# Accepted Manuscript

Please cite this article as: D.J.J. Farnell, H. Popat, S. Richmond, Multilevel Principal Component Analysis (mPCA) in Shape Analysis: A Feasibility Study in Medical and Dental Imaging, *Computer Methods and Programs in Biomedicine* (2016), http://dx.doi.org/10.1016/j.cmpb.2016.01.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Multilevel Principal Component Analysis (mPCA) in Shape Analysis: A Feasibility Study in Medical and Dental Imaging

DJJ Farnell, H Popat, and S. Richmond

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

## Abstract

**Background and Objective:** Methods used in image processing should reflect any multilevel structures inherent in the image dataset or they run the risk of functioning inadequately. We wish to test the feasibility of multilevel principal components analysis (PCA) to build active shape models (ASMs) for cases relevant to medical and dental imaging.

**Methods:** Multilevel PCA was used to carry out model fitting to sets of landmark points and it was compared to the results of "standard" (single-level) PCA. Proof of principle was tested by applying mPCA to model basic peri-oral expressions (happy, neutral, sad) approximated to the junction between the mouth/lips. Monte Carlo simulations were used to create this data which allowed exploration of practical implementation issues such as the number of landmark points, number of images, and number of groups (i.e., "expressions" for this example). To further test the robustness of the method, mPCA was subsequently applied to a dental imaging dataset utilising landmark points (placed by different clinicians) along the boundary of mandibular cortical bone in panoramic radiographs of the face.

1

**Results:** Changes of expression that varied between groups were modelled correctly at one level of the model and changes in lip width that varied within groups at another for the Monte Carlo dataset. Extreme cases in the test dataset were modelled adequately by mPCA but not by standard PCA. Similarly, variations in the shape of the cortical bone were modelled by one level of mPCA and variations between the experts at another for the panoramic radiographs dataset. Results for mPCA were found to be comparable to those of standard PCA for point-to-point errors via miss-one-out testing for this dataset. These errors reduce with increasing number of eigenvectors/values retained, as expected.

**Conclusions:** We have shown that mPCA can be used in shape models for dental and medical image processing. mPCA was found to provide more control and flexibility when compared to standard "single-level" PCA. Specifically, mPCA is preferable to "standard" PCA when multiple levels occur naturally in the dataset.

**Keywords:** multilevel PCA; active shape models; dentistry

Corresponding Author: Dr. Damian JJ Farnell, farnelld@cardiff.ac.uk

2

# Introduction

Methods used in image processing should reflect the underlying structures not only within the images themselves but also between images. This is particularly important for repeated images taken from the same person or different "patches" from the same image. Methods that reflect these relationships between images at multiple levels, should function more efficiently than those that do not shape models (ASMs) and active appearance models (AAMs) [1-8] are a common technique of image processing that are used to search for specific features or shapes in images.

Central to these methods is the principle components analysis (PCA) of sets of "landmark" points that define a specific shape in a set of images. Often these sets of points for the images are identified or defined by an expert (or experts) using a graphical user interface (GUI). For example, Fig. 1 shows a typical landmarked medical image and GUI for a panoramic dental radiograph [8].
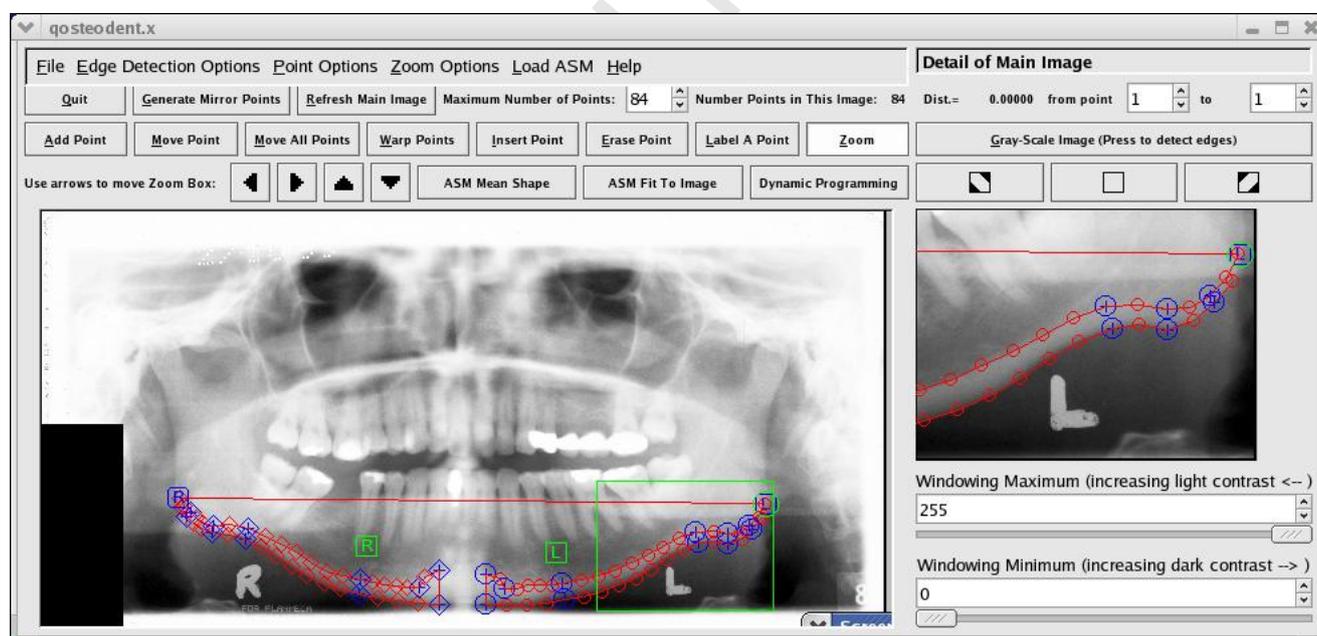


**Figure 1:** Mark-up points on a panoramic dental radiograph as part of the OSTEODENT project (see Ref. [8]).

Once an image has been landmarked, the principle components form a point distribution model and one uses this information to detect shapes in new images. A distinct advantage of ASMs is that one

3

constrains the types of shapes that are allowed in any image search. This is achieved by allowing any image search to go only so far along directions defined by the eigenvectors (major modes of variations) based on PCA, e.g., for ASMs, of a covariance matrix for sets of landmark points. Therefore any shape found in the new image using this approach is never too extreme when compared to the training set, which is very useful for medical images for which artefacts and noise are common.

If clustering or multilevel data structures exist naturally in the data set, the eigenvectors and eigenvalues from PCA will only be partially reflective of the true variation in the set of shapes. Multilevel modelling is a method of statistics that has been used extensively, e.g., to model naturally occurring clusters in the data or to model the effects of repeated-measures in longitudinal studies. A previous application of mPCA in ASMs was for segmentation of the human spine [9]. The results of this study showed that mPCA offers more flexibility and allows deformations that classical statistical models cannot generate. As mPCA decomposes data into a within-individual and a between-individual component clear benefits to the analysis of multi-level images is seen.

