UNIVERSITY OF GLASGOW

Methods for the Analysis of Three-Dimensional Anatomical Surfaces

by

Joanna Smith

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

School of Mathematics and Statistics

College of Science and Engineering

July 2012

Declaration of Authorship

I, JOANNA SMITH, declare that this thesis titled, 'Methods for the Analysis of Three-Dimensional Anatomical Surfaces' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Abstract

Shape is an inherent feature in everything around us and one that we, as humans, can process very efficiently. The question of how to analyse shape in an objective manner, however, is a more complex one. With continually improving and more readily available imaging technologies, there are an increasing number of fields in which it is of interest to have a more quantitative means of analysing the resulting images. Shape analysis is a rapidly growing branch of statistics that aims to meet these needs.

In contrast to the early methods of shape analysis that were based on distances or angles from an object, shape is now generally assessed in terms of the full geometry of the object. This does not necessarily mean an analysis of the object in its entirety however; a simplified representation of the surface is more often used. This representation has traditionally been in the form of a set of landmarks - points of anatomical or mathematical interest on an object that act as key descriptors of its shape. A key feature of these points is therefore that they are in positions that correspond across all images. Clearly this makes them a very powerful tool for analysis, particularly useful for comparison across shapes, and as such many methods have been presented for their analysis.

However, due to the fact that they are based on significant anatomical points and are more often than not manually placed on an image, landmark points tend to be fairly small in number. A major disadvantage to this type of approach is therefore that they give a very sparse description of the object of interest's shape. In order to improve on this, alternative methods have been developed that instead present the object in terms of a set of curves or, more recently, representative surface points. These representations give a richer description of shape but, provided they are created so that they also correspond across objects, can equally be analysed by means of the many existing landmark-based techniques. Nevertheless, although a surface-based approach clearly utilises a greater deal of the shape information of an image and hence provides a more comprehensive analysis, it remains a far less common technique for the analysis of shape.

This thesis therefore aims to develop tools for the analysis of three-dimensional surface data, with focus lying specifically in the field of medical imaging. Three distinct studies are conducted, each of which has its own questions of interest and hence necessary techniques. The first study presented is based on a cohort of unilateral mastectomy and reconstruction patients, where interest lies in evaluating the breast asymmetry that is present post-surgery. A novel method is presented for the creation of corresponding surface points, and an analysis of these is conducted based on an established approach to the study of asymmetry.

The second study then looks to investigate the 'normal' patterns of facial growth that are seen in young children, specifically between the ages of 3 months and 5 years. From the multiple, longitudinal images that are available for each child, a set of corresponding surface representations are created by means of a well-established technique known as the Thin Plate Spline. A principal components analysis is then applied to the surfaces in order to reduce the dimensionality of the data, and the resulting principal components scores are modelled by a linear mixed effects approach.

The third and final study presented is an investigation into the soft-tissue changes that are seen as a result of craniofacial surgery. A system is devised to index the location of all points on the surface, and this information is then used to model the changes taking place at various positions on the face. While most previous approaches to this problem have been based on complex finite element models, this study aims to investigate whether a simpler and more efficient statistical modelling approach can instead prove useful.

The diversity of these studies hints at the wide-ranging applications of shape analysis within the medical imaging field alone, as well as many of the issues that can arise in the analysis of surface data. While it is intuitive that more informed conclusions can be drawn through these surface based analyses than would be possible by a more traditional landmark-based approach, it is also seen that the use of a surface representation allows for an improved visualisation and interpretation of the results.

The techniques developed here are illustrated on specific medical applications, although it is hoped that they would prove similarly useful in a wider variety of shape analysis settings.

Sections of this thesis have been presented at a variety of conferences, and can be found in published form in Smith and Bowman (2010), Smith and Bowman (2011), Henseler et al. (2011) and Henseler et al. (2012).

Acknowledgements

First and foremost, immeasurable thanks go to my supervisor Professor Adrian Bowman, whose knowledge and support over the last three years have been invaluable. Thank you for always being a constant and calming presence, without which this thesis would not have been possible.

My deep gratitude also goes to the Engineering and Physical Sciences Research Council and the (former) Department of Statistics for jointly funding this research project, as well as to the Glasgow Dental Hospital for provision of the data.

To the remainder of the staff and students of the department, my 8 years spent here have been a pleasure and I thank everyone for their help and warmth over that time. Particular thanks go to Maria and Massi - for adventures, laughter and piadina. Ci vediamo nella friccateria, amici!

To my amazing friends, I thank you all for the years of abuse and the many sanity-reviving retreats to the north. In particular to Erin and Jackie, for being like sisters for long over a decade, and to the 'dunkers', for keeping me in tea and terrible movies.

To my wonderful family, you are a constant source of love and support and I thank you for always being there. Special thanks go to Rebecca and Niamh, who have kept me fed, watered and generally sane in the final fraught few months. I will miss our little home deeply, and in particular the laughter and ridiculous dancing that fills it.

And finally, to Mum. You were, and continue to be, my rock, my inspiration and the type of person that I aspire to be. Thank you for that and so much more. This is for you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Shape is an inherent feature in everything around us and one that we, as humans, can process very efficiently. A lot of information can be held in the shape of an object, and as such there are many fields in which the ability to analyse shape is of great interest, with applications being found in medicine, biology, computer vision and genetics to name but a few. The human brain has an impressive capacity for visual perception, allowing us to understand and describe shape in a variety of ways. To analyse shape objectively, however, a scientific and quantitative method of assessment is clearly required. The field of shape analysis, or alternatively morphometrics, is a relatively new and growing branch of statistics that attempts to meet these needs.

