The Average Smile - Modeling Continuous Shape Change for Facial Animation

Julian J. Faraway
Technical Report # 386
Department of Statistics
University of Michigan

May 22, 2002

Abstract

The movement of landmarks on the human face can be recorded in 3D using motion capture equipment. We describe methods for the analysis of data collected on groups of subjects with a view to describing and assessing the differences between the facial motions of those groups. We focus on the smile motion in particular. The methods presented can be used more generally for continuous shape change data.

We introduce a novel parameterization of shape change that allows the parsimonious description of facial motion. We allow for a distinction between static facial shape and dynamic facial motion. We describe statistical methods for modeling differences in facial motion including a comparison of mean motions, principal components for describing the variation in motion and linear models for describing the effects of predictors.

keywords: functional data analysis, shape data, registration, B-splines

1 Introduction

Surgical procedures for repairing and correcting facial injuries and defects have advanced considerably in recent years. Surgeons are able to make impressive improvements to facial appearance. However, more than just a pretty face is needed — the face must function well too. Eating, speaking and facial gestures are an essential part of life. Function, as well as form, must be considered when evaluating surgical outcomes.

Cleft lip or palate is a relatively common birth defect. The standard treatment involves a series of surgeries through childhood to correct the problem. The incremental improvements in facial form tend to decrease with each successive surgery while increasing the risk of scarring and nerve damage that can degrade facial function. Decisions need to be made by the surgeon and parents about whether to continue. Aesthetic considerations tend to have greater weight than they perhaps

deserve and small improvements in form may be obtained at the cost of function. The full quantification of facial function that would allow for more informed decisions has not been available. In earlier work - Trotman, Faraway, and Essick (2000), we proposed some measures but the ones presented here represent a full step beyond to represent the full dynamic motion of the face.

Another type of surgery where additional understanding of facial function is needed involves persons with unusually small (retrognathic) or large (prognathic) jaws. Surgical techniques have been developed to modify jaw size closer to the norm but less is understood about how the movement of persons with retrognathic or prognathic jaws differs from the the norm and how surgery affects this movement. The data we analyze below is drawn from a study investigating these issues.

The methods we shall present below are useful for quantizing and modeling facial motion for clinical purposes. They also have some application to non-clinical applications in animation and in modeling shape change in general. We discuss these possibilities in the conclusion.

Three dimensional motion cannot be satisfactorily represented statically on a page. To appreciate our motion models, they must be seen in action. We have developed standalone viewer software that must be seen to understand the outcomes of our analyses.

2 Data

The sample analyzed below consisted of 48 healthy subjects recruited from the University of North Carolina School of Dentistry Orthodontic and Dentofacial Clinics. Some of the subjects had retrognathic and prognathic jaws. The extent of these variations in facial form can be quantified in various ways. We will use a measure called MFLF which represents the ratio of a midface distance (nasion to anterior nasal spine) to a lower face distance (anterior nasal spine to menton). There are other measures that would be considered in a full analysis including age and gender, but we aim to present only the method of analysis here. Subjects were asked to perform various standard facial motions — eye open, eye close, cheek puff, lip purse, grimace and smile. We shall focus on the smile only here although a full analysis would study the other animations. Subjects were instructed to bite on the back teeth and smile as much as possible. This is artificial but a previous study indicated that this smile had similar characteristics to natural smiles — see Mendez (1999). Each animation was repeated 3 times.

Data may be collected on facial motion using motion capture equipment. Twenty four retroreflective markers, each with a diameter of 2mm were attached using eyelash adhesive to specific sites on the face shown in Figure 1.

Some markers can be placed quite consistently. For example, the position of the commissures (lip corners — markers 16 and 19) is well defined. Other markers are placed on landmarks that are less well defined — for example, markers 13 and 14 on the upper lip. This is important in our choice of measures in the data analysis to follow.

