In: International Workshop on Growth and Motion in 3D Medical Images, European Conference on Computer Vision, Copenhagen, Denmark, 2002. (to appear)

Estimating Average Growth Trajectories in Shape-Space using Kernel Smoothing

Tim J. Hutton
Biomedical Informatics Unit
Eastman Dental Institute
University College London, UK

Bernard F. Buxton
Computer Science Department
University College London, UK

Peter Hammond
Biomedical Informatics Unit
Eastman Dental Institute
University College London, UK

***Abstract*— In previous work we showed how a dense surface point distribution model of the human face can be computed and demonstrated the usefulness of the high-dimensional shape-space for tackling classification problems. In this paper we show how trajectories in the same space can be used to express the shape changes associated with growth and how average growth trajectories for the human face can be computed in the absence of longitudinal data by using kernel smoothing across a population. A training set of 3D surface scans of 199 male and 201 female subjects with ages between 0 and 50 is used to build the model.**

***Index Terms*— medical image registration, facial growth, deformable models, morphometrics**

I. Introduction

There are many changes in the shape of an individual's face over time. In young adults there is considerable growth of the skeletal structures, along with an increase in muscle tissue and changes in the volume of fatty tissues. In middle life there is little change in the bone structures but continued growth in cartilage, especially in men, affecting amongst other things the shape of the nose. In later life changes in both muscle tone and skin elasticity affect the outer shape of the face considerably. The modelling and prediction of these changes is a challenging task.

A previous paper [9] introduced dense surface point distribution models of the human face. The core of the procedure is the computation of a high-dimensional *face-shape-space* in which each example in the training set is represented as a point. Because each vertex in a dense mesh of the face is a landmark in the model, the shape-space is of sufficiently high dimensionality to capture subtle features of the face such as those associated with gender differences and facial dysmorphologies. This makes it useful for delineating between different subgroups in a population, as was described in [9]. This high-dimensionality of the shape-space also makes it useful for analysing continuous parameters, such as the age of an individual, which we will study in this paper.

While some studies have analysed three-dimensional longitudinal data [5], [2], to date there has not been much human data available. This is partly because the development of useful acquisition technologies has only occurred recently. Another factor is that x-ray and x-ray CT imaging as used in [5], [2] has an associated radiation dosage that precludes its use on healthy individuals.

Dean et al. [5] used three-dimensional landmarks acquired from the Bolton standards collection of frontal and lateral head radiographs taken repeatedly for the same individuals over time. They used Procrustes methods of shape analysis to register and compare sets of landmarks across the skull of 32 individuals (16

Correspondence to: T.Hutton@eastman.ucl.ac.uk

male, 16 female) at yearly intervals between the ages of 3 and 18.

Andresen et al. [2] used CT scans of the mandible of six children (4 male, 2 female) up to the age of 12. A surface model was built to compare the changing shape of the mandibles over time.

The type of data we use in this study comes from non-contact surface scanners that use natural, infra-red or laser light. The data consists of a dense polygonal surface representing the captured area, with a sample spacing typically of the order of between 0.1mm and 2mm across the face. Although we have data from a few hundred subjects the scanning techniques were only industrially developed over the last few years. It has therefore not (yet) been possible to carry out longitudinal studies. Without good longitudinal data we thus look at a method that computes an aging trajectory for the average individual, from a population that spans a large age range. If the pattern of growth is not significantly different between healthy individuals of the same sex, as was found in [5], then the average aging trajectory should capture the pattern of normal growth and could in theory be used to predict the change in shape of an individual's face with aging.

A trajectory in the space spanned by the principal components of shape variation (a shape-space) is a natural way of expressing growth. The technique is used in both [5] and [2] to show the correlation between age and the first principal component of shape variation.

In this paper, we first briefly review construction of the dense surface point distribution model for computing the face-shape-space [9] and then look at the use of kernel smoothing to compute growth trajectories in this space.

II. Materials and Method

The training set consisted of 400 face scans of different people, acquired on a DSP400 face-scanner (`http://www.3dmd.com`). Each scan results in a triangulated surface mesh model of the subjects face with between 2,000 and 10,000 vertices (see Fig. 2). The nature of the acquisition is such that structures other than the face area of interest are frequently included. The model-building method deals with these automatically as we shall show.