This study takes Monte Carlo simulations (called the "smile" dataset as a shorthand) and data from the OSTEODENT project [8] to explore an application of mPCA to a practical problem. The OSTEODENT project aims to investigate the link between dental measurements (e.g., bitewing and panoramic radiographs) and osteoporosis, and the reader is referred to Refs. [8,10] for more detailed information about the project. The present work follows on to some extent from previous analyses of placement errors of landmark points [11] for the OSTEODENT dataset. The focus of this previous work was to use the OSTEODENT data and also simulated data in order to explore the effects of measurement errors [12-14] on ASMs [11]. Although there are some parallels with this previous work, mPCA is the correct method for those cases where multiple levels occur naturally in the data, and where one wishes to PCA in models of shape. For the OSTEODENT project, mPCA describes

4

mark-ups of landmark points in these images for all subjects as one level in the data. The mark-ups are provided by two independent experts for these images and therefore the "experts" provide the next level. The approach of mPCA discussed in Ref. [9] is used to study the effects of multiple levels in the dental images. The aim is to test if it is possible to apply such methods to topics in dentistry and if clinically applicable results can be obtained when compared to results of "standard" single-level PCA. Finally, we explore practical issues relating to mPCA such as how the number of mark-up points, the number of shapes, and the number groups will affect the results. The formalism for mPCA is presented in the methods section and an explicit calculation of covariance matrices and mPCA is given in an appendix for a very simple example. Results for the application of mPCA to the "smile" and OSTEODENT datasets is then given. We present our conclusions in the final section.

5
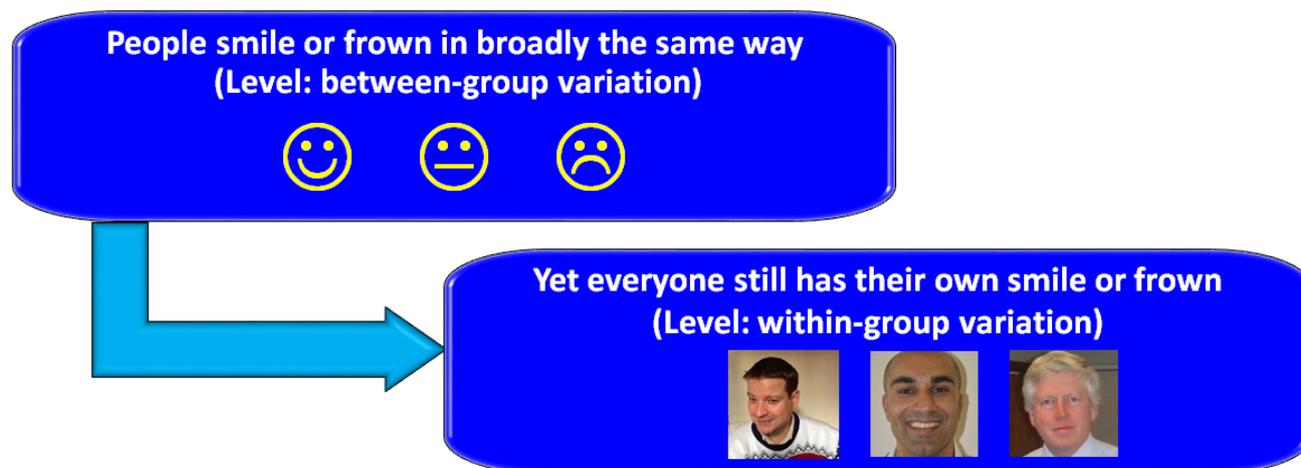
# Methods



**Figure 2:** Flowchart illustrating the "nested" nature of multilevel data.

*Mathematical Formalism*

The ASM method has been extensively documented in the literature (see, e.g., Refs. [1-8]), and therefore this topic is not discussed here. Instead, we use the approach of mPCA discussed in Ref. [9] in order to study the effects of multiple levels in the dental and medical images in ASM image searches. A flowchart illustrating the nature of multilevel data is presented in Fig. 2. Note that the within-group level might be thought of as being "nested" within the broader between-group level. We reflect this structure explicitly in multilevel methods. For example, we carry out PCA at both within-group and between-group levels independently for mPCA, as explained in Appendix A. Apart from Ref. [9], the mathematical formulation and implementation of the mPCA approach for ASMs has not been discussed extensively in the literature. Therefore, the mathematical formalism and analytic example of mPCA is also given in Appendix A and B respectively. As well as a feasibility study of this method to problems in dental image analysis and research, the explicit presentation of the mathematical formalism and the numerical solution of resulting equations is itself a "novel" aspect of the present work.

6

*The "Smile" Dataset*

One of the simplest idealisations of a "true problem" that one might consider in dentistry is that of peri-oral expression, i.e., the centre line of a lips. A quadratic function $y = x^2$ is used to represent the centreline of the lips. Monte Carlo sampling of a normal distribution is used to model expression for up to $K$ groups. For example, we use $K = 3$ and so we might interpret the groups as (say): $j = 1$, "very happy" $c = N(0.6, 0.01)$; $j = 2$, "neutral" $c = N(0, 0.01)$; and $j = 3$, "very sad" $c = N(-0.6, 0.01)$. Within-group variability is modelled by varying the width of the lips, where: $width = N(1, 0.2)$ and $x = [-0.5 \times width, 0.5 \times width]$. It is also a simple matter to produce data for additional "test cases" under the same conditions as those used for the training set. We compare results of mPCA with "standard" PCA, where no distinction between within-group and between-group variability is made. All calculations for the Smile and OSTEODENT data sets were carried out by using MATLAB.

*The OSTEODENT Dataset*

Monte Carlo simulation has the advantage that we can precisely control the nature of the shapes being employed in the model, and that error and artefacts do not occur unless we wish to build them into the simulations explicitly. However, such simulations are also idealised versions of shapes seen in the "real world" and so they might not reflect the "true variability" of these shapes. Furthermore, common artefacts or noise might be neglected or ignored. It is therefore wise to test the method against a "real" dataset.

A secondary analysis of the OSTEODENT "shape data" is shown by way of illustrating the method. 133 female patients aged 45-55 attended the University Dental Hospital of Manchester for routine dental treatment [10]. Mark-up points were placed on the cortical bone edges of panoramic mandible images (8000 pixels x 4000 pixels) [8] by $K = 2$ two experts separately using a custom-written GUI (see Fig. 1). There were $n = 84$ mark-up points in total. All shapes were scaled in size by the width of

7

the first shape so that the overall scale is of order 1, although differences in orientation was not corrected. No other pre-processing steps were carried out.

The shape of the cortical bone edges across all subjects therefore provides the within-group variability and any systematic biases and/or random error in placement of the mark-up points by the two experts gives the between-group variability. Such systematic biases and random errors are expected to be small in comparison with shape changes between individuals in the dataset, and so we expect within-group effects to dominate in this case.

8

# Results

## *Monte Carlo Simulated "Smile" Dataset*
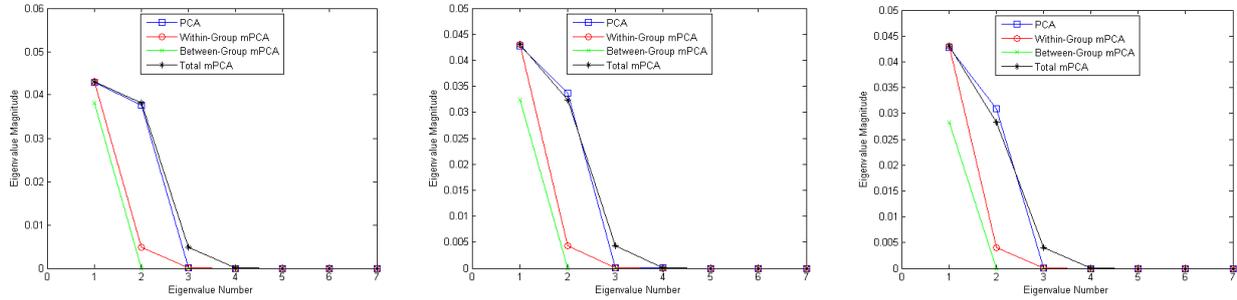


**Figure 3:** Eigenvalues for the "smile" dataset for $n$ = 11 mark-up points, $N$ = 200 shapes per groups, and for (left) $K$ = 7 groups, (middle) $K$ = 11 groups, and (right) $K$ = 21 groups.