The *Motion Analysis* video tracking system used four cameras to track the motion at a rate of 60 frames per second. All the motions were recorded for 3 seconds for a total of 180 frames per motion. At least two cameras must have a line of sight to a given marker for its position in three dimensions to be recorded. Due to the redundancy of having four cameras, missing data was relatively rare although occasional outliers are generated due to problems with tracking the data were encountered. Note that the head was not constrained so that the motion of face was

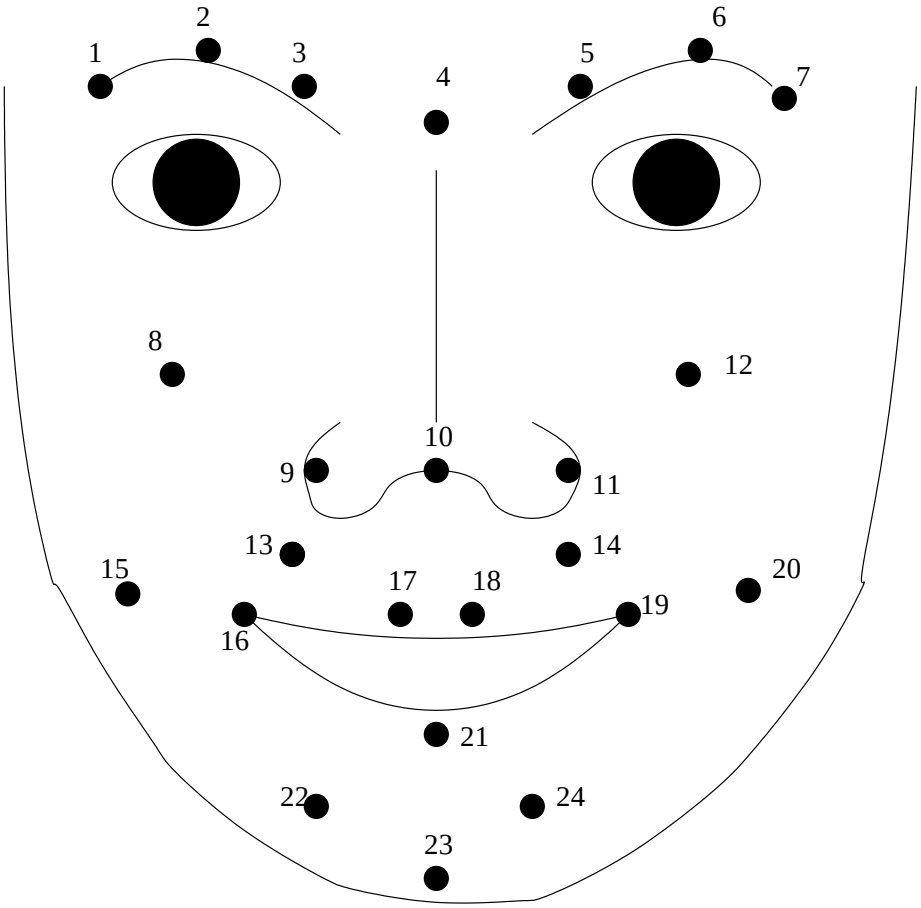

Figure 1: Schematic of facial marker placement.

confounded with motion of the whole head.

Each motion is represented by 24x3x180=12960 numbers and there were 144 smiles observed for a total of around 1.9 million numbers in the complete dataset. The size of the dataset has implications for the types of analysis that are computationally feasible and appropriate.

We show two smiles by the same person in Exhibit 1. (Please see Appendix for description of how to obtain and operate the viewing software. It's important to view the motion from the side as well as the front). In just these two observations, we see several important characteristics of the data. The head is not constrained. The initial positions are not exactly (nor could they be) aligned. The smiles take place at different times during the 3 second period. A small amount of noise due to measurement error is visible.

We have presented only a representative selection of analyses of this data. Our purpose is just to describe what is possible in a concrete manner but without attempting a comprehensive analysis of the data.

3 Methods

3.1 Shape

The configuration of the 24 markers constitute a shape. See Dryden and Mardia (1998), Bookstein (1991) and Small (1996) for introductions to statistical shape analysis. However, we are not so much interested in the shape of the face itself, but rather in how this shape changes when the face moves. We already know that facial shapes differ from person to person and what distinguishes those with retrognathic jaws from those with the normal jaws. Our interest is in the motion itself independent of the shape — essentially the derivative of the shape as it changes over time. This explains the necessity of developing new techniques.