Early approaches to the study of form saw the analysis of selected distances or angles from the object in question, and a brief discussion of studies by such methods can be found in Dryden and Mardia (1998). However, this approach reduces the shape information to solely a series of positive values or measurements, meaning that the topology of the object itself is lost. As such, favour now lies strongly with analyses that are based on the full geometry of the object. This does not necessarily mean an assessment of the object in its entirety however; a simplified representation that adequately describes its shape might be used.

The theoretical foundations of this geometrical approach can be attributed to Kendall (1984), although its first application in a shape analysis setting was presented by Bookstein (1986). The connection to the underlying mathematical theory was however unclear at the time, and was only explicitly outlined as the basis for the statistical methodology around 10 years later (Bookstein 1996*b*). Since that time, shape analysis has been a very active field of research.

As discussed in Dryden and Mardia (1998), the statistical approach to shape analysis takes into account the 'presence of randomness', thereby considering the objects under study as a sample from the wider population. This allows for the estimation of mean shapes, as well as for inference on the variability of shapes that is seen across the population. Shape can intuitively be thought of as the geometrical information of an object that is unchanged by a relocation, rotation, enlargement or reduction of the object. When comparing across different objects it is therefore necessary to remove these effects. This is an issue which is very important within shape analysis, and some of the standard methods for the removal of these effects are discussed later in this thesis.

As stated by Loncaric (1998),

> 'The goal of shape analysis is to provide a simplified representation of the original shape so that the important characteristics of the shape are preserved.'

To date, the most common approach to achieving this simplified representation has been to define an object by a set of landmarks, i.e. a set of points that act as key descriptors of shape. These are usually, although not always, anatomically significant points. An example of some landmarks of the human face can be seen in the left hand panel of Figure 1.1. The vast majority of work in the shape analysis field has been based on the analysis of these landmarks, with the methods taking advantage of the corresponding nature of their locations across the population. Applications have been far reaching and have included investigations into the physical effects of many disorders, such as Foetal Alcohol Syndrome and epilepsy (Mutsvangwa

and Douglas 2007, Sonat et al. 2009), as well as for classification purposes, for example in the assessment of sexual dimorphism (Takemura et al. 2004). Other interesting studies based on this type of data include that of Kent et al. (2001), who looked at the development of growth models for landmark data, using roughness penalties to investigate the smoothness of both the rate and direction of growth. Mardia et al. (2000) used landmark data as a means of analysing the asymmetry of shapes, which was assessed by the comparison of a set of landmarks and their reflections. Although these examples are by no means a comprehensive description of the landmark based methods in the literature, they provide an idea of the diverse applications and techniques that exist.

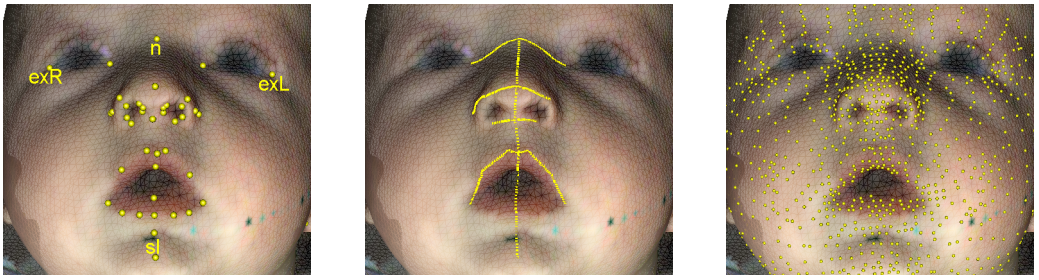

FIGURE 1.1: An image of a child's face showing: *Left:* a set of landmark points *Centre:* a set of surface curves *Right:* a set of representative surface points. A selection of the landmarks that are used in later analysis have been labelled.

A disadvantage of this landmark based approach, however, is that only a relatively small number of points are taken into account and hence much of the information of the shape is lost. An object with a very simple structure may indeed be adequately described by a finite and relatively small set of points such as this. For more complex and varied shapes, however, as is often the case with medical and biological images, a more thorough description of the shape may be required. As such, curves and, more recently, surface representations have instead been utilised as a means of analysis. The attractive quality of landmarks, however, and the underlying basis for their analysis, lies in the standardised locations that they depict on the surface. It is therefore necessary to ensure that any curves or surface representations have this same feature if they are to allow for as powerful an analysis of shape, or indeed for the use of standard landmark techniques with this richer type of data.

With shape data, curves are generally formed by a set of pseudo-landmarks, meaning that they are created by the interpolation of two existing landmark points. Provided a regularised interpolation is used across all images, this will clearly serve to achieve the necessary correspondence of the points, making this a fairly straightforward extension to implement. Some examples of this type of surface curve can be seen in the centre panel of Figure 1.1. Studies that have been based on this type of data include Barry and Bowman (2008), who described a series of facial curves by a B-spline parameterisation and subsequently modelled the spline coefficients by a linear mixed effects approach. The method was used to compare facial shapes in a group of children with a cleft palate with those of a group of age-matched controls. Another example of the analysis of surface curves can be seen in Samir et al. (2006), where geodesic paths were calculated between repeatedly captured facial curves with the aim of developing an automated facial recognition system.