Eigenvalues for standard PCA and mPCA for $n$ = 11 points, $N$ = 200 shapes per group are shown in Fig. 3. Only two large eigenvalues are found for PCA, whereas mPCA, indicates that three large eigenvalues might occur. We see that the eigenvalues for the within-group and between-group covariance matrices agree with the first two eigenvalues of PCA in this case. Importantly, the largest eigenvalue for the within-group variability (i.e., relating to the width of mouth/lips) is larger than that of the between-group variability (i.e., expression) for $K$ = 7, 11, and 21. Interestingly, the first eigenvalue for the within-group matrix reduces in magnitude with increasing $K$, which is due to the fact that increasing $K$ will reduce magnitude of variability in the coefficient $c$ in the Monte Carlo simulations (e.g., $K$ = 3, s.d. of $c$ = 0.495, $K$ = 5, s.d. of $c$ = 0.426, and $K$ = 7, s.d. of $c$ = 0.401). Although the accuracy of the covariance matrices in expected to increase with increasing $K$, our results are remarkably similar for all values for $K \geq 3$ groups. Indeed, results appear similar to larger values of $K$ even for the extreme limit of $K$ = 3. Profiles for the cumulative proportion of variability are shown in Fig. 4 for $K$ = 7, 11, and 21. The cumulative proportion of variability for between-group effects has a value of 1 after a single eigenvalue, whereas the cumulative proportion of

9

variability for within-group effects has a value of 1 after a two eigenvalues. Results for the profile of

the cumulative proportion of variability are similar for $K \geq 3$ groups.
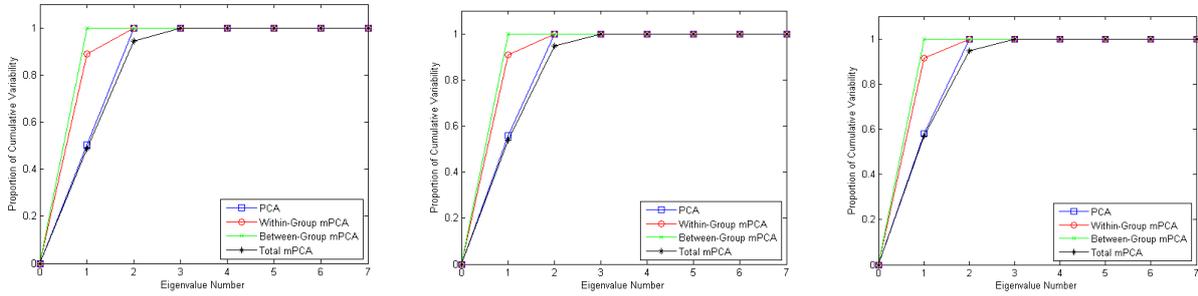


**Figure 4:** Proportion of cumulative variability for the "smile" dataset for $n = 11$ mark-up points, $N = 200$ shapes per groups, and for (left) $K = 7$ groups, (right) $K = 11$ groups, and (right) $K = 21$ groups.

The broad profile for the eigenvalues and cumulative proportion of variability do not change greatly

as one changes the number of points, $n$, and the number of shapes in each group, $N$. For example, the

number of large eigenvalues for both PCA and mPCA remains the same. Clearly though, sampling

error will decrease as $N$ is increased and so the covariance matrices will become more accurate. In

other words, reducing $N$ increases sampling error and eigenvalues therefore become less accurate

compared to their theoretical "population" values. Similarly, changing the number of points does not

greatly affect the number of large eigenvalues, although the overall magnitude of eigenvalues is

found to increase with increasing numbers of points, $n$, as one might expect.
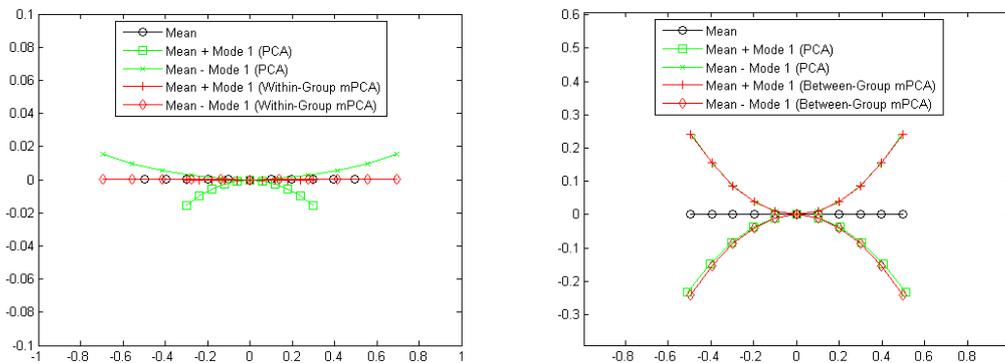


**Figure 5:** Major modes of variation for $n = 11$ mark-up points, $N = 200$ shapes per groups, and $K = 5$ such group for the "smile" dataset: (left) within-group effects relating to changes in scale (i.e., the width of "mouth"). (right) between-group effects relating to changes in expression (i.e., happy $c > 0$, neutral $c = 0$, and sad $c < 0$).

10

The major modes of variation are explored in Fig. 5 for number of points $n =$ 11, number of shapes per group $N = 200$, and $K = 5$ such groups. The mode of variation from the within-group terms for mPCA are compared to the first major mode from standard PCA, and both predict changes in the width of the mouth/lips. However, we see that the major mode from PCA also has some element of "between-group" variation in it (e.g., changes in curvature). Between-group terms in mPCA are also shown in in Fig. 5 for $K = 5$ groups and the results are compared to the second major mode from standard PCA. The mode of variation from the between-group terms for mPCA correctly predicts changes in the curvature of the lips (i.e., variability in coefficient $c$), as required.
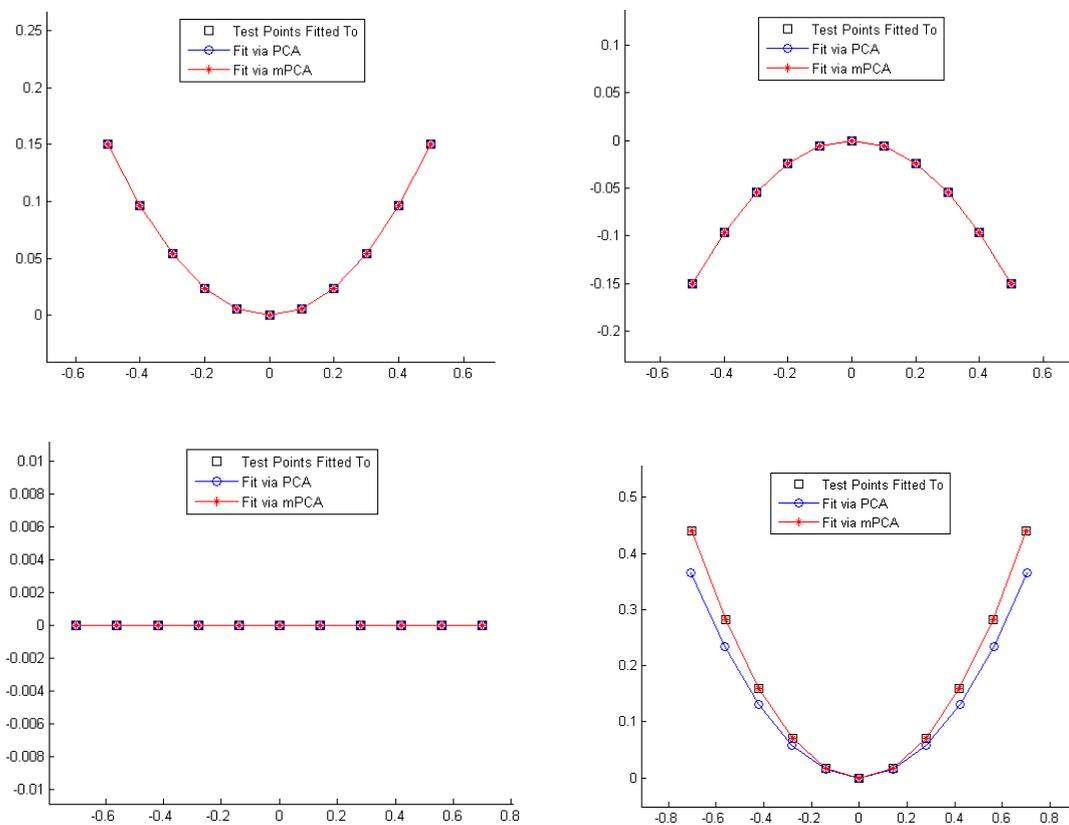