We restricted our study to the range of 0-50 years because of a relative lack of data on people aged above 50. Figure 1 shows the distribution of our 400 subjects across different ages. 199 of the subjects were male, 201 female. The people were from all ethnic backgrounds and most scans were taken with an approximately neutral facial expression.

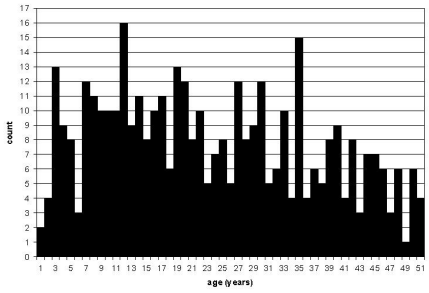

Fig. 1. Age distribution

A. Building the dense surface model

The first step is to place landmarks manually on each surface. We found that just 9 landmarks were enough to build a good model with our dataset. As described in [9], experimentation showed that too few (5) landmarks resulted in anatomically incorrect registration of scans, while too many (20) introduced noise into the model since there is insufficient visual guidance for their accurate placement. Figure 2 shows an example mesh with the landmarks overlaid.

The next step in building our statistical face-shape-space is to establish a dense correspondence between the surface meshes. This could be done using any set of landmarks as a frame of reference but it is desirable that the landmarks are typical of the shape distribution, so we have used the generalized Procrustes algorithm [7] to compute the mean landmarks. The key step in this process, used repeatedly in this Procrustes algorithm, is a least-squares alignment of two sets of 3D

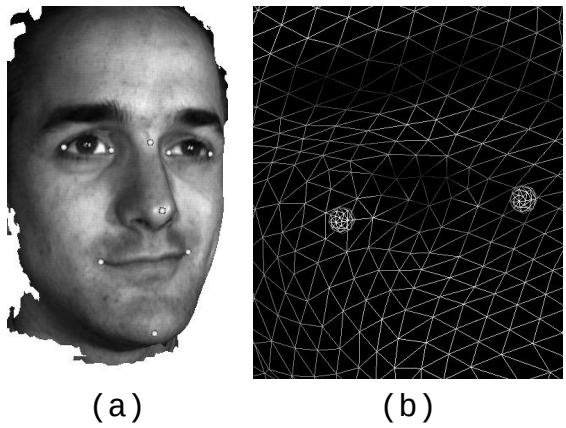

Fig. 2. An example scan with our nine landmarks *(a)* and a detail of the mesh around the eye with two landmarks visible *(b)*.

landmarks for which we use the quaternion method described for example in [8].

Each surface is then warped onto the mean landmarks using the thin-plate spline (TPS) technique [3]. This brings the landmarks into exact alignment and interpolates a smooth transform for the other parts of the mesh by minimizing a bending energy. This is intended to ensure that, while we take advantage of all the information implicit in the landmarks, as little spurious variation as possible is introduced, especially in the vicinity of each landmark.

Having brought all the surfaces into close alignment, a dense correspondence is made by taking the closest point on each surface from each vertex in a base mesh. We used as a base mesh one of the examples in the training set. Experiments showed that as long as the area of interest is covered by an adequate triangulation, the choice of which example to use to provide the mesh is not critical.

The scans in the training set (including the base mesh) often included significant neck and ear areas that were not present in all the examples. We snip off these areas by using only those vertices where the *maximum* distance from the base mesh to the surface of each scan (after alignment) is less than 20mm (found through experiment). While this distance is, of course, application-specific, this technique is very effective in restricting the model to those regions of the face surface that are well-represented in the training data. Essentially we are taking an *intersection* of the meshes of the subject's scans; keeping only those areas that are well covered in *all* the examples.

Additionally, the scans used in the training set often have occasional small holes or triangulation errors, causing them not to be locally manifold. Because our method uses one mesh to sample the others, it is robust to such errors - any vertex displacements caused by these errors tend not to be correlated with the major shape changes and are thus not represented in the principal components.

