age (in arbitrary units) to all points in this template equally. All other "subjective" variation was included by altering the width of the face for all subjects randomly (irrespective of age). A small amount of normally distributed random error was added to all shapes additionally. However, there were essentially just two factors affecting facial shape in dataset 1. We expect the overall change in scale to be reflected at level 1 (age) of the multilevel model shown in Figure 1 and changes in the width to be reflected at level 2 (all other sources of variation). All shapes were centered on the origin and the average scale across all shapes was set to be equal to 1. Note that 30 age groups were used here with 300 subjects per group in the "training set" used in forming the original model and 100 per group for a separate testing dataset. The effects of outliers in the training set for single-level PCA and mPCA was explored by carrying out additional calculations with an extra 5% of the data containing strong outliers. Such shapes were outlying in terms of facial width and overall scale. Dataset 2 contained real data of 195 shapes from 27 white, male subjects (aged 11 to 16) selected from two large comprehensive schools in the South Wales Valleys area in Rhonda Cynon Taf. Those with craniofacial anomalies were excluded. Ethical approval was obtained from the director of education, head teachers, school

committees, and the relevant ethics committees of Bro Taf. Written informed consent was obtained before obtaining the 3D laser scans. 12 landmark points along the centerline of the face and between the eyes were then used to describe facial shape. All shapes were GPA transformed and the average scale across all shapes was again set to be equal to 1.

3 Results

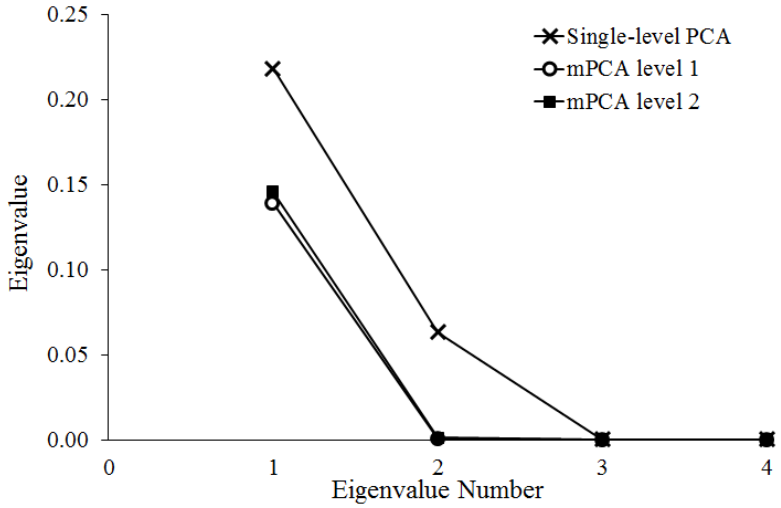

Figure 3. Eigenvalues for single-level PCA and mPCA level 1 (age) and level 2 (all other variations) for dataset 1.

Eigenvalues for single-level PCA and mPCA are shown in Figure 3 for dataset 1. These results for mPCA demonstrated a single non-zero eigenvalue for the level 1 (age) and a single large eigenvalue for the level 2 (all other variations), as expected. Results for the eigenvalues for single-level PCA are of comparable magnitude to those results of mPCA, as one would expect, and they follow a very similar pattern.

Modes of variation of shape for dataset 1 are presented in Figure 4. The first mode at level 1 (age) via mPCA and mode 1 via single-level PCA both capture increases in overall size of the face. The first mode at level 2 (all other variations) for mPCA clearly corresponds to changes in the width of the face. However, mode 2 via single-level PCA clearly mixes the effects of overall changes in size and width of the face. Such "mixing" is a limitation of single-level PCA.

Results for the standardized component 'scores' for mPCA are shown in Figure 5. Component 1 for level 1 (age) via mPCA demonstrates differences due to age clearly because the centroids are strongly separated. Indeed, there is a clear progression of these centroids with age. By contrast, component 1 for level 2 (all other variations except age) via mPCA does not seem to reflect changes due to age very strongly (not shown

here). Results for both components 1 and 2 via single-level PCA again demonstrate a clear trend with age.