**Figure 6:** Fits of PCA ($M = 4$) and mPCA ($M^w = 3$ and $M^b = 1$) to "typical" test shapes for the "smile" dataset for $n = 11$ mark-up points, $N = 200$ shapes per groups, and $K = 5$ such groups: (top left) "very happy" ($c = 0.6$, *width* $= 1$); (top right) "very sad" ($c = -0.6$, *width* $= 1$); (bottom left) "neutral" though extremely "wide" ($c = 0.0$, *width* $= 1.4$); and, (d) extremely "happy" <u>and</u> extremely "wide" ($c = 0.9$, *width* $= 1.4$).

11

Results of a model fit for PCA ($M = 4$) and mPCA ($M^w = 3$ and $M^b = 1$) to "test shapes" are shown in Fig. 6 again for number of points $n = 11$, number of shapes per group $N = 200$, and $K = 5$ groups. We see that both mPCA and PCA provided adequate fits to the test points for cases that are not too extreme i.e., moderately "curved" examples for $c = 0.6$ and $c = -0.6$ (*width* = 1 in both cases). Both mPCA and PCA provide a good fit to the data points for an extremely wide, albeit neutral expression ($c = 0$ and *width* = 1.4). However, we see that only mPCA provides an excellent fit to the data points for a case that is extreme (albeit still possible in the "training" dataset) in both width and curvature ($c = 0.9$ and *width* = 1.4). By contrast, PCA provides only a poor fit for this extreme case. Indeed, it underestimates the level of curvature of this extremely and "wide" and "happy" expression. These results were typical of all values of $n$, $N$, and $K$ considered in the simulations.

The inclusion of normality distributed random "placement error" in the landmark points was also investigated for the Smile dataset. This source of random error was added to the training set of 1000 shapes used in forming the basic models with for $n = 7$ landmark points and $K = 5$ groups. The average point-to-point errors evaluated over test set of 1000 shapes are given in Table 1. The point-to-point errors are seen to increase with the magnitude of error added to the training set, although they reduce with the number of eigenvalues/vectors retained. Results of PCA and mPCA are broadly comparable in terms of point-to-point errors for all cases.

12

| Magnitude of placement error in training set | $M$ | PCA | mPCA with $M^w = M - 1$ & $M^b = 1$ | mPCA with $M^w = M$ & $M^b = M$ |
|---|---|---|---|---|
| 0.00 | 2 | 0.0000 ± 0.0000 | 0.0000 ± 0.0000 | 0.0000 ± 0.0000 |
| | 4 | 0.0000 ± 0.0001 | 0.0000 ± 0.0001 | 0.0000 ± 0.0001 |
| | 8 | 0.0000 ± 0.0001 | 0.0000 ± 0.0001 | 0.0000 ± 0.0001 |
| 0.025 | 2 | 0.0153 ± 0.0001 | 0.0153 ± 0.0001 | 0.0151 ± 0.0001 |
| | 4 | 0.0145 ± 0.0001 | 0.0148 ± 0.0001 | 0.0144 ± 0.0001 |
| | 8 | 0.0080 ± 0.0000 | 0.0096 ± 0.0000 | 0.0079 ± 0.0000 |
| 0.05 | 2 | 0.0304 ± 0.0001 | 0.0305 ± 0.0001 | 0.0300 ± 0.0001 |
| | 4 | 0.0284 ± 0.0001 | 0.0296 ± 0.0001 | 0.0271 ± 0.0001 |
| | 8 | 0.0206 ± 0.0001 | 0.0211 ± 0.0001 | 0.0206 ± 0.0001 |

**Table 1:** Mean point-to-point errors (± standard errors of the mean) were found for $n = 7$ landmark points with $K = 5$ groups and $N = 1000$ training shapes and for $N_{test} = 1000$ "test" shapes. $M$ is the number of eigenvalues/vectors retained for PCA, and $M^w$ and $M^b$ are the number of within and between eigenvalues/vectors retained for mPCA, respectively.

### The OSTEODENT Dataset
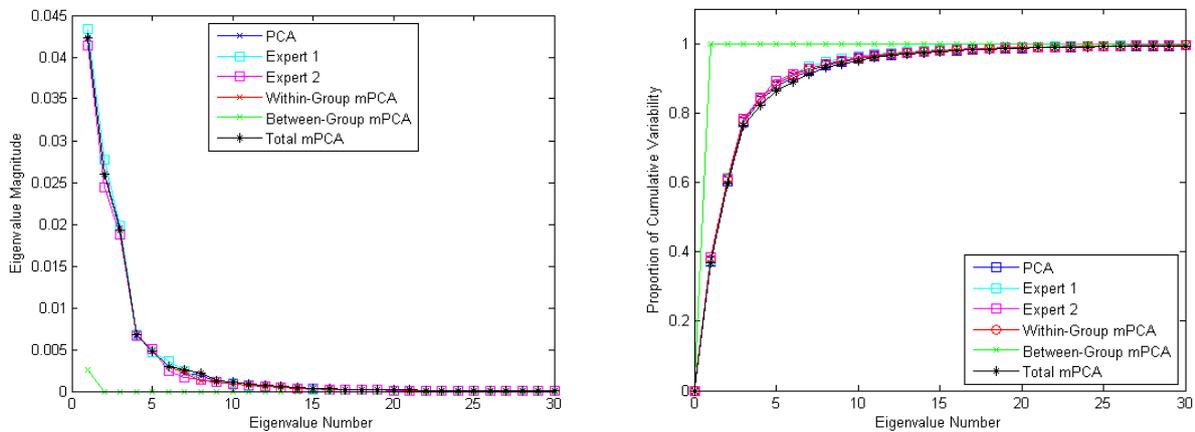


**Figure 7:** Results of PCA and mPCA for the OSTEODENT dataset ($n = 84$, $m = 133$ and $K = 2$): (left) eigenvalues; and (right) proportion of cumulative variability.