While curves can provide a more extensive description of a shape than landmark points alone, there often remain many areas of the surface that are not included in this type of representation (as can be seen in Figure 1.1). In situations where any subtle or more localised changes are of interest, it is therefore preferable to have a representation of the object at a more global level. As such, many approaches have been suggested to extend the existing methods in order that a more comprehensive set of surface data can be analysed. This extension has generally been in the area of creating correspondence between different surfaces, and several of the proposed methods are discussed in later chapters.

To date, a number of studies have been presented with a surface-based approach. For example Hennessy et al. (2007) and Brignell et al. (2010) used surface data in order to assess the effects of schizophrenia on facial and cerebral shape respectively, while Hutton et al. (2001) examined the patterns of variation seen in adult facial shapes. The latter study reported an increased sensitivity to factors such as age and gender when compared to a landmark based model, as would intuitively be expected from an analysis that is using more of the available shape information. However, despite recent movements towards this approach, surface based methods still remain a far less common technique for the analysis of shape.

The aim of this thesis is therefore to develop tools for the analysis of three-dimensional surface data, with a specific focus in the field of medical imaging. This is an area where images are increasingly being captured as a means of assessment, and where the technologies for doing so are continually improving. It is therefore of great interest to be able to assess these images in an effective fashion. It should be emphasised that, despite this specific focus, the methods explored here could equally be applied outwith this field.

Three different practical applications are considered throughout this thesis, leading to a variety of issues and necessary techniques. Alongside the more general objective, each of these three studies is distinct and has its own aims and agenda. These aims are discussed more fully in the following chapter, but can be summarised as follows:

- The development of methods for the assessment of breast asymmetry, in this instance applied to a cohort of mastectomy and reconstruction patients.
- The analysis of facial shape, investigated in terms of a model for 'normal' facial growth patterns, as well as in an evaluation of the outcomes in cleft palate repair surgery.
- The development of a model to determine the soft tissue changes seen as a result of craniofacial surgery, examined specifically for the case of orthognathic surgery.

The structure of the thesis is as follows. Chapter 2 introduces in greater detail the three datasets that form the basis of the analyses, along with a discussion of the data validation procedures that were necessary. A review of some standard methods in shape analysis is then given in Chapter 3, where both the foundations of the landmark-based approach and the extension required for their application to surface data are detailed. Chapter 4 then deals with a variety of issues that can arise in the analysis of surface data, specifically detailing ways in which the surface can be reduced to the area of biological interest alone, thereby leading to a more focused analysis.

The following chapters then aim to answer the questions posed by each of the three studies. Chapter 5 demonstrates methods for the analysis of asymmetry, illustrated on both the cleft palate and the breast reconstruction datasets. The standard methods of Mardia et al. (2000) are utilised, with some application specific extensions introduced. The cleft dataset is then looked at once more in Chapter 6, where the model developed for the assessment of normal facial growth is detailed. The approach involves a principal components analysis of the data, the scores of which are modelled by means of a linear mixed effects model. Chapter 7 then considers the modelling of soft tissue reactions in facial surgery, where a novel approach to this type of problem is presented. Finally, Chapter 8 concludes with some discussion of these techniques, and of course suggestions for possible future work.

Chapter 2

Data Description and Validation

As mentioned in Chapter 1, the techniques developed throughout this thesis are illustrated on data from three distinct clinical studies. Widely speaking, all three studies are concerned with assessing the effectiveness of a particular type of reconstructive surgery. However, the motivations behind their collection and questions of interest are different, giving rise to a variety of issues and necessary techniques. The background to each of these studies is given here, along with a description of the validation procedures that were necessary for one of the datasets.

2.1 Breast Reconstruction Data

The first study explored looks to examine the extent of asymmetry present in the breasts of cancer patients who have undergone a unilateral mastectomy and reconstruction procedure. This procedure involves the removal and subsequent rebuilding of one breast, with the aim of making it as similar to the other breast as possible. The study involved 43 female patients, each of whom was captured by a stereophotogrammetry camera system. This is a type of system which takes multiple, simultaneously captured two-dimensional images and pieces them together to construct a three-dimensional representation of an object, a method first used in a clinical setting by Burke and Beard (1967). For this study, a four-pod, eight-camera system was used to capture the breasts simultaneously from the right, left, frontal and

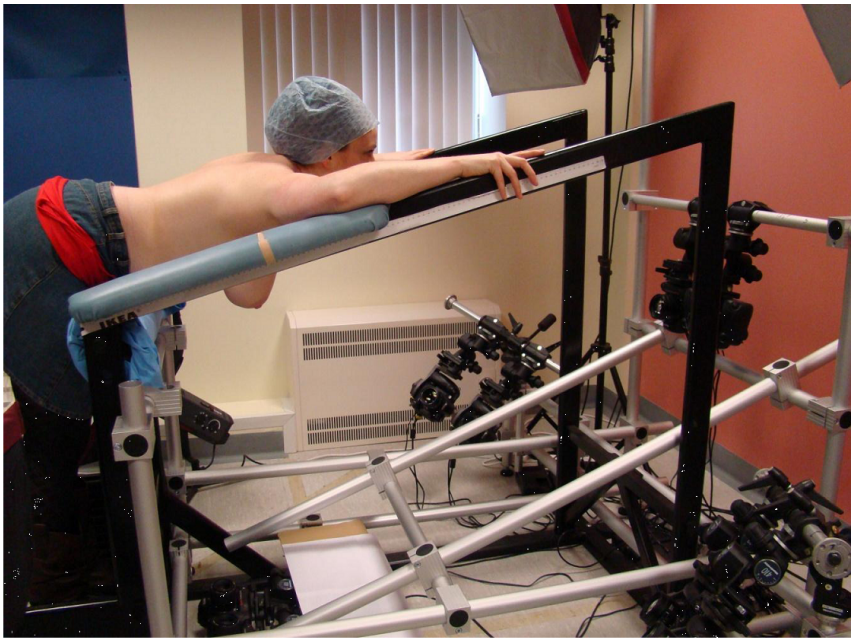

FIGURE 2.1: The stereophotogrammetry system utilised for image capture of the breasts.