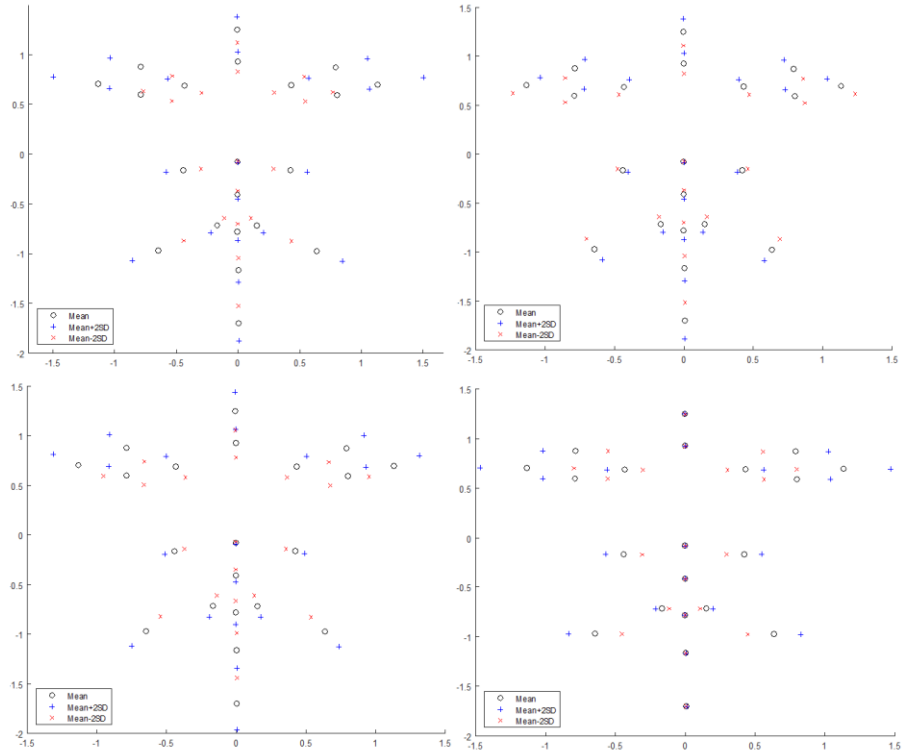

Figure 4. Modes of shape variation in the frontal plane only for dataset 1: (upper left) = mode 1 via single-level PCA; (upper right) = mode 2 via single-level PCA; (lower left) = mode 1 at level 1 (age) via mPCA; (lower right) = mode 1 at level 2 (all other variations) via mPCA. (Landmark points are illustrated in Figure 3.)

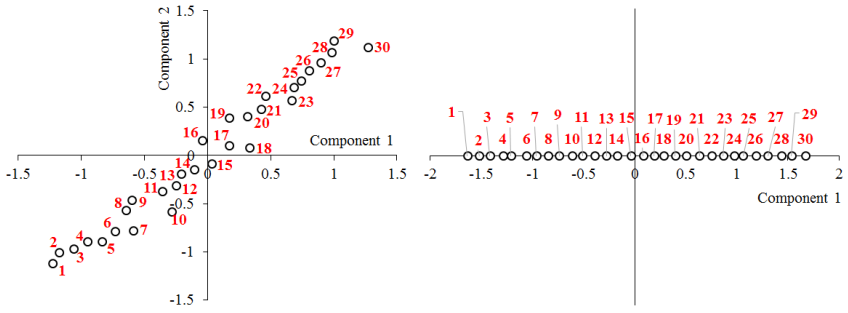

Figure 5. Centroids of standardized component scores for each of the 30 age groups (indicated by labels) for the test shapes in dataset 1 for (left) single-level PCA (modes 1 and 2) and (right) mPCA for mode 1 at level 1 (age).

Results for the (conditional) probabilities of group membership of Eq. (8) are shown as a heat map in Figure 6 for dataset 1. (Note that very similar results are seen by using Eq. (10) for a multivariate normal distribution for $P(l)$ and so these results are not presented here.) A strong trend in the maximal probabilities is observed in Figure 6 that clearly reflects the groupings by age. Age-group membership for each shape was predicted by choosing the group for which the conditional probability was highest. The Pearson correlation coefficient of actual versus predicted age group (from 1 to 30) is given by $r = 0.99$.