Eigenvalues and cumulative proportion of variability for standard PCA and mPCA for the OSTEODENT are shown in Fig. 7. We see that the eigenvalues for within-group terms are an order of magnitude larger than the eigenvalue for the between-group terms. The result is logical because we expect those variations between experts to be much smaller in magnitude than the variations in shape of the mandible between subjects. Indeed, only a single "large" eigenvalue is seen for the

13

between-group terms for mPCA, as expected for only two sets of landmark points for two "raters" or "experts." The results for PCA agree with results of the within-group terms and also "total mPCA" for both the eigenvalues and proportion of cumulative variability.
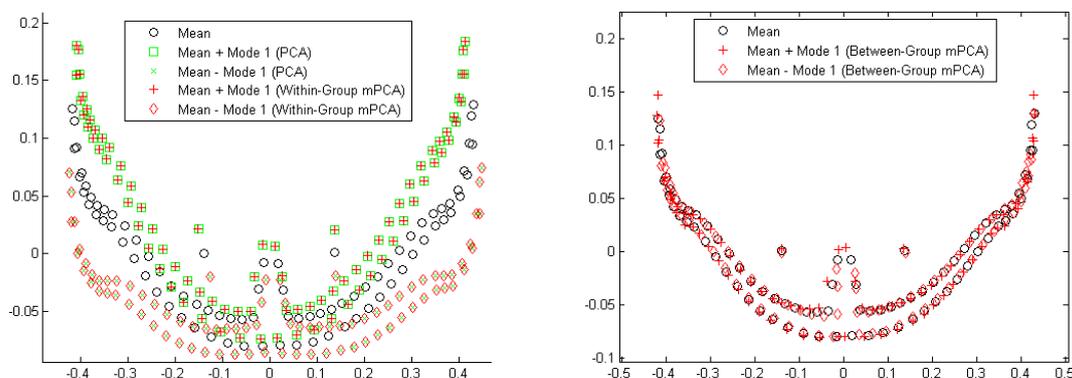


**Figure 8:** Major modes of variation for the OSTEODENT dataset. (left) within-group mPCA relating to variability in shape between subjects; (right) between-group mPCA relating to variability in placement of points along the edges of the feature between the two experts.

The major modes of variation are shown in Fig. 8. The within-group terms from mPCA correctly capture variations in shape between subjects, as expected. The first major mode from "standard" PCA agrees well with the first major mode from within-groups mPCA. It has been noted anecdotally for this dataset previously [8] that both experts were able to position landmark points perpendicular to strong edges with accuracy, although they found to harder to place points these edges in a consistent manner. The major mode from between-groups mPCA correctly identifies this inter-expert variability because we see that the major mode clearly corresponds to variability along these edges. No equivalent mode from "standard" PCA can be plotted in this case.

14
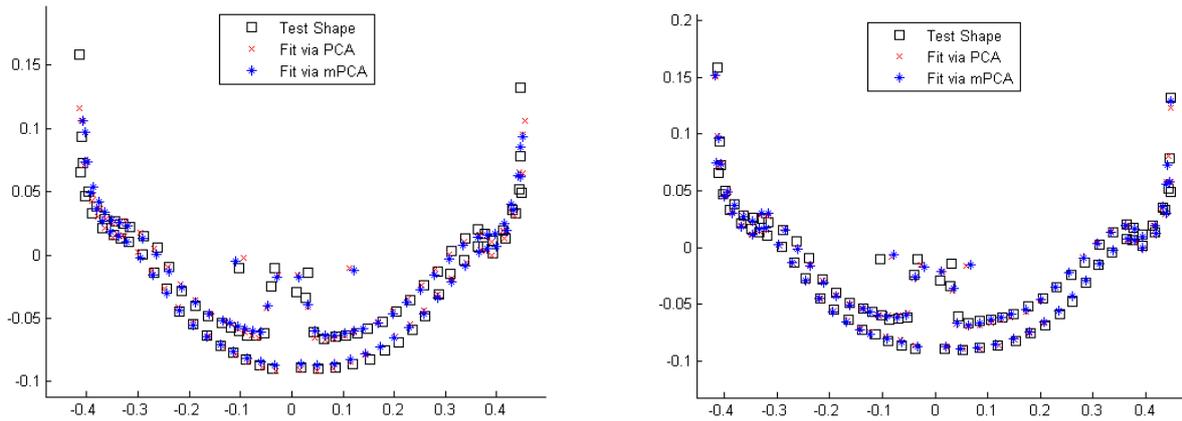
**Figure 9:** Fits of PCA and mPCA to the first test shapes for the OSTEODENT dataset using "miss-one-out" testing (i.e., the model is formed on all other shapes in the dataset excepting this specific "test shape"). (left) PCA ($M = 12$) and mPCA ($M^w = 11$ and $M^b = 1$); and, (right) PCA ($M = 30$) and mPCA ($M^w = 29$ and $M^b = 1$).

Finally, model fits of PCA ($M = 12$) and mPCA ($M^w = 11$ and $M^b = 1$) to the first shape for the OSTEODENT dataset using "miss-one-out" testing (i.e., where the model formed on all other shapes in the dataset excepting this specific "test shape") is given in Fig. 9. Results for PCA ($M = 12$) and mPCA ($M^w = 11$ and $M^b = 1$) are shown in the left-hand figure, and we see that both approaches give an adequate fit to the test data. However, there is some error in the model fit for the more "extreme" points in this test set, e.g., for the mandibular condyle for the left-hand figure. The average point-to-point error evaluated over all points and all shapes using miss-one-out testing is equal to $0.0055 \pm 0.0015$ for both PCA and mPCA. Results for PCA ($M = 30$) and mPCA ($M^w = 29$ and $M^b = 1$) for the same shape produce a much better fit to the test points, as also shown in Fig. 9. Furthermore, the average point-to-point error evaluated over all points and all shapes using miss-one-out testing is now equal to $0.0029 \pm 0.0011$ for both PCA and mPCA. As expected, the point-to-point distances reduce as we increase $M$ for PCA and also $M^w$ and $M^b$ for mPCA.

15

## Discussion

The effects of multiple levels of structure in shape data have been explored in this article, particularly in relation to dental image analysis research. The formalism for mPCA has been given and it was shown that mPCA allows between- and within-group variations to be modelled correctly. We have shown that mPCA allows us to model variations at different levels of structure in the data and an explicit calculation of within-group and between-group covariance matrices used in mPCA (Appendix A and B). We have also demonstrated that the initial results of mPCA from the OSTEODENT project can be transferrable clinically. This is an excellent first step in evaluating the usefulness and feasibility of mPCA in dental image analysis.

For example, it was seen that mPCA successfully managed to decompose the within-group variation (variation in the width of lips between subjects) and between-group variation (e.g., $K = 3$ groups: very happy, very neutral, and sad) for the Monte Carlo simulated "smile" dataset. We were able to determine the relative importance of these two sources of variation in the data. For $K \geq 5$, it was found that within-subject effects were slightly more important than between-subject effects by inspection of the pattern of eigenvalues. Note that meaningful information could still be extracted even for very small numbers of groups such as $K = 3$ groups. For example, major modes of variation for within-groups effects correctly identified changes in the width of the line (between subjects) and major modes of variation for within-groups effects correctly identified changes in the curvature of the line even for extreme limiting case of $K = 3$ groups.