When the dense correspondence with a base mesh has been made, the connectivity of the base mesh can be applied to the points of correspondence in each mesh to give a set of new meshes. We can then dispose of both the landmarks and the original meshes. Finally, we apply the inverse of the aligning TPS warp to return each surface to its original location, carrying with it the new mesh derived from the base mesh. (While TPS does not have an analytical inverse, an approximation can be computed using Newton's method, see Acknowledgements section for an available implementation of inverse TPS.)

Now that we have constructed a dense mesh of corresponding vertices on the surface of each of the scans, we can treat all the vertices as landmarks. Following [4], we first apply the Procrustes algorithm to align all the shapes and produce a mean shape. Because our data is calibrated for size, we do not include scaling in the Procrustes alignment, but instead build a size-and-shape model [6].

A statistical model and face shape-space is constructed by applying a principal components analysis (PCA) to the data obtained from the above Procrustes alignment. As usual [4], each example can be represented by a shape vector of concatenated x, y and z coordinates for all n vertices:

$$\mathbf{x}_i = [x_1, y_1, z_1, \ldots, x_n, y_n, z_n]^T \qquad (1)$$

The mean shape vector is given by:

$$\overline{\mathbf{x}} = \frac{1}{s}\sum_{i=1}^{s} \mathbf{x}_i \qquad (2)$$

where $s = 400$ is the number of examples in the training set.

We compute the $3n \times s$ matrix $\mathbf{D}$ using:

$$\mathbf{D} = [(\mathbf{x}_1 - \overline{\mathbf{x}})| \dots |(\mathbf{x}_s - \overline{\mathbf{x}})] \quad (3)$$

And then, using the Eckart-Young theorem [10], we eigen-decompose not the usual $3n \times 3n$ covariance matrix but instead (for efficiency) the $s \times s$ matrix $\mathbf{T}$:

$$\mathbf{T} = \frac{1}{s-1}\mathbf{D}^T\mathbf{D} \quad (4)$$

to get the eigenvalues $\lambda_1 \dots \lambda_s$. From [10], the eigenvectors we are interested in are given by $\phi_i = \mathbf{D}\mathbf{e}_i$, where $\mathbf{e}_i$ are the eigenvectors of $\mathbf{T}$.

The computed eigenvectors ϕ_i can be treated as deformations of the whole mesh and can be added directly to the coordinates of the vertices of the mean mesh to synthesise new faces (see for example [4]):

$$\mathbf{x}' = \overline{\mathbf{x}} + \Phi\mathbf{b} \quad (5)$$

In (5), $\Phi = [\phi_1|\phi_2| \dots |\phi_s]$ is the matrix of the eigenvectors and $\mathbf{b} = [b_1, b_2 \dots b_s]$ is a set of parameters controlling the modes of shape variation. As usual, each eigenvector is normalized to unit length.

The coordinates in shape-space of any example can be found using:

$$\mathbf{b}_i = \Phi^{\mathrm{T}}(\mathbf{x}_i - \overline{\mathbf{x}}) \quad (6)$$

B. Computing average age trajectories using kernel smoothing

The examples in the training set can now be represented as points in a shape-space. Crucially, the *metric* of the space is a measure of the similarity in shape - the Procrustes distance. This means that intuitive operations such as interpolation and partitioning are meaningful.

Kernel smoothing is a standard statistical tool for filtering out high-frequency noise from signals with a lower frequency variation. Here we use the technique across the population of the training set to compute an average face for any given age, using weighted support from the examples that are close to the target age. The kernel serves both to interpolate between the examples (since we may have no examples at exactly the age we want) and to average out the variation due to individual subjects in which we are not interested and treat as noise.

Rather than compute an average aging trajectory for all the individuals we instead treat them as two groups - males and females. We do this because there is known to be a difference in the timing and types of facial growth between men and women and we would like to see this emerge from the data.

The path of the average aging trajectory is parameterised by the target age, t, and given by:

$$\mathbf{a}(t) = \frac{\sum_{i=1}^{n} w(\mathrm{age}_i, t)\mathbf{b}_i}{\sum_{i=1}^{n} w(\mathrm{age}_i, t)} \quad (7)$$

where n is the size of the population, age_i and $\mathbf{b}_i$ are the age and location in shape-space of the i'th subject.

We use a triangular kernel, defining:

$$w(x, t) = \max(1 - \frac{|x-t|}{\mathrm{width}}, 0) \quad (8)$$

but a Gaussian kernel was found to give very similar results.