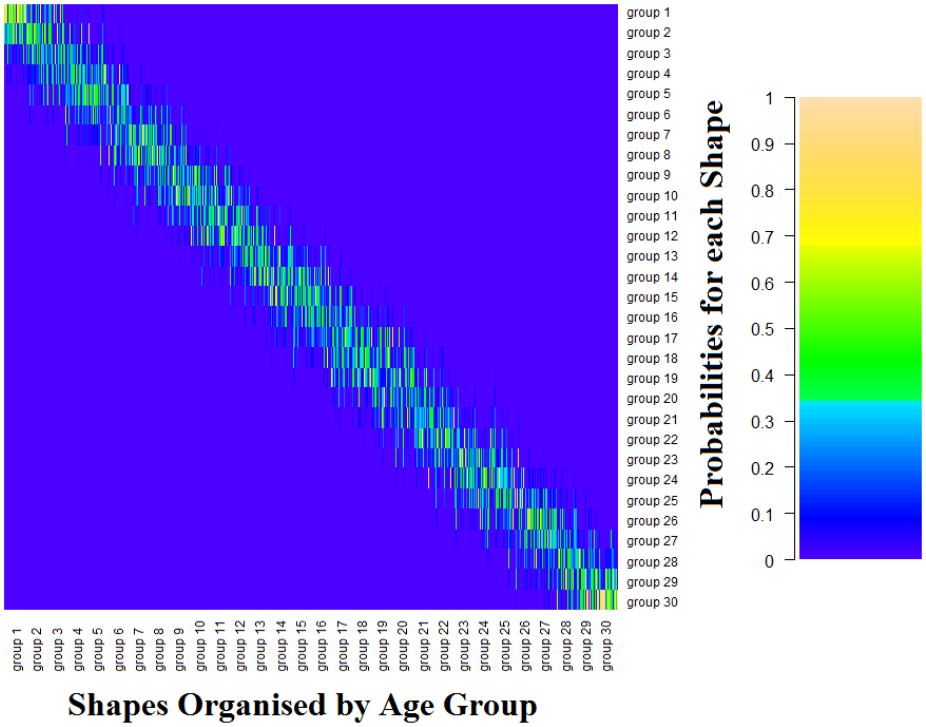

Figure 6. Heat map of (conditional) probabilities of group membership of Eq. (8) for the 3000 test shapes used in dataset 1 (30 age groups and 100 shapes per group).

The effect of adding outlying shapes to the training shape data in dataset 1 was to increase the magnitude of eigenvalues and to add "random scatter" to points in the major modes of variation for both single-level PCA and mPCA. Model fits for the test set were seen to demonstrate a progression with age with that was less clear than in Figure 5 due to this source of additional error. Furthermore, the overall scale of the (standardized) component scores was increased and conditional probabilities of Eqs. (8) and (10) became less efficient at predicting group membership, e.g., Pearson's r was reduced. Robust covariance matrix estimation and robust (median) averaging of covariance matrices was found to reduce the effects of outliers in these initial studies.