Monte Carlo simulated datasets run the risk of being too simplistic when compared to "real-world" problems. It is therefore necessary to apply such methods also to "real" datasets. Shape data from the OSTEODENT project was used here for this purpose. It was found by inspection of eigenvalues and cumulative proportions of variability found that between-group effects were very much smaller than within-group effects. The first major mode for between-group effects was seen to model differences

16

in shapes between subjects and the major mode for within-group effects was seen to model differences in shapes between experts (i.e., placement of points along strong edges). This is a somewhat remarkable result considering that only two groups (i.e., experts) were used in this study of the OSTEODENT data. The major modes for the within-group and between-group covariance matrices were still seen to give meaningful results and insight into the data.

The usefulness of Monte Carlo simulation is that one can consider quickly the effects of the number of groups, number of shapes per group, and number of mark-up points on results of a new technique. These simulations showed that increasing the number of shapes per group reduced sampling error and so results for the eigenvalues and eigenvectors became more accurate with increasing numbers of shapes per group. Broadly, consistent and reliable results were found numbers of shapes per group of order $N \approx 50$ to 100 for the "smile" dataset and for $K \geq 5$ groups, although these are very much *ad hoc* estimates only. The number of points was one parameter that was found to set the overall scale for the eigenvalues. Hence, some of the issues relating to the implementation of mPCA for shape models have been considered.

Sample-size estimation is well-known in statistics for simple methods of statistical inference (e.g., *t*-tests), although rather less research has been carried out for more sophisticated methods such as PCA. Such analyses were carried out mostly for factor analysis, see, e.g., Refs. [15,16] or in non-PCA versions of multilevel modelling [17,18]. Virtually no research has been carried out on the effects of sample sizes for PCA applied to models of shapes. However, it is well-known that the rank of a covariance used in ASMs or AAMs (including the "mean" shape) cannot be greater than the number of subjects or shapes used in the forming the model. However, this tells us nothing about how large our sample should be in order to obtain good results, which is the essence of sample-size estimation in statistics. Common sense would tell us that the inclusion of more data in forming a model (perhaps multiple images or shapes from the same subject) ought to lead broadly to more

17

accurate models as long (as long as we reflect any multilevel data structures within these models). However, this is a complex issue, especially for data structures than contain multiple levels. Although Monte Carlo simulation is a potential method of exploring such issues [18], this subtle and important topic lies beyond the scope of this article and is therefore a topic of research that might be studied in future.

Results of model fits for the "Smile" and OSTEODENT datasets were also considered. It was seen for the "Smile" dataset that standard PCA (i.e., PCA that makes no distinction between within- and between-group effects in the data) and mPCA both performed adequately for those test cases that were not too extreme when compared to the training set. However, PCA performed badly for "extreme cases" in the "Smile" dataset that are nonetheless still possible in the "training set," whereas mPCA performed well even for such extreme cases. This result suggests that standard PCA misses some of the multilevel structure inherent in the data and so leads to poorer, as one would expect. As such, these problems can be avoided completely by using mPCA from the outset.

Random normally distributed placement error was added to the Smile (training) dataset and average point-to-point errors were found for both PCA and mPCA approaches. Both approaches showed that average point-to-point errors increases with magnitude of error added to the training set of shapes and that it decreases as we increase the number of eigenvectors/values retained. It has also been seen that mPCA provide sensible results when to compared to those results of standard PCA for the OSTEODENT dataset, e.g., the accuracy (i.e., measured by average point-to-point errors) of both methods again increases as we increase the number of eigenvectors/values retained in initial calculations for the OSTEODENT dataset. We remark that the mPCA formalism could be incorporated into existing ASM or AAM software straightforwardly in future studies.

mPCA is the correct method for combining sets of landmark points provided by different experts for the same images because it allows us firstly to model inter-rater effects in point-placement and then

18

secondly to control these effects in any subsequent model fit. Multilevel models might therefore also be useful in addressing the point-correspondence problem for ASMs and AAMs. Previous methods have used optimised procedures for the placement of points, such as using the minimum descriptor length approach [19,20] from a training set of example boundaries or surfaces. mPCA might be used, in principle, to combine results of automated procedures in a statistically well-controlled manner. This is another possible future application of mPCA.

The analysis of 3D facial movement is a subject of considerable interest in dentistry [21-24]. For example, it might be use to aid in craniofacial diagnosis and dental treatment planning. A sensible "next step" in this research relating to mPCA in shape or appearance models might be to use facial movement datasets (see, e.g., Refs. [25-27], for more information) in which multiple images from the same subject are taken. Another potential avenue of research is that of lip shape [28] and also any associated dynamics [29].

## Conclusions

It has been shown that mPCA in ASMs can be used in dentistry and that this approach gives clinically applicable results. Explicit calculations were given, which can be a useful guide to other researchers who wish to employ this approach. We have explored some of the practical issues relating to using such methods in these initial calculations, e.g., how changing the number of mark-up points affects results. This initial investigation for the Monte Carlo "Smile" dataset suggests that a better model should also lead to better model fits, especially for "extreme" cases with respect to the training set. An important point is that mPCA allows one, in principle, to have much more flexibility and control when forming models and then fitting these models in subsequent images searches. There appear to be many future areas for research relating to mPCA in ASMs and AAMs, including applications in medicine and dentistry.

19

## Acknowledgements

20

**Appendix A – Mathematical Formalism**

For ASMs, one carries out PCA for a covariance matrix as discussed in Ref. [8]. The $l^{\text{th}}$ eigenvalue is denoted $\lambda_l$ and its eigenvector is denoted by $u_l$. Note that we retain $M$ of the PCA eigenvectors in the expansion, where $M \leq 2n$. It is often useful to determine the proportion of cumulative variability by

$$\Lambda_M = \frac{\sum_{l=1}^{M} \lambda_l}{\sum_{l=1}^{2n} \lambda_l} \tag{A1}$$

If the eigenvectors $u_l$ are orthonormal then we can determine the coefficients, $a$, for a fit of the model to a new shape vector, $z$, readily by using

$$a_l = u_l \cdot (z - \bar{z}) \tag{A2}$$

where the symbol $\cdot$ indicates the scalar product between two vectors. Constraints are often placed on the $a$-coefficients, such as $|a_l| \leq 3\sqrt{\lambda_l}$, in active shape models (ASMs) and so the model fit is always "plausible" given the training set for the shape model. This property is an advantage of the method.

For multilevel PCA (mPCA) we represent the "subject" by the index $i$ and the "group" by the index $j$. The landmark points (i.e., mark-up points) may be represented by a vector, $z_{ij}$. The total number of such points is $n$. The index $i$ refers to a particular shape (of $N$ such shapes in each group) and $j$ represents the group number (of $K$ such groups). Let $z_{ijk}$ now refer to the $k^{\text{th}}$ element of this vector of size $2n$. Any shape vector can now be expressed (again trivially) as

$$z_{ij} = \bar{z} + (z_{ij} - \bar{z}_j) + (\bar{z}_j - \bar{z}) \tag{A3}$$

21

The grand mean shape for the $k^{\text{th}}$ element in the vector is evaluated over all of the shapes in our data set (all images and all groups), and is given by

$$\bar{z}_k = \frac{1}{NK} \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ijk} \tag{A4}$$