inferior perspectives, in order to ensure the entire surface was adequately described (Figure 2.1). An overview of the data collection procedure is given in Henseler et al. (2011). From the resulting three-dimensional images the breast volume could be determined using the Breast Analysis Tool (BAT) software, which was developed at the University of Glasgow (Oehler 2006).

While stereophotogrammetry is a relatively well established technique, its application to capturing the breast is novel with only a small number of other studies reported, such as Loughry et al. (1987) and Losken et al. (2005). It was therefore necessary to assess the accuracy of the system in this setting prior to analysis, in order to ensure that any findings were valid. A discussion of the required validation procedures is now given.

2.1.1 Validation of the Camera System

Prior to the image capture of the patients, a pilot study was conducted in order to investigate the stereophotogrammetry system. The accuracy and reproducibility of

the system in recording breast volume were evaluated by comparison to measurements obtained from the water displacement method, which was taken to be the gold standard. The accuracy of the two methods was compared in two stages - firstly using measurements from specially constructed plaster models, and secondly using those from the breasts of volunteer live models. The data collection and volume measurements were conducted by Dr Helga Henseler at the Glasgow Dental Hospital, and further details of the two methods of measurement can be found in Henseler et al. (2011).

Systematic Differences for Plaster Models

For the first stage of the study, nine plaster casts were used for comparison. The casts had been made in order to represent the wide variety of shapes and sizes of the female breast, and their individual volumes were calculated ten times by each of the methods. The average values found by each method, as well as the variability across the 10 measurements, can be seen in Table 2.1 and Figure 2.2.

	Water Displacement		3-d Imaging	
Plaster Model	Mean	St. Dev.	Mean	St. Dev.
1	364.60	3.89	354.40	6.77
2	434.70	2.41	422.30	2.41
3	579.00	2.36	613.40	7.96
4	231.10	3.96	209.50	0.85
5	433.40	2.07	465.10	1.20
6	591.60	1.96	610.50	4.14
7	596.60	5.13	620.70	9.52
8	100.40	2.01	96.10	0.74
9	863.10	6.62	902.60	6.50

TABLE 2.1: Mean and standard deviation of measurements for the plaster models, by both water displacement and 3-d imaging.

The overall mean volume obtained from all casts was 466.06cc from water displacement and 477.18cc from the BAT giving a mean difference of 11.12cc. While this suggests that the differences in measurement between the two methods were not overly large, this will now be looked at in more detail.

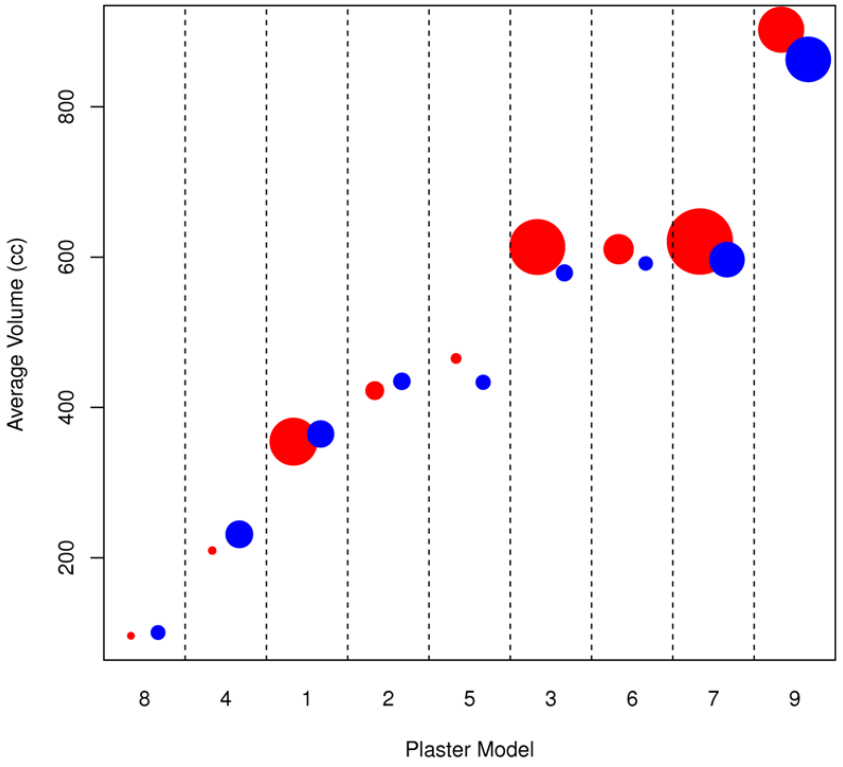

FIGURE 2.2: The mean volume found for each plaster model by both water displacement (blue) and 3-d imaging (red), plotted in order of ascending overall mean. The size of each circle is proportional to the variability (i.e. standard deviation) for that model across the ten repeat measurements.