We choose to base our analysis on the following type of measure: Let $d_{ij}(t)$ be the (Euclidean) distance between marker i and j at time t. Now let

$$r_{ij}(t) = \frac{d_{ij}(t)}{d_{ij}(0)} - 1,$$

which represents the relative change in the distance from rest. This measure has several desirable properties:

1. It is invariant to whole head motion.

2. It is approximately invariant to small variations in marker placement. As we mentioned above, the landmarks for several of the markers are not precisely defined. Because of the relative scaling, small variations in placement will only have a second order effect.

3. It is not dependent on local shape. For example, consider the distance between the commissures, 16-19. Of course, some people have bigger mouths than others but we have little interest in this. We are more concerned with how this distance changes during, say, a smile. By scaling using the initial distance, we remove much of the this effect.

In Lele and Richtsmeier (2000) and other research articles, an approach to statistical shape analysis based on d_{ij} over all pairs of distances called EDMA (Euclidean Distance Matrix Analysis) is developed. Given the matrix d_{ij}, it is possible to reconstruct the shape (up to rotation, translation and reflection). The important difference in our analysis is that we observe that not all pairs of distances are needed to reconstruct the shape. This is important because there are $n(n-1)/2$ pairs of distances given n markers. For n of any size, this is a substantial number of measures which is increasingly cumbersome for the types of analyses we demonstrate below. For the $n = 24$ in this example, there are 276 total pairwise distances for just a single frame of one motion.

We now describe how we reconstruct the face (shape) given only a subset of the pairwise distances:

1. Choose four landmarks. Given the six pairwise distances, we can reconstruct the position of these landmarks up to rotation and translation. Of the two possible reflections, we shall be able to choose the correct one given our knowledge of the relative positions of these landmarks on the face. The translation is irrelevant and the rotation, although arbitrary, can be chosen to place the face in, perhaps, an upright position for viewing.

2. Choose a new landmark and obtain the three pairwise distances of this landmark to 3 of the already positioned landmarks. Given this information, we can reconstruct the (irregular) tetrahedron of the one new and three old landmarks. Two possible tetrahedra satisfy the distance requirements. We shall rely on knowledge of the general shape of the face to select the correct one.

3. Keep adding new landmarks in the same manner until the face is complete.

The procedure is illustrated in Figure 2.

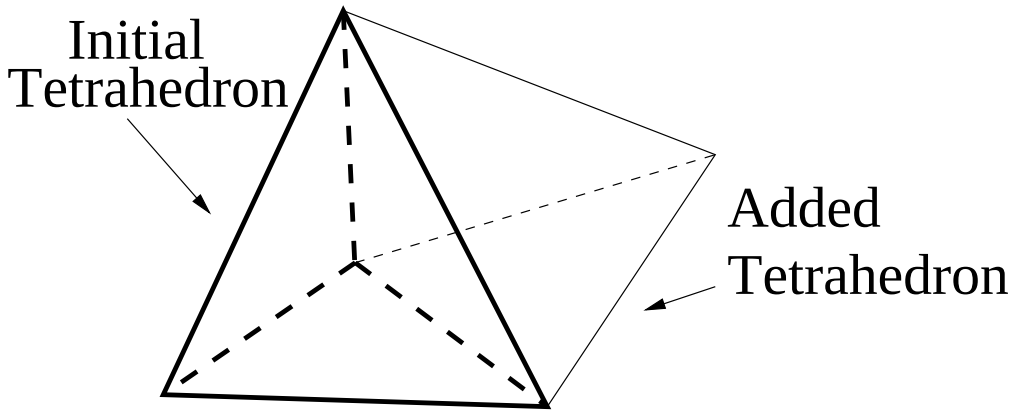

Figure 2: Six distances are required to construct the initial tetrahedron. Three additional distances are required to place each subsequent tetrahedron.