The mean shape for the $k^{\text{th}}$ element of the vector for a specific group $j$ is given by

$$\bar{z}_{jk} = \frac{1}{N} \sum_{i=1}^{N} z_{ijk} \tag{A5}$$

The *within-group* covariance matrix is given by

$$C_{k_1 k_2}^{w} = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{N-1} \sum_{i=1}^{N} (z_{ijk_1} - \bar{z}_{jk_1})(z_{ijk_2} - \bar{z}_{jk_2}) \tag{A6}$$

where $k_1$ and $k_2$ indicate elements of the covariance matrix. We carry out PCA for the (positive semi-definite) covariance matrix of the above equation and the eigenvalues are positive, real numbers. The $l^{\text{th}}$ eigenvalue is denoted $\lambda_l^w$ and its eigenvector is denoted by $u_l^w$. We now expand the vector $(z_{ij} - \bar{z}_i)$ in terms of the eigenvectors $u_l^w$, where

$$z_{ij} - \bar{z}_i = \sum_l a_l^w u_l^w \tag{A7}$$

where the $a_l^w$'s are just scalar coefficients. Constraints may be placed on these *a*-coefficients, such as $|a_l^w| \leq 3\sqrt{\lambda_l^w}$. The *between-group* covariance matrix is given by

$$C_{k_1 k_2}^{b} = \frac{1}{K-1} \sum_{j=1}^{K} (\bar{z}_{jk_1} - \bar{z}_{k_1})(\bar{z}_{jk_2} - \bar{z}_{k_2}) \tag{A8}$$

where $k_1$ and $k_2$ indicate elements of the covariance matrix. We carry out PCA for the (positive semi-definite) covariance matrix given above and the eigenvalues are positive, real numbers. The $l^{\text{th}}$ eigenvalue is denoted $\lambda_l^b$ and its eigenvector is denoted by $u_l^b$. We now expand the vector $(\bar{z}_j - \bar{z})$ in terms of the eigenvectors $u_l^b$, where

$$\bar{z}_j - \bar{z} = \sum_l a_l^b u_l^b \tag{A9}$$

22

where the $a_l^b$'s are just scalar coefficients. Constraints may again be placed on these $a$-coefficients, such as $\left| a_l^b \right| \leq 3\sqrt{\lambda_l^b}$. Any new shape vector may be represented by

$$z = \bar{z} + \sum_{l_1=1}^{M_w} a_{l_1}^w u_{l_1}^w + \sum_{l_2=1}^{M_b} a_{l_2}^b u_{l_2}^b + \varepsilon \tag{A10}$$

Thus, we have modelled the two levels in the data independently. Note that $\varepsilon$ that is an appropriate residual error term. Here we set the values of $M_b$ and $M_w$ by ranking all of the eigenvalues $\lambda^b$ and $\lambda^w$ into descending order and we choose the $M$ largest eigenvalues from both the within-group and between-group covariance matrices. Thus, $M_b$ and $M_w$ do not need to be equal necessarily. Similar to Eq. (A1), the proportion of cumulative variability with to respect to the eigenvalues of within covariance matrix and between covariance matrix separately, or for all eigenvalues (both within and between), which is denoted "total mPCA." An example of PCA and mPCA for the simplest type multilevel problem (i.e., the positioning of a single point in ten images by five observers) is given in Appendix B. The 2×2 covariance matrices, eigenvalues and cumulative proportion of variable are described in detail. The two-level model presented here ought to be generalizable to three or even more levels.

The covariance matrices are symmetrical and so all "within" eigenvectors $u_l^w$'s are orthogonal to all other "within" eigenvectors (and similarly for the "between" eigenvectors). However, the eigenvectors $u_l^w$ and $u_l^b$ do not necessarily have to orthogonal with respect to each other, and so an equivalent projection to Eq. (A2) for mPCA becomes problematic. A gradient descent method such as that described in Ref. [30] may be implemented straightforwardly to fit a model to a new set of candidate points. As is normal in (active) shape models, we define the usual distance measure between the new shape $z$ and the model fit $z^{\text{model}}$,

$$H = \sum_{k=1}^{2n} (z_k - z_k^{\text{model}})^2 = \sum_{k=1}^{2n} (z_k - \bar{z}_k - \sum_{l=1}^{M_w} a_l^w u_{lk}^w - \sum_{l=1}^{M_b} a_l^b u_{lk}^b)^2 \tag{A11}$$

(Note again that $z_k$ is the $k^{\text{th}}$ element of the shape vector $z$, etc.) A solution [30] is obtained by iterating the equations

$$a_l^w \leftarrow a_l^w - \kappa \frac{\partial H}{\partial a_l^w} \quad \text{and} \quad a_l^b \leftarrow a_l^b - \kappa \frac{\partial H}{\partial a_l^b} \tag{A12}$$

23

from some "starting point" (and with an appropriate choice of κ such as κ = 0.01) until convergence. We note that

$$\frac{\partial H}{\partial a_l^b} = -2\sum_{k=1}^{2n} u_{lk}^b (z_k - \bar{z}_k - \sum_{l'=1}^{M_w} a_{l'}^w u_{l'k}^w - \sum_{l'=1}^{M_b} a_{l'}^b u_{l'k}^b) \tag{A13}$$

and

$$\frac{\partial H}{\partial a_l^w} = -2\sum_{k=1}^{2n} u_{lk}^w (z_k - \bar{z}_k - \sum_{l'=1}^{M_w} a_{l'}^w u_{l'k}^w - \sum_{l'=1}^{M_b} a_{l'}^b u_{l'k}^b) \tag{A14}$$

If the "starting point" is the mean shape then we set $a_l^b = 0$ and $a_l^w = 0$ as our initial choices for these coefficients. This approach converged readily and rapidly (i.e., 100 to 1000 iterations for κ = 0.01) for all of the cases considered here. The constraints on the coefficients $\left|a_l^w\right| \leq 3\sqrt{\lambda_l^w}$ and $\left|a_l^b\right| \leq 3\sqrt{\lambda_l^b}$ can be computationally implemented easily in any such iterative scheme, which is a strong advantage of this approach. In this case, we constrain the between and within components individually. Note again that all studies were carried out here by using MATLAB.

24

## Appendix B – Illustration of finding covariance matrices for mPCA

The simplest possible (idealised) example that mPCA can be applied to is given by the positioning a single Cartesian point in ten images. This process is replicated five times (e.g., by five different "experts"), thus giving us five sets of points across each of the ten images:

*Set 1:* (-0.58, 0.16), (-0.63, -0.6), (0.43, -0.55), (-2.27, -0.65), (-2.03, -0.71), (-2.83, -0.56), (-2.26, -0.74), (-1.74, -0.37), (-2.03, 0.69), (-1.7, -0.4)

*Set 2:* (-0.31, 1.3), (-0.64, 1.51), (-0.19, 1.92), (-0.26, 1.37), (0.2, 1.44), (-0.05, 1.26), (-1.84, 0.98), (-0.6, 1.15), (-0.73, 1.34), (-0.34, 1.28)

*Set 3:* (1.86, -0.78), (0.28, -1.77), (1.48, 0.12), (2.25, -1.6), (0.05, -0.37), (1.19, -0.28), (0.36, -1.63), (3.43, -0.76), (2.98, 0.9), (3.31, -0.65)

*Set 4:* (-0.49, 0.76), (-1.75, -0.65), (-0.91, 0.6), (-1.22, 1.06), (1.01, 0.89), (-0.92, 1.43), (1.22, 1.79), (0.22, -0.69), (-1.33, 1.51), (-0.66, 0.37)

*Set 5:* (1.95, 0.1), (-0.66, -0.97), (1.64, 0.1), (0.9, -0.94), (-0.44, -0.25), (-0.43, -1.56), (2.39, -0.82), (2.35, 0.58), (0.38, -0.74), (-1.28, -0.26)