From Figure 2.2 it appears that, at least in the case of 3-d imaging, there may be a tendency for a higher level of variation as breast size increases. This trend was confirmed by a correlation test (p=0.033) with a confidence interval for the correlation coefficient of (0.08, 0.93). This trend was unique to the 3-d imaging method however, as no significant correlation was found between breast volume and reproducibility for the water displacement method (p=0.168).

The average volumes for each cast are shown in the left panel of Figure 2.3, along with the line of equality. As can be seen, the points all lie close to the line, suggesting that the two methods give fairly similar results. It also seems that the smaller plaster casts are estimated to have a greater volume by water displacement than 3-d imaging, whereas as the size of the cast increases the opposite becomes the case. This can be examined by fitting a regression model and seeing if the slope of the line of best fit (i.e. the coefficient for water displacement) is greater than 1. Upon fitting the

model, the 95% confidence interval for this coefficient was found to be (1.02, 1.14), thereby confirming this trend.

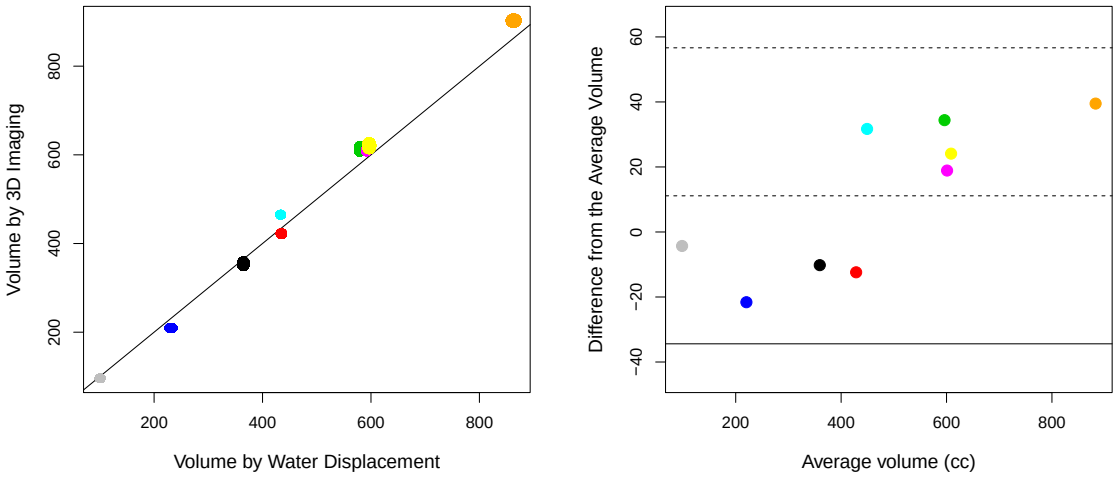

FIGURE 2.3: *Left:* Scatterplot of average volumes found by water displacement and 3-d imaging for the plaster models, plotted alongside the line of equality. The ellipses represent the variability seen by each method. *Right:* Bland Altman plot for the plaster models.

As can be seen in Table 2.2 below, the differences between the two methods did not exceed 40cc for any cast. The relative differences shown here were calculated in each case as the average volume found by 3-d imaging subtracted from that found by water displacement, divided by the overall mean for that case. The values for water displacement were between 6% smaller and 10% larger, with the volume being on average 0.5% smaller by water displacement.

Plaster Model	Difference between means	Relative difference (%)
1	10.2	2.8
2	12.4	2.9
3	-34.4	-5.8
4	21.6	9.8
5	-31.7	-7.1
6	-18.9	-3.1
7	-24.1	-4.0
8	4.3	4.4
9	-39.5	-4.5

TABLE 2.2: Relative differences between the two methods for the plaster models.

As the variance does not appear to be constant across all models (Figure 2.2), comparing the methods by a t-test may not be appropriate here. However, we can instead calculate the distribution for the average difference and investigate its expected value. Taking x_i and y_i to be the mean volumes obtained for plaster model i by water displacement and 3-d imaging respectively, x_i and y_i are then distributed as

$$\bar{x}_i \sim N(\mu_i, \hat{\sigma}_i^2), \qquad \bar{y}_i \sim N(\nu_i, \hat{\tau}_i^2) \tag{2.1}$$

The distribution of the difference for the i-th model is then given by

$$\bar{x}_i - \bar{y}_i \sim N(\mu_i - \nu_i,\ \ \hat{\sigma}_i^2 + \hat{\tau}_i^2), \tag{2.2}$$

with the distribution of the mean difference over all models being

$$\bar{D} \sim N(\bar{\mu} - \bar{\nu},\ \ \hat{\bar{\sigma}}^2 + \hat{\bar{\tau}}^2), \tag{2.3}$$

where $\bar{D} = \frac{1}{n}\Sigma_{i=1}^{n}(\bar{x}_i - \bar{y}_i)$. A 95% confidence interval for $\bar{\mu} - \bar{\nu}$ can then be calculated by

$$\bar{D} \pm 1.96 * \sqrt{(\hat{\bar{\sigma}}^2 + \hat{\bar{\tau}}^2)} \tag{2.4}$$

This led to a confidence interval of (-15.44cc, -6.81cc) for the plaster model data, thereby indicating a significant, albeit small, difference between the two sets of measurements.