The particular order of reconstruction can be encapsulated in a matrix C_{ij} where $i = 1, \ldots, n-3$ and $j = 1, 2, 3, 4$. The first row of C indicates which four landmarks form the initial tetrahedron while successive rows indicate which 3 now known landmarks should be used to reconstruct the fourth new landmark. A total of $3n-6$ pairwise distances are needed. If we subtract an additional

parameter for a size factor we have $3n-7$ parameters — the same number that would be required if the usual shape coordinates were used. So we see that this is an alternative choice of coordinates adapted to the particular requirements of this type of data. Furthermore, we can be sure that no reconstruction would be possible with fewer pairwise distances.

This reconstruction algorithm has a speed advantage over that used on the full pairwise distance matrix. The latter requires a singular value decomposition on an n^2 matrix while only tetrahedra need to be reconstructed in our approach. Given that the algorithm will need to be applied repeatedly to reconstruct motion, this speed difference is magnified. Note also that the amount of computation and storage grows only linearly (compared to quadratically) in n for our method.

There are many possible choices of the reconstruction rule, C. Three considerations drive the particular choice:

1. The fourth point of each tetrahedron should not tend to be close to collinear with the other three points. Given the pairwise distances, there are two possible tetrahedra and we rely on our knowledge of facial shape to make the correct choice. In cases of near collinearity, the choice might become ambiguous.

2. We prefer to use pairwise distances between landmarks that are close because these will better preserve the objective of local shape invariance. Furthermore, some pairwise distances between adjacent landmarks represent particular muscles so it is natural to work with these.

3. We prefer that the tetrahedra be close to regular. If the pairwise distances become rather unequal, it is possible that it may not be possible to reconstruct a tetrahedron (the 3-D analogue of the triangle inequality is violated).

Even so, further refinement of the choice will be necessary. A selection of candidate reconstruction rules were evaluated for stability under random perturbations of the shape and a choice made.

3.2 Registration

The subjects are instructed to start from a neutral facial pose, assume the required pose, such as "smile", and then relax. This should be completed within 3 seconds starting from a cue to the subject that the cameras are recording. Subjects are given the opportunity to practice and if they miss the 3 second time slot, they have the opportunity to repeat the animation.

There are five phases to the motion:

1. Neutral pose

2. Assume the required pose

3. Hold

4. Relax from the required pose

5. Neutral pose

This means there are 4 transition times. These times will vary from animation to animation so it would not make sense simply to average several motions cross-sectionally in time. Motions need to be "registered" with each other so that comparable points in the motion are averaged. A further difficulty is that it is difficult to precisely identify these transition times.