The mean values of each set are given by: *Set 1* (-1.56, -0.37), *Set 2* (-0.48, 1.36), *Set 3* (1.72, -0.68), *Set 4* (-0.48, 0.71), *Set 5* (0.68, -0.48). Thus, we see that there are offsets for each of the five "experts." We now apply PCA for a case in which no distinction between the five datasets (e.g., experts) is made. The covariance matrix for all 50 points is given by

$$C = \begin{pmatrix} 2.293 & -0.184 \\ -0.184 & 1.000 \end{pmatrix}$$

The eigenvalues for this matrix are given by $\lambda_1 = 2.319$ and $\lambda_2 = 0.974$, and unit eigenvectors are found readily. We apply mPCA and the "between" covariance matrix is given by

25

$$C^b = \begin{pmatrix} 1.264 & -0.377 \\ -0.377 & 0.622 \end{pmatrix}$$

The eigenvalues for this matrix are given by $\lambda_1^b = 1.438$ and $\lambda_2^b = 0.448$, and unit eigenvectors are found readily. The "within" covariance matrix is given by

$$C^w = \begin{pmatrix} 1.029 & 0.193 \\ 0.193 & 0.377 \end{pmatrix}$$

The eigenvalues for this matrix are given by $\lambda_1^w = 1.082$ and $\lambda_2^w = 0.325$, and unit eigenvectors are found readily. We now write all mPCA eigenvalues in descending order: $\lambda_1^b = 1.438$, $\lambda_1^w = 1.082$, $\lambda_2^b = 0.448$, and $\lambda_2^w = 0.325$. Thus if we were to restrict mPCA to 2 eigenvectors, e.g., in model fitting, then we would choose the first "between" eigenvector and the first "within" eigenvector.

26

# References

1. T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam, Use of Active Shape Models for Locating Structure in Medical Images, Image and Vision Computing 12 (1994) 355-365.

2. 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 (1995) 38-59.

3. A. Hill, T.F. Cootes, and C.J. Taylor, Active shape models and the shape approximation problem, Image and Vision Computing, 14 (1996) 601-607.

4. C.J. Taylor, T.F. Cootes, A. Lanitis, G. Edwards, P. Smyth, and A.C.W. Kotcheff, Model-based interpretation of complex and variable images, Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 352 (1997) 1267-1274.

5. T.F. Cootes and C.J. Taylor, A mixture model for representing shape variation, Image and Vision Computing 17 (1999) 567-573.

6. T.F. Cootes, G.J. Edwards, and C.J. Taylor, Active appearance models, Ieee Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 681-685.

7. T.F. Cootes and C.J. Taylor, Anatomical statistical models and their role in feature extraction, British Journal of Radiology 77 (2004) S133-S139.

8. P.D. Allen, J. Graham, D.J.J. Farnell, E.J. Harrison, R. Jacobs, K. Nicopolou-Karayianni, C. Lindh, P.F. van der Stelt, K. Horner, and H. Devlin, Detecting reduced bone mineral density from dental radiographs using statistical shape models, IEEE Trans.Inf.Technol.Biomed. 11 (2007) 601-610.

27

9. F. Lecron, J. Boisvert, M. Benjelloun, H. Labelle, S. Mahmoudi, Multilevel statistical shape models: A new framework for modeling hierarchical structures, 9th IEEE International Symposium on Biomedical Imaging (ISBI) (2012) 1284-1287.

10. K. Horner, H. Devlin and L. Harvey, Detecting patients with low skeletal bone mass. Journal of Dentistry 30 (2002) 171-175.

11. D.J.J. Farnell, A. Pickles, and C. Roberts, Measurement Error in Statistical Models of Shape. Computer Methods and Programs in Biomedicine 104 (2011) e29-e44.

12. G. Dunn, Design and Analysis of Reliability Studies Oxford University Press, New York, 1989.

13. R.J. Carroll, D. Ruppert, and L.A. Stefanski, L. A. 1995, Measurement Error in Nonlinear Models, CRC Press, 1995.

14. G. Dunn, Statistical Evaluation of Measurement Errors, 2nd edn, Oxford University Press, New York, 2004.

15. R.C. MacCallum, K.F. Widaman, S. Zhang, and S. Hong, Sample size in factor analysis, Psychological methods 4 (1999) 84.

16. H.W. Marsh, J.R. Balla, and R.P. McDonald, Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size, Psychological bulletin 103 (1998) 391.

17. T.A.B. Snijders, Power and sample size in multilevel linear models. Encyclopedia of statistics in behavioral science, John Wiley & Sons Ltd, 2005.

18. C.J.M. Maas and J.J. Hox, Sufficient sample sizes for multilevel modeling. Methodology 1.3 (2005) 86-92.

19. R.H. Davies, C.J. Twining, T.F. Cootes, C.J. Waterton, and C.J. Taylor, A minimum description length approach to statistical shape modelling, IEEE Transactions on Medical Imaging 21 (2002) 525–537.

20. H.H. Thodberg, Minimum description length shape and appearance models, Information Processing in Medical Imaging: Lecture Notes in Computer Science 2732 (2003) 51–62.

28

21. H. Popat, S. Richmond, L. Benedikt, D. Marshall, and P.L. Rosin, Quantitative analysis of facial movement—A review of three-dimensional imaging techniques. Computerized Medical Imaging and Graphics 33 (2009) 377-383.

22. H. Popat, S. Richmond, R. Playle, D. Marshall, P.L. Rosin, and D. Cosker, Three-dimensional motion analysis–an exploratory study. Part 1: Assessment of facial movement. Orthodontics & craniofacial research 11 (2008) 216-223.

23. H. Popat, S. Richmond, R. Playle, D. Marshall, P.L. Rosin, and D. Cosker, 2008. Three-dimensional motion analysis–an exploratory study. Part 2: Reproducibility of facial movement. Orthodontics & craniofacial research 11 (2008) 224-228.

24. H. Popat, E. Henley, S. Richmond, L. Benedikt, D. Marshall, and P.L. Rosin, A comparison of the reproducibility of verbal and nonverbal facial gestures using three-dimensional motion analysis. Otolaryngology--Head and Neck Surgery 142 (2010) 867-872.

25. L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, A high-resolution 3D dynamic facial expression database. In Automatic Face & Gesture Recognition 08 (2008) 1-6.

26. L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato, A 3D facial expression database for facial behavior research. In Automatic face and gesture recognition 06 (2006) 211-216.

27. A list of 3D facial databases is given at: http://www.face-rec.org/databases/

28. L. M. de Freitas, K. M. de Freitas, A. Pinzan, G. Janson, M.R. de Freitas. A comparison of skeletal, dentoalveolar and soft tissue characteristics in white and black Brazilian subjects. J Appl Oral Sci. 18 (2010) 135-42.

29. P. Chetan, P. Tandon, G. K. Singh, A. Nagar, V. Prasad, V.K. Chugh. Dynamics of a smile in different age groups. Angle Orthod. 83 (2013) 90-6.

30. J. Hertz, A. Krogh, and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley Publishing Company, 1991, pp. 82-87.

29

## Highlights

- mPCA in ASMs can be used in medical and dental image analysis
- Results provided by mPCAin initial studies appear to be sensible
- Between- and within-subject variations are modelled correctly using mPCA
- mPCA has more flexibility, control, and accuracy than standard PCA
- mPCA is the method of combining sets of landmark points from different experts