Finally, the agreement between the two methods was also assessed using a Bland-Altman plot. This is a technique for comparing two distinct methods of measurement on a certain parameter, as discussed in Bland and Altman (1986). The limits of agreement (calculated as $d \pm 2s$, where d is the mean difference and s is the standard deviation of the differences) are shown in the right panel of Figure 2.3. As all points

lie within the limits, this suggests that it is valid to replace the water displacement method by 3-d imaging, provided that differences within the limits of agreement would be clinically acceptable. In this instance, the limits were found to be $2s = 46$cc, a relatively small value in comparison to the overall volumes of the breasts. It is therefore likely that discrepancies of this magnitude would not be problematic, and hence the two methods appear to be interchangeable.

Systematic Differences for Live Models

In the second stage of assessment, the differences between the two methods were assessed for the breasts of live models. To do this, six volunteers had their breast volume measured six times each, and the results can be seen in Table 2.3. The overall mean volume obtained was 687.19cc from water displacement and 480.14cc from the BAT, giving a mean difference of 207.05cc - markedly higher than the difference seen for the plaster models. This greater separation between the two methods is also apparent in Figure 2.4.

	Water Displacement		3-d Imaging	
Live Model	Mean	Standard Deviation	Mean	Standard Deviation
1	1083.08	34.60	618.93	62.61
2	1128.42	32.37	854.23	42.09
3	390.33	21.52	206.57	12.10
4	295.00	12.81	176.28	22.64
5	220.42	24.02	89.17	28.08
6	1005.92	141.64	935.68	25.23

TABLE 2.3: Mean and standard deviation of measurements for the live models, by both water displacement and 3-d imaging.

The differences between the two methods exceeded 400cc when assessing the live models, with relative differences ranging from between 7% and 85% and an average relative difference of 47.7% (Table 2.4). Clearly this is far higher than the relative differences that were seen previously.

In contrast to the plaster models, no significant correlation was found between breast volume and the variability of the measurements for either method.

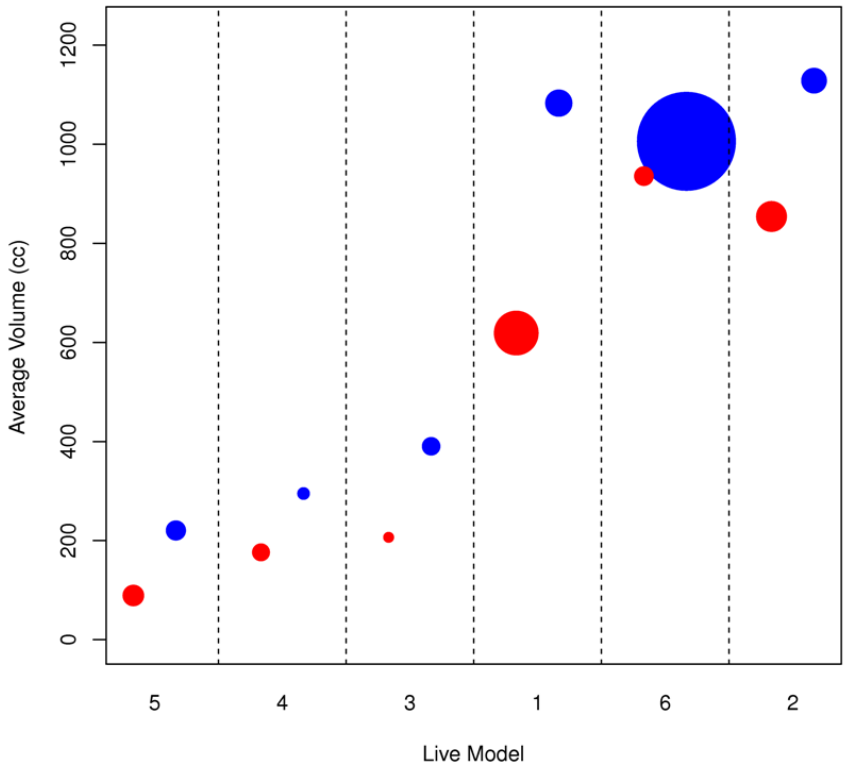

FIGURE 2.4: The mean volume found for each live model by both water displacement (blue) and 3-d imaging (red), plotted in order of ascending overall mean. The size of each circle is proportional to the variability (i.e. standard deviation) for that model across the six repeats.

Live Model	Difference between means	Relative difference (%)
1	464.15	55.0
2	274.18	28.0
3	183.77	62.0
4	118.72	50.0
5	131.25	85.0
6	70.23	7.0

TABLE 2.4: Relative differences between the two methods for the live models.

As can be seen in the left panel of Figure 2.5, the points do not lie as close to the line of equality for this data. Furthermore, the mean measurements for water displacement are larger than those for 3-d imaging in all cases. As before, the differences between the two methods were formally assessed by calculating the expected value of the mean difference. Here, $\bar{D} \sim N(207.05, 869.09)$, leading to a confidence interval for the mean difference of (149.27cc, 264.83cc). This indicates that the volumes found by water displacement tended to be higher than those found by 3-d imaging

on average, with the true difference likely to be somewhere within this range.