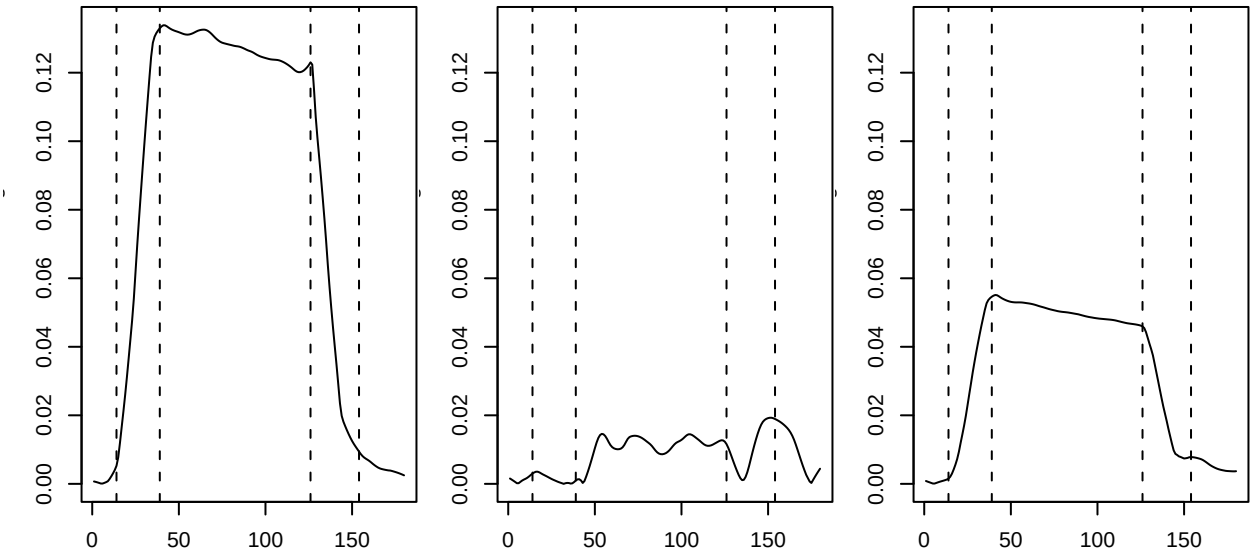

Figure 3: Selected relative change from rest for a smile. Left - 13-17 (Upper Lip). Center - 4-5 (Eyebrow). Right - mean. Selected transitions shown

We illustrate the issues in Figure 3, where information from a smile is depicted. The first panel shows a Lowess-smoothed $r_{13,17}(t)$ which represents a distance on the upper lip. The five phases of the motion are clearly identifiable although the precise transitions are not unambiguous. The center panel shows $r_{4,5}(t)$ which represents a distance above the eye. In this case the phases are not identifiable, not perhaps unsurprisingly as movement during a smile is confined to the lower part of the face while other unrelated activity may take place above the eyes. Clearly, we'd prefer to use the first plot to choose the transitions in this example. Unfortunately, patterns will differ from animation to animation and from individual to individual. What may be a good pair for one motion might not be for another. For this reason, we compute the average $r_{ij}(t)$ over all 66 pairwise distances as shown in the right panel and use this to select the transitions.

Various methods for automatically registering curves have been developed — see for example Ramsay and Li (1998). The drawback of automatic methods in our experience is that, although good ones may work well enough most of the time, they can fail badly when confronted with unexpected features. Given the outliers and missing values in the data, such anomalous cases are not uncommon. For this reason, we manually identify the transitions in the data. This was done anonymously to avoid bias in the selection. It would have been necessary to manually check the selections even had an automatic method been used so little time was lost and many errors avoided.

3.3 B-Spline representation

The curves shown in Figure 3 can be approximated by B-splines. We use standard cubic B-splines. We represent the angle curves as linear combinations of these basis functions, $B_j(t)$ for

$j = 1, \ldots, m$. The i^{th} curve $r_i(t)$ is represented as

$$r_i(t) = \sum_{j=1}^{m} R_{ij} B_j(t) + \varepsilon(t)$$

where the coefficients R_{ij} are found by minimizing a least squares criterion

$$\int_0^1 (r_i(t) - \sum_{j=1}^{m} R_{ij} B_j(t))^2 dt$$

Perhaps another more robust criterion could be used be we chose to remove outliers and other aberrant points earlier in the process and so our earlier smoothing makes this unnecessary in this case. The particular B-spline basis is determined by the choice of knot location. We evenly space our knots within the five phases described above. Furthermore, we know that $r_i(0) = r_i(end) = 0$. We can impose this restriction directly by omitting the first and last B-spline basis functions. The B-spline basis functions corresponding to Figure 3 with just one interior knot for each phase are shown in Figure 4.