The width of the kernel is a parameter that is determined by the amount of input data and its distribution. Smaller values tend to introduce noise into the trajectory since it becomes influenced by individual examples, while values that are too large will tend to smooth out the variation in which we are interested. For this dataset we found that a $\mathrm{width} = 20\mathrm{yrs}$ retained the difference in shape between the two groups while giving smooth curves. This seems rather large and perhaps reflects the fact that 200 faces per group over a span of 50 years is not very much data when trying to model the entire surface of the face at small time intervals. Larger populations would allow the use of smaller width values.

One effect of using kernel smoothing is that the width of the kernel limits how close to each end of the parameter range you can estimate. When we compute $\mathbf{a}(0)$ (new-born baby) we will get a weighted average that will

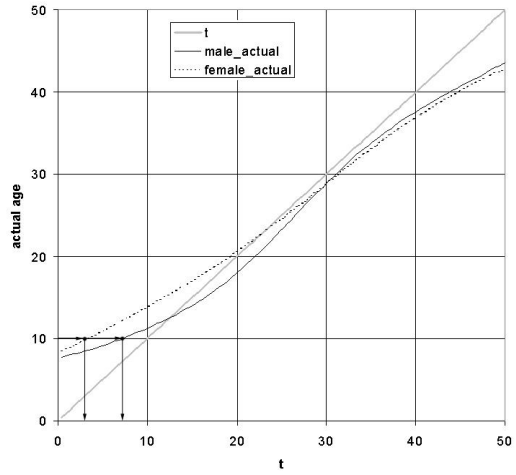

Fig. 3. A plot of t versus $\mathrm{age}(\mathbf{a}(t))$ shows how the extremes of the age-range get rounded-off. At the lower end the faces are older than their corresponding t values while at the upper end they are younger. We allow for this by using the value of t needed to the get the face of the correct age.

necessary be unevenly biased towards individuals from the population that are older than the target age since these are the only ones that were available to average across. This results in a face that is older than one would expect at the lower end, and a younger than expected face at the upper end of the age range. We can compute a value for a more realistic age of each average using:

$$\mathrm{age}(\mathbf{a}(t)) = \frac{\sum_{i=1}^{n} w(\mathrm{age}_i, t)\mathrm{age}_i}{\sum_{i=1}^{n} w(\mathrm{age}_i, t)} \qquad (9)$$

The plot of t versus $\mathrm{age}(\mathbf{a}(t))$ in Fig. 3 shows the typical rounding-off at each end of the age-range. We can use this relationship to infer the necessary t value for a desired age. For example, to get an acceptable 10-year old average face we need $t = 7.3$ for the male subgroup and $t = 3.4$ for the female subgroup. This difference is caused by varying numbers of examples at different ages for the two sexes. Table I summarises the values used to produce Fig. 6.

desired age (yrs)	t male	t female
10	7.3	3.4
15	16.4	11.9
20	22.0	19.3
25	26.5	25.5
30	31.0	31.4
35	36.4	37.4
40	43.6	44.8

TABLE I

TABLE OF DESIRED AGES AND THE NECESSARY VALUES OF t FOR THE MALE AND FEMALE SUBGROUPS.

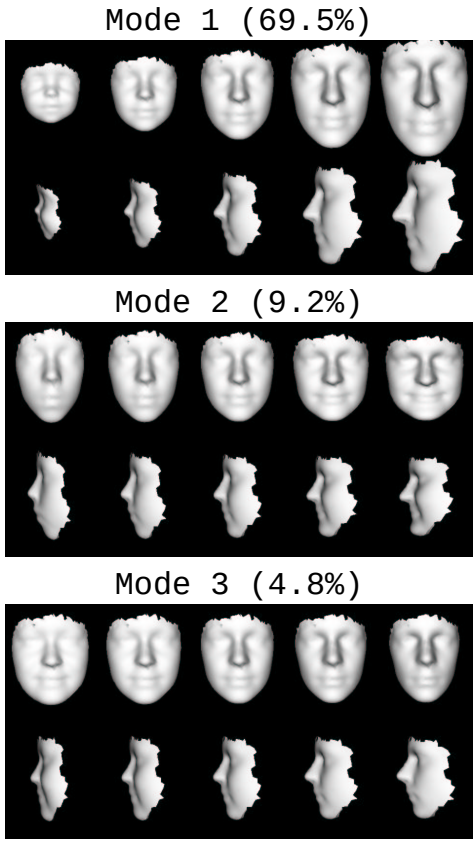

Fig. 4. The first three modes, between -3 and +3 standard deviations (columns), side and front views (rows)