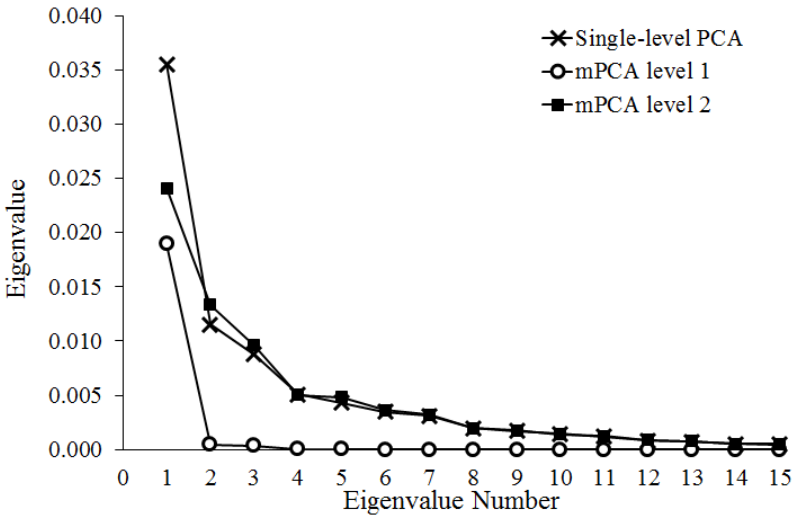

Figure 7. Eigenvalues for single-level PCA and mPCA level 1 (age) and level 2 (all other variations) for dataset 2.

Eigenvalues for single-level PCA and mPCA are shown in Figure 7 for dataset 2. The results for mPCA demonstrate a single large non-zero eigenvalue for the level 1 (age) only, which is presumably due to the small number of landmark points and / or the small number of groups at this level. However, level 2 (all other variations) does now have many large non-zero eigenvalues, which is reasonable for "real data." Results for the eigenvalues via single-level PCA are again of comparable magnitude to those results of mPCA and they follow a very similar pattern. mPCA calculations suggest that age contributed approximately 21% of the total variation for this 3D shape dataset.

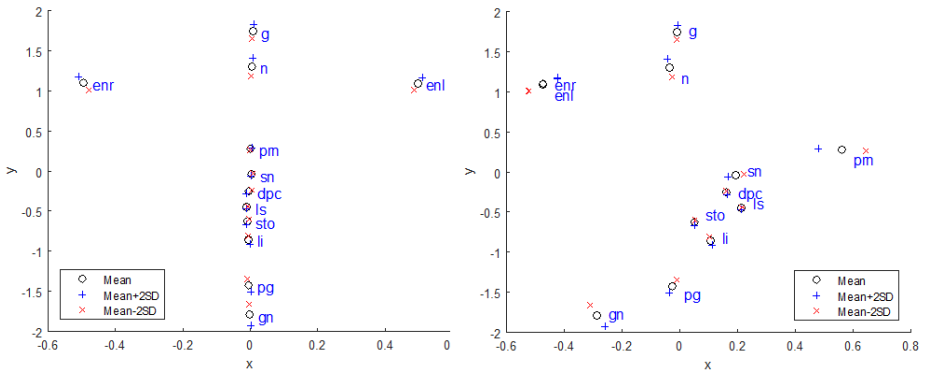

Figure 8. Mode 1 of shape variation at level 1 (age) via mPCA for dataset 2: (left) frontal plane; (right) sagittal plane. (Glabella (g), nasion (n), endocanthion left (enl), endocanthion right (enr), pronasale (prn), subnasale (sn), labiale superius (ls), labiale inferius (li), pogonion (pg), gonion (gn), philtrum (dpc), and stomion (sto).)

Mode 1 at level 1 (age) via mPCA is shown in Figure 8 for dataset 2. This mode represents an overall increase in length of the face (and distance between the eyes) and a decrease in the distance between the endocanthion and pronasale. Broadly, one might interpret this as an elongation in facial shape, which is consistent with the growth of children [11]. Subtle differences are observed only between modes 1 at levels 1 (age) and 2 (all other variations) via mPCA, although we believe that these differences would become more apparent with increased number of landmark points. Mode 1 via single-level PCA is similar to both of these modes via mPCA. Mode 2 via single-level PCA and mode 2 at level 2 via mPCA are similar; both modes appear to relate to shape changes relating to the eyes and prominence of the chin. However, all modes are difficult to resolve with so few landmark points, and future studies of age-related changes in facial shape in adolescents will include more such landmark points.