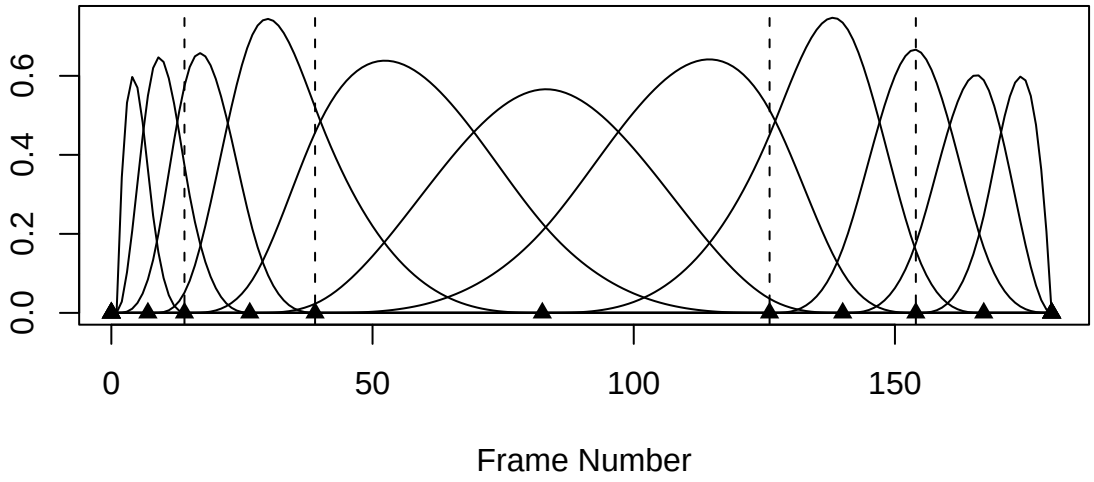

Figure 4: B-spline basis functions corresponding to transitions in Figure 3. Knot locations are shown on the horizontal axis. Note the zero values at the two endpoints.

Because the transition points will differ from motion to motion, the placement of the knots will also differ. However, we will be able to directly compare and compute statistics on the coefficients R with the assurance that $R_{i_1 j}$ and $R_{i_2 j}$ represent the same part of the motion. The knot positioning ensures the appropriate registration.

We chose m=16. This allows for 6 knots at the endpoints and transitions and two interior knots in the phases. Note that the choice of the number of knots is now an approximation rather than a smoothing issue. The observed data has already been smoothed — see Figure 3. We simply need enough knots to adequately approximate the curves without using more than necessary which would increase the computational and storage burden.

4 Statistics

For convenience, we unroll the matrix R_{ij} into a vector R_k where $k = 1, \ldots, m(3n-6)$ which represents one complete motion. For our choices of $m = 16$ and $n = 24$, we have a vector of length 1056. What follows are essentially statistical methods for such a multivariate response with some special adaptations and interpretations for this particular use.

Also, we have the distance between markers at rest $d_{ij}(0)$ which can also be unrolled into a vector d_k where $k = 1, \ldots, (3n-6)$ which represents the face at rest.

We can compute

$$d_{ij}(t) = d_{ij}(0)(1 + R_{ij}(t))$$

and thereby reconstruct the whole motion.

Most of the intermarker pairs used in the reconstruction rule C have a symmetric counterpart on the other side of the face. We have chosen to symmetrize the facial shape and motion by averaging these pairs in the examples that follow. Note that such symmetrization will not always be appropriate.

4.1 Means

It is straightforward to compute means within different subgroups in the data. For example, we may compute the average smile on the average face by simply averaging R_k and d_k over all the observed smiles. To compute the resulting motion, we must also specify the 4 transitions. We could compute the means of the observed transitions but for ease of comparison between different displays, we have set these transitions at $t = 1/6, 2/6, 4/6, 5/6$. Such a smile is shown in Exhibit 2.

It is not necessary to use the same subgroups for computation of the motion and the static face. We could compute

$$d_{ij}(t) = d_{ij}^A(0)(1 + R_{ij}^B(t))$$

where A and B represent means computed over different groups of individuals or even just a single individual. For example, we can impose the group average smile on the face of any given individual. Within this particular application, this would be useful for comparing the actual motion of a patient with what might considered normal motion.

We can also decompose the effects of static shape and dynamic motion. We might ask whether a particular subgroup differs from another because the motion or the shape or both are different. To illustrate this idea consider a division of our subjects into two groups based on their MFLF score. One group consists of subjects considered normal and the other with higher scores generally tending to indicate smaller jaws. We call the higher MFLF group, "patient" for ease of reference.