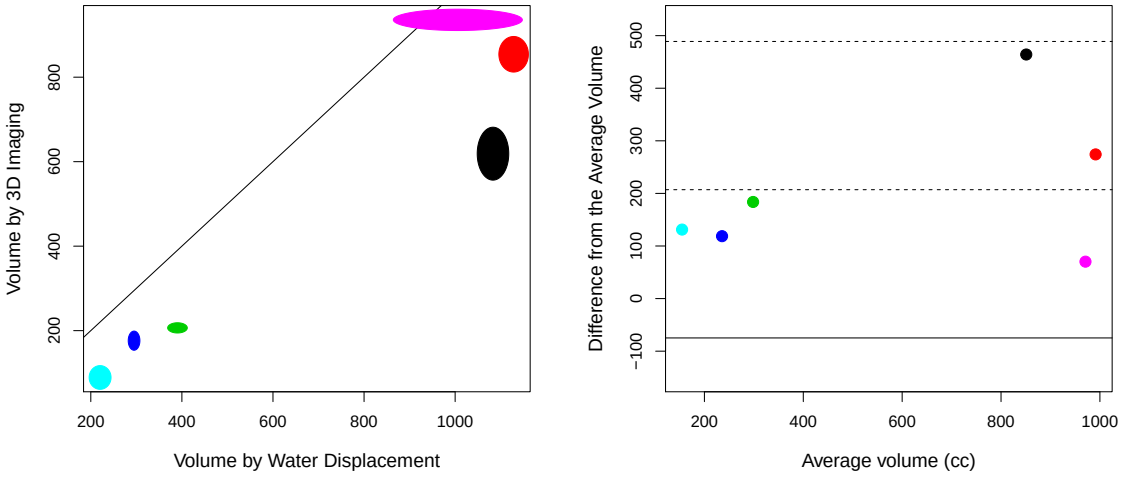

FIGURE 2.5: *Left:* Scatterplot of average volumes found by water displacement and 3-d imaging for the live models. The ellipses represent the variability seen by each method. *Right:* Bland Altman plot for the live models.

Using the Bland-Altman method to again assess the agreement, the differences were once more all found to lie within the limits of agreement, as can be seen in the right panel of Figure 2.5. However, in this case the bands are far wider and the margin of error is 281.88cc, which is substantial when compared to the sizes of the breasts.

However, while the previous tests have shown that there is a significant difference between the measurements found by each method for the live models, it is possible that this is a constant difference, i.e. that water displacement finds a volume Xcc higher than 3-d imaging due to some systematic difference, for example. From Figure 2.4, this at least seems plausible for the first five models. To investigate this, a regression model was fitted to the two variables, from which it was found that there is a strong relationship between the measurements arising from the two methods. The resulting 95% confidence interval for the coefficient of 3-d imaging was (0.93, 1.23). As the interval contains the value 1, this suggests it is plausible that the discrepancies between the two methods are constant. Upon setting the slope to be 1 and refitting the regression, the coefficient for the intercept shows what this constant difference is likely to be, and its 95% confidence interval was found to be (156.99, 257.11).

This difference between the two methods is perhaps to be expected with the live models, as it is very possible that the point to which the breast is immersed in the water is not the same as where the chest wall is calculated to be by the BAT software. The fact that the difference is reasonably constant, however, suggests that the volume is being adequately described by the software.

Reproducibility of 3-d Measurements for Live Models

In addition to comparing the measurements obtained from the two measurements, it is also of interest to examine how consistent the repeated measurements were within each method.

Overall, the volume measurements were found to be more reproducible by 3-d imaging than by water displacement in the live models, with standard deviations of 36cc and 62.6cc respectively. This was largely due to model 6 however who had a hugely increased standard deviation for water displacement, whereas almost all other models had a lower standard deviation and therefore more reproducible results by this method. This can be clearly seen in Figure 2.4.

For the live models, however, there is considerably more information available about the variability of the 3-d measurements. Each of the live models underwent six repeat captures by the 3-d stereophotogrammetry system. This was done at two separate timepoints, with three captures on each occasion. From each of the six images the operator used the BAT software to calculate the breast volume three times separately, and the average of these measurements was then used. It is therefore of interest to examine where most of the variability in the 3-d imaging results arose - was it from the pose taken in the rig on the two separate occasions, the individual image captures of the models or the operators use of the BAT software for the volume estimation?

In order to look at this, a linear mixed effects model was fitted to the data. This type of model assumes an overall mean, and finds how far the results deviate from this mean for each model, pose, capture and individual BAT measurement. The model for a particular measurement y_i can be written as

$$y_i = \mu + a_i + b_{ij} + c_{ijk} + d_{ijkl}, \tag{2.5}$$

where $i = 1, ..., 6$, $j = 1, 2$, $k = 1, ...3$ and $l = 1, , , 3$ represent the live model, pose, capture and BAT measurement respectively.

Upon fitting the model, by far the largest component of variability was of course due to the differences in the live models themselves, with a standard deviation of 369.73cc. While these genuine size differences are of course a source of variability in the data, they are not a source of error in the volume measurement and are therefore not of interest here. From the three possible sources of error, the largest contribution to variation was introduced by the models assuming the two different poses on the rig, with a standard deviation of 32.95cc. The individual BAT measurements and the differences between the repeat captures then gave slightly smaller contributions to the variation, with standard deviations of 28.32cc and 19.43cc respectively. The interval estimates for the variability of each of the components is shown in the plots below. The second plot is simply the first plot without the model component, in order to give a clearer picture of what is happening at a lower level.