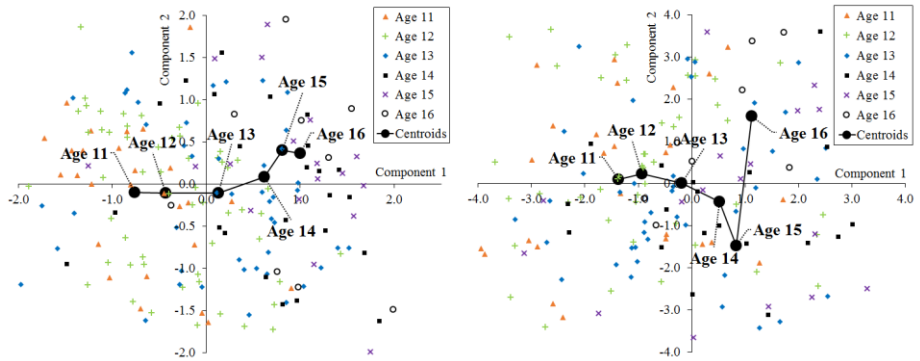

Figure 9. Standardized component scores with respect to shape for dataset 2 for (left) single-level PCA and (right) mPCA at level 1 (age).

Results for the standardized component 'scores' via mPCA are shown in Figure 9. Component 1 for level 1 (age) via mPCA demonstrates differences due to age clearly because the centroids are strongly separated. Indeed, a clear progression of these scores with age is again seen via mPCA at level 1. Component 2 for level 1 (age) shows a possible difference between ages 15 and 16, although this is probably due to random error because the sample size for age 16 was quite small. Component scores for level 2 (all other variations) for mPCA again do not seem to reflect changes due to age very strongly (not shown here). This is an encouraging result given that we found only very subtle differences between modes 1 at levels 1 and 2 via mPCA. A clear trend with age is also seen in Figure 9 for both components 1 and 2 via single-level PCA.

Results for the probabilities of group membership of Eq. (8) are shown as a heat map in Figure 10 for dataset 2. (Note that very similar results are again seen by using Eq. (10) for a multivariate normal distribution for $P(l)$ and so these results are not presented here.) "Miss-one-out" testing was used here, i.e., the model in each case was formed from all shape data except for the shape being tested. A trend is again observed in the maximal probabilities in Figure 11 that reflects the groupings by age, although this trend is not quite as clear as for the MC-simulated data. We expect that this trend will

become clearer with increased number of landmark points. Again, age-group membership for each shape was predicted by choosing the group for which the conditional probability was highest. The Pearson correlation coefficient of actual versus predicted group membership (from 11 to 16 years old) is given by $r = 0.63$.

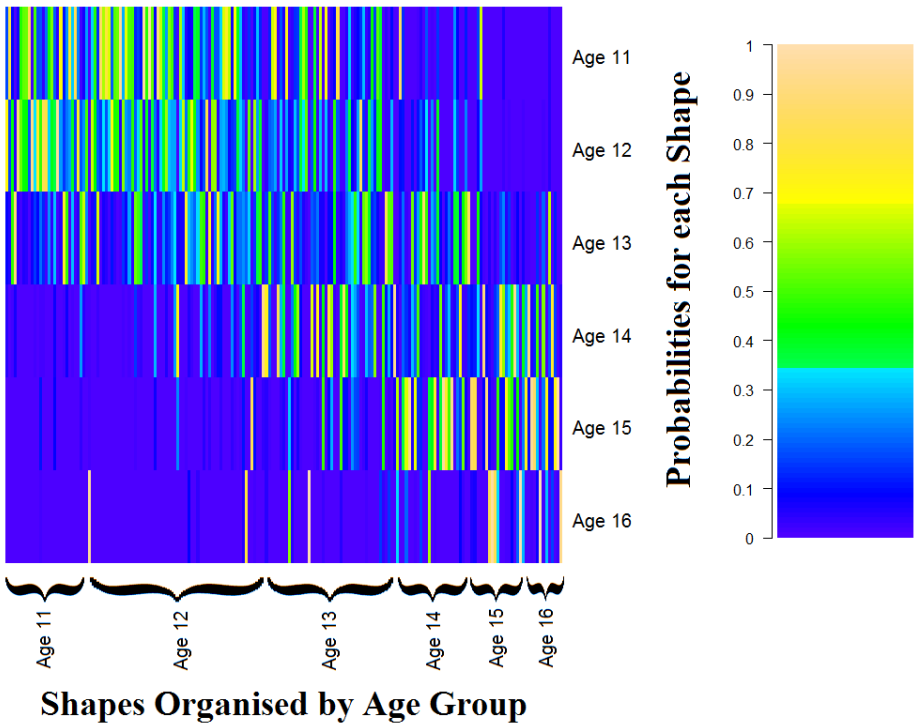

Figure 10. Heat map of (conditional) probabilities of group membership of Eq. (8) for the 195 shapes in dataset 2 (27 white, male subjects, aged 11 to 16 years old) using "miss-one-out" testing.