In Exhibit 3, we show a comparison of average smiles of normal subjects on the average normal face with the average smile of the patient subjects on the average patient face. We observe some clear differences in these motions (ignoring questions of statistical significance for now). But is the difference because the patient group have different shaped faces or because they smile differently? In Exhibit 4, we show a comparison of average smiles of normal subjects on the average normal face with the average smile of the patient subjects on the average *normal* face. The differences we now see are just those of motion which are far less substantial.

This observation has some practical significance since surgical techniques exist for extending the jaws of retrognathic patients. One might be concerned that altering the shape of the face would not be sufficient if the motion were abnormal. However, our exhibits suggest that the differences are due to abnormalities in shape, not in motion and that provided the surgery did not alter the motion of the patient, the outcome would be normal shape *and* motion. However, since this conclusion is derived from just a single type of motion (a smile) and a single measure of facial shape (MFLF) it is just speculation. A broader analysis is under way.

4.2 Variance

Of course, there is substantial natural variation in facial motion. We can describe the nature of this variation with a principal components analysis on the R_k. In this case, we have 1056 variables but even counting all the smiles separately, we have only 145 cases. Nevertheless, we may still compute the principal components. We find the percentage of variation explained by the first 5 components are 32.9, 11.1, 6.8, 5.7 and 5.0% respectively. We compute

$$\bar{R} \pm 2\sqrt{s_i}v_i$$

where s_i and v_i are the i^{th} eigenvalue and eigenvector respectively. We have applied these motions for the first principal component to the average face as shown in Exhibit 5. We see that the first principal component reflects the variation between people who tend to open their mouth when smiling as opposed to those who tend to keep their mouths closed.

The principal component scores are useful to identify to unusual motions. These can be used to detect faulty or exceptional motions that were not found at an earlier stage. Furthermore, these can be used to rate new patients with respect to a standard group. This provides a quantitative measure of abnormality and could also be used to objectively assess any changes due to surgery.

4.3 Inference

We could simply apply the standard techniques of multivariate analysis (see for example Johnson and Wichern (1992)) to the R_k. However, the dimension is large (1056 in our example) and so the power of such tests would be poor in that they would reflect unimportant differences between the groups arising in the smaller principal components. A similar problem of dimensionality with functional data was observed in Faraway (1997). Instead, we contend that it is both better and simpler to perform the inference on the first few principal component scores.

In Figure 5, we show the first three principal component scores plotted against MFLF. No relationship is apparent although some groupings of the three replicates per subject can be seen. We can fit a linear mixed effects model for the j^{th} replicate of subject i:

$$pc_{ij} = \beta_0 + \beta_1 \mathrm{MFLF}_i + \gamma_i + \varepsilon_{ij}$$

where γ_i is the random subject effect with variance σ^2_γ while within subject variation ε_{ij} has variance σ^2_ε.

As might be expected, the test $H_0 : \beta_1 = 0$ is not rejected (p=0.34). This confirms our earlier observation about the small difference in the two motions. We find $\hat{\sigma}_\gamma = 7.7$ while $\hat{\sigma}_\varepsilon = 6.9$. This