III. RESULTS

A. Examining the model

Figure 4 shows the first three modes of our model based on all 400 scans (males and females together). Computation of this model takes roughly six minutes on a 1.7GHz machine. Synthesis of a new example takes under a tenth of a second. The first mode dominates (at 69.5%) because we left scaling in the model in order to capture the correlation between face size and shape.

The first mode shows a strong correlation with the age of the subject while modes 2 and 3 show changes in face shape that involve identity and facial expression. Mode 3

is quite strongly correlated with ethnic origin, with Asian faces being found on the left of this mode and Caucasian faces more on the right. The appearance of these modes is essentially unchanged from the previous study [9] even though the number of examples has doubled.

In this model, 43 modes are required to account for 98% of the total variation in our training set. Two previous models required 33 modes (193 examples) and 25 modes (72 examples) to account for the same 98% of the variation, showing how extracting the strong correlations can give a compact representation of the data.

A visual examination of these modes forms part of our evaluation of the efficacy of our method and the correctness of the correspondence. All of the images are plausible human faces and don't appear to have any artefacts such as ridges or distortion.

B. Visualizing the age trajectory

Figure 5 is a plot of the two trajectories we computed against the two strongest principal components, ϕ_1 and ϕ_2. The population is represented in the plot as a scatter of crosses and circles for males and females respectively.

Several features of this graph are worth noting. Firstly, a general divergence of the trajectories with age can be observed, with the two trajectories moving further away from each other as age increases. This correlates well with the accepted patterns of growth, where male and female face shapes become differentiated following puberty. Secondly, the trajectories for the years of growth up to 15 or so are approximately linear and correlated with the first principal component. This matches the findings of [5] and [2] (even though the datasets concerned are of different parts of the head). Later in life the shape-change follows a different direction, showing that over the entire lifetime the change in shape of the face is distinctly non-linear.

A third feature is that the male group shows significantly more movement along the first principal component than the female group. This mode of variation expresses an increase in size of the face, especially in the nose, eyebrows and chin and the top-right image in Fig. 4 shows a distinctly masculine appearance.

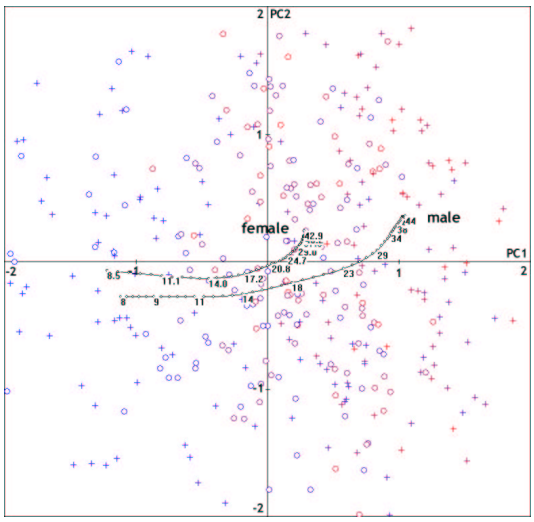

Fig. 5. The average growth trajectories for male and female, with labels marking the age in years. The axes are the first two principal components, expressed in standard deviations. The scatter plot in the background shows the overall distribution of males (crosses) and females (circles).

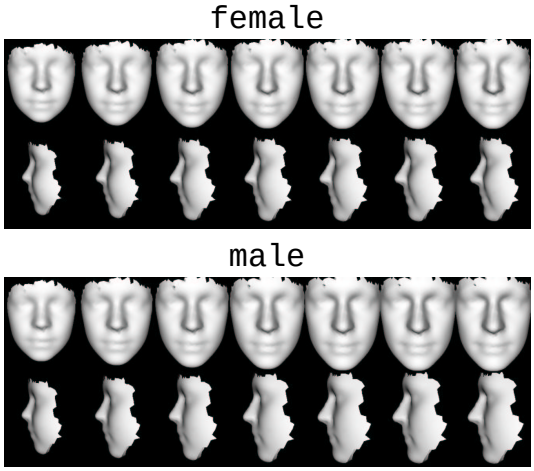

Fig. 6. Synthesised faces at ages 10,15,20,25,30,35,40 along the aging trajectories for females and males. The scale is the same throughout.