4 Conclusions

The effect of age on 3D facial shape data has been explored in this article. The formalism for mPCA has been described and it was seen that mPCA allows us to model variations at different levels of structure in the data, i.e., age at one level of the model and all other variations at another level. Two datasets were considered, namely, MC-simulated data of 21 3D landmark points and real data for 195 shapes of 12 3D landmark points for 27 white, male subjects aged 11 to 16 years old. Eigenvalues appeared to make sense for both datasets. In particular, examination of these eigenvalues suggested that age contributed approximately 21% to the total variation in the shapes for the real data in dataset 2. Modes of variation also appeared to make sense for both datasets. Evidence of clustering by age group was seen in the component scores for both the simulated data and also the real data. An initial exploration of the associated multivariate probability distribution for such multilevel architectures was presented.

Conditional probabilities were used to predict group membership. Results for the predicted and actual group memberships were positively correlated and the Pearson correlation coefficients were $r = 0.99$ and $r = 0.63$ for the MC-simulated and the real data, respectively. These results are an encouraging initial exploration of the use of multilevel statistical methods to explore and understand age-related changes in facial shape.

References

1. Claes P, Hill H, Shriver MD (2014) Toward DNA-based facial composites: Preliminary results and validation, Forensic Science International: Genetics **13**, 208–216.
2. MacCormick IJ, Williams BM, Zheng Y, Li K, Al-Bander B, Czanner S, Cheeseman R, Willoughby CE, Brown EN, Spaeth GL, Czanner G (2019) Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. PloS One **10**, e0209409.
3. Doersch C (2016) Tutorial on variational autoencoders, arXiv:1606.05908.
4. Pu Y, Gan Z, Henao R, Yuan X, Li C, Stevens A, Carin L (2016) Variational autoencoder for deep learning of images, labels and captions. In Advances in Neural Information Processing Systems, Curran Associates, Inc., Red Hook, USA. pp. 2352–2360.
5. Wetzel SJ (2017) Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. Physical Review E **96**, 022140.
6. Lecron F, Boisvert J, Benjelloun M, Labelle H, Mahmoudi S (2012) Multilevel statistical shape models: A new framework for modeling hierarchical structures, 9th IEEE International Symposium on Biomedical Imaging (ISBI) 1284-1287.
7. Farnell DJJ, Popat H, Richmond S (2016) Multilevel principal component analysis (mPCA) in shape analysis: A feasibility study in medical and dental imaging, Computer Methods and Programs in Biomedicine **129**, 149-159.
8. Farnell DJJ, Galloway J, Zhurov A, Richmond S, Perttiniemi P, Katic V (2017) Initial Results of Multilevel Principal Components Analysis of Facial Shape, Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, pp. 674–685.
9. Farnell DJJ, Galloway J, Zhurov A, Richmond S, Pirttiniemi P, Lähdesmäki R (2018) What's in a Smile? Initial Results of Multilevel Principal Components Analysis of Facial Shape and Image Texture. In Medical Image Understanding and Analysis. Springer: Cham, Switzerland; Volume **894**, pp. 177–188.
10. Farnell DJJ, Galloway J, Zhurov A, Richmond S, Marshall D, Rosin PL, Al-Meyah K, Pirttiniemi P, Lähdesmäki R (2019) What's in a Smile? Initial Analyses of Dynamic Changes in Facial Shape and Appearance, Journal of Imaging **5**, 2.
11. Lin AJ, Lai S, and Cheng F (2010) Growth simulation of facial/head model from childhood to adulthood. Computer-Aided Design and Applications **7**, 777–786.