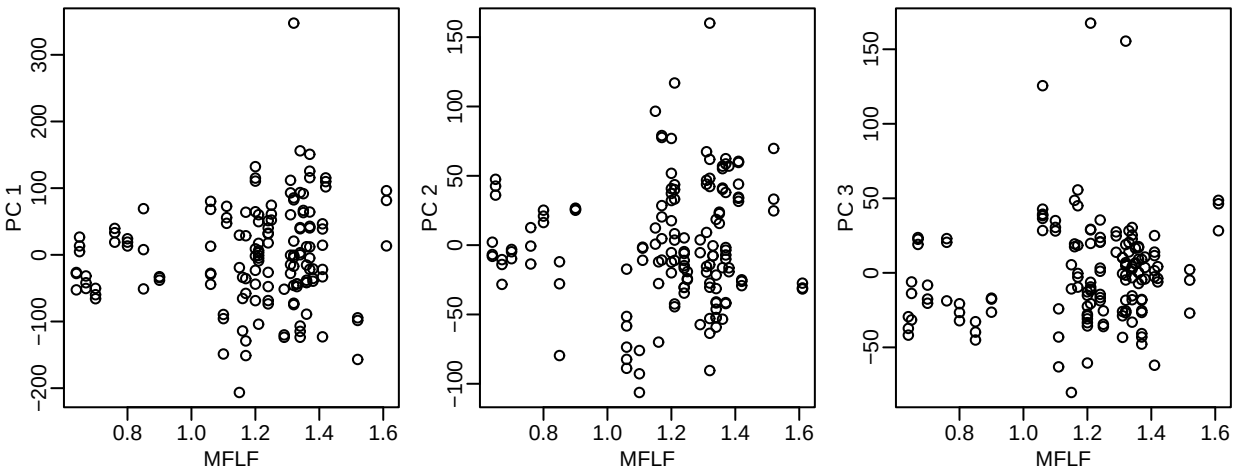

Figure 5: First three principal components of the smile against MFLF

indicates that while there is some consistency in smiles of the same subject, there is also a fair amount of variation in those smiles.

More extensive modeling and inference is possible but not explored here.

5 Discussion

For this particular application and other similar clinical problems, we have developed methods for

- Comparing and testing the dynamic motion of groups of subjects independent of their static facial shape.
- Describing and visualizing variation within the facial motions of groups of subjects
- Scoring the difference between individual subjects and the population motion

We also believe the analysis presented here has application to fields beyond facial modeling in clinical applications. Any data on continuously changing shapes might be modeled in a similar manner. However, some modification would be required for shapes that change substantially because of the necessity of being able to definitively choose between the two valid tetrahedra during the reconstruction process. For example, instead of using distances to three known landmarks for the reconstruction of each new point, four or more might be considered. This might reduce potential ambiguities and at the cost of some additional computation.

Acknowledgments

Dr. Carroll-Ann Trotman of University of North Carolina Dental School collected the data and motivated the analysis discussed in this article. This work was supported in part by grant DE 05215-2151 from the National Institute of Dental Research. The author also thanks the Department of Mathematical Sciences at the University of Bath which hosted the author during the period in which this article was written.

Appendix

Motion Viewer

Facial motions are difficult to represent statically and in two dimensions so they are best viewed dynamically. We have constructed a viewer that can display these motions that can be view at any angle. The viewer may be downloaded from

`http://www.stat.lsa.umich.edu/~faraway/face/`

The viewer keyboard commands are

- Function keys F1-F5 - Load Exhibit 1-5 respectively
- Arrow keys rotate the view
- a - show first (or only) face moving
- b - show second (if available) face moving
- c - show both (if available) faces moving
- </> - increase/decrease face size

List of Exhibits

1. Two smiles by the same individual.

2. The average smile.

3. Average face/smile of normal MFLF group compared with average face/smile of patient MFLF group.

4. Average face/smile of normal MFLF group compared with average face of normal with average smile of the patient MFLF group.

5. Smiles two standard deviations above and below the average in the direction of the first principal component.

References

Bookstein, F. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.

Dryden, I. and K. Mardia (1998). *Statistical Shape Analysis*. Wiley.

Faraway, J. (1997). Regression analysis for a functional response. *Technometrics 39*, 254–261.

Johnson, R. and D. Wichern (1992). *Applied Mulivariate Statistical Analysis* (3rd ed.). New Jersey: Prentice Hall.

Lele, S. and J. Richtsmeier (2000). *An Invariant Approach to Statistical Analysis of Shapes*. Chapman & Hall.

Mendez, M. (1999). A three-dimensional analysis of facial movement in normal adults: Methodological variation and characterization of natural expressions. Master's thesis, University of Michigan.

Ramsay, J. and X. Li (1998). Curve registration. *Journal of the Royal Statistical Society, Series B 60*, 351–363.

Small, C. (1996). *The Statistical Theory of Shape*. Springer.

Trotman, C.-A., J. Faraway, and G. Essick (2000). Three-dimensional nasolabial displacement in repaired unilateral and bilateral cleft lip and palate patients. *Plastic and Reconstructive Surgery 105*(4), 1273–1283.