Plots of the trajectories against the other principal components show less obvious variation so have not been reproduced here. A PCA on the shape of the aging trajectory itself would help identify which modes if any show as much correlation with aging as the first two.

Figure 6 shows synthesised faces from along the two trajectories. The seven images correspond to ages 10,15,20,25,30,35 and 40 and were computed using the t values shown in table I.

IV. Summary and Conclusions

We have described a method for creating detailed deformable models from a set of surface meshes. The method described will work on surfaces of any topology, including those with holes and unconnected regions, as long as the topology is consistent across the training set. If the topology is not consistent then the surface used to provide the base mesh will be used to sample the others and may give incorrect results. The method will work with surfaces of regular or irregular triangulation, as well as with surfaces of varying densities of triangulation. Indeed, non-polygonal surface representations could also be used, as long as they are amenable to TPS warping and querying for finding the closest point on the surface. If required, a triangulation could be created to use as a base mesh.

The method is robust to occasional errors in the training set data such as noisy or missing vertices. Such errors are not likely to be correlated with the other shape changes and should therefore only appear in the less significant modes.

We have shown how separate growth trajectories for male and female subgroups can be computed in the absence of longitudinal data through the use of kernel smoothing. The faces from along the two trajectories correlate well with what we think of as male and female faces of different ages. The kernel smoothing tends to round-off the extremes of the age-range but we can correct for this by using the parameter value that gives the face of the age we want.

Having a model of the pattern of normal growth is essential for studying syndromes or pathologies that cause abnormal growth. One example is Noonan Syndrome, where the growth patterns of the face change over time [1].

An interesting open question is to what extent the trajectory for an individual over their lifetime would be parallel to the average. While we would expect the underlying patterns of growth to be broadly similar in all healthy individuals, factors such as ethnic origin, diet and lifestyle will have an effect.

Acknowledgements

The face scanner was acquired with funding from the Birth Defects Foundation (`http://www.birthdefects.co.uk`).

Many of the techniques used in this work are implemented in the Visualization Toolkit (VTK) (`http://public.kitware.com/VTK`).

Our thanks to Dr Henry Potts of the Cancer Research UK London Psychosocial Group, King's College, London for his valuable input. Thanks are also due to the many volunteers who have allowed their scans to be used in our work.

References

[1] J.E. Allanson, J.G. Hall, H.E. Hughes, M. Preus, and D. Witt. Noonan syndrome: An evolving phenotype. *Am. J. Genet.*, 21:507–514, 1985.

[2] P.R. Andresen, F.L. Bookstein, K. Conradsen, B.K. Ersbøll, J.L. Marsh, and S. Kreiborg. Surface-bounded growth modeling applied to human mandibles. *IEEE Transactions on Medical Imaging*, 19(11):1053–1063, 2000.

[3] F.L. Bookstein. Shape and the information in medical images: A decade of the morphometric synthesis. *Computer Vision and Image Understanding*, 66:99–118, 1997.

[4] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] D. Dean, M.G. Hans, F.L. Bookstein, and K. Subramanyan. Three-dimensional Bolton-Brush growth study landmark data: Ontogeny and sexual dimorphism of the Bolton standards cohort. *Cleft Palate - Craniofacial*, 37:145–156, 2000.

[6] I.L. Dryden and K.V. Mardi. *Statistical Shape Analysis*. Wiley, 1998.

[7] J.C. Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–51, 1975.

[8] B.K.P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4:629–642, 1987.

[9] T.J. Hutton, B.F. Buxton, and P. Hammond. Dense surface point distribution models of the human face. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (part of Computer Vision and Pattern Recognition)*, pages 153–160, Kauai, Hawaii, 2001.

[10] R.M. Johnson. On the theorem stated by Eckart and Young. *Psychometrika*, 28:259–263, 1963.