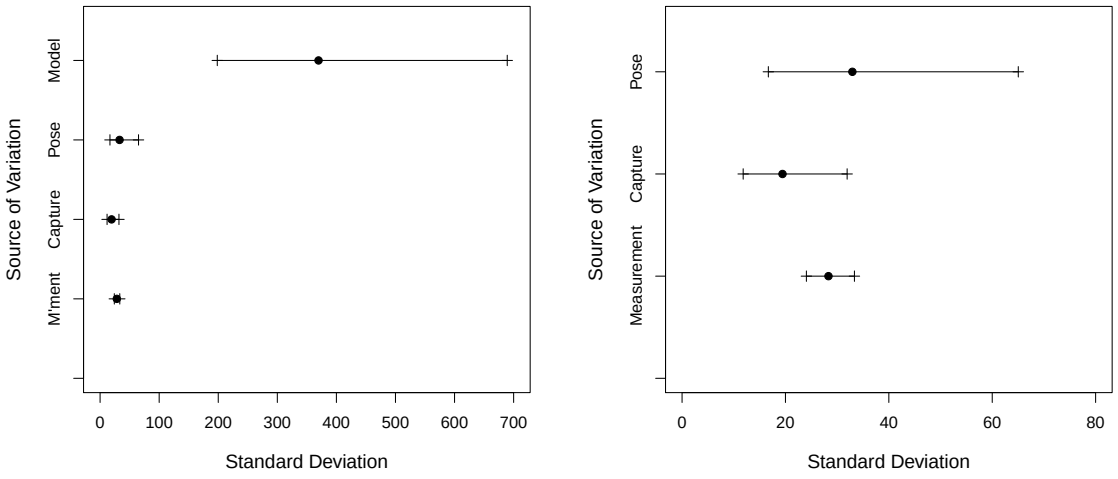

FIGURE 2.6: Interval estimates for the variability in volume measurements attributable to model, pose, capture and BAT measurement

2.1.2 Reproducibility of Landmarks

As well as assessing the accuracy of the system in terms of volume calculation, it was also necessary to investigate the reproducibility of the landmark points that were placed on these images and used in later analysis. Each of the 3-d surfaces was marked with 10 anatomically significant landmarks, which can be seen in Figure 2.7. The choice of surface landmarks was taken following established literature on land-

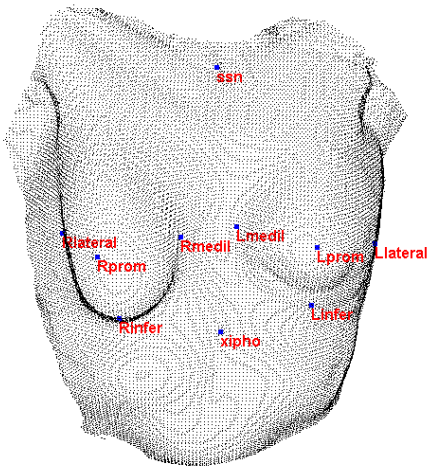

FIGURE 2.7: Image of a chest with the ten landmarks indicated: suprasternal notch (*ssn*), xiphoid (*xipho*), and on each breast (left and right breast respectively) the most medial point (*Lmedial*, *Rmedial*), the most lateral point (*Llateral*, *Rlateral*), the most prominent point (*Lprom*, *Rprom*) and the most inferior point (*Linfer*, *Rinfer*).

mark placement on the breast by Brown et al. (1999). The landmarks were placed manually by a clinician using Facial Analysis Tool software (FAT), an established 3-d manipulation software originally developed for landmark identification on 3-d images of the face by Hajeer et al. (2005). The software allows the user to view, rotate and magnify the images in 3-d in order to place the landmarks, and records the landmark coordinates for future analysis.

An investigation into the reproducibility of the landmarks was conducted, for which ten images were landmarked four times, twice on each of two separate days. As

discussed by Gwilliam et al. (2006) and Toma et al. (2009), to be of clinical use the reproducibility should be assessed in all three planes of space. For each landmark, the within-day, between-day and overall technical error measurement (TEM) was calculated for each of the three coordinate positions as a measure of reproducibility. TEM is calculated as the square root of measurement error variance when a number of repeat measurements are taken on the same subject, as discussed in Mueller and Martorell (1988) and Ulijaszek and Kerr (1999). For the comparison of two sets of measurements (i.e. the within-day and between-day TEMs), the necessary formula is

$$TEM = \sqrt{\frac{\Sigma_{i=1}^{N} d_i^2}{2N}} \tag{2.6}$$

where d_i represents the differences in the two measurements for case i and N is the number of subjects. For the case where more than two sets of measurements are being compared, namely when calculating the overall error between all four sets of landmarks, the corresponding formula is

$$TEM = \sqrt{((\Sigma_{i=1}^{N}((\Sigma_{j=1}^{K} M_{ij}^2) - ((\Sigma_{j=1}^{K} M_{ij})^2/K)))//N(K-1))} \tag{2.7}$$

where K is the number of repeat measurements and M_{ij} is the j-th measurement on subject i. The results for the within-day, between-day and overall landmarking are displayed in Tables 2.5 - 2.7.

As can be seen from the tables, using only the two sets of landmarks from the same day was the most reproducible method, followed by using all four sets of landmarks. However, it was thought that perhaps the points marked on the same day could be influenced by memory, as the marker may recall where they placed the first landmark and be inclined to place the second point in the same place. It was therefore decided that all cases should be landmarked four times over two distinct timepoints, in order to increase reproducibility but minimise the bias arising from this memory effect.

Landmark	x-coordinate	y-coordinate	z-coordinate
ssn	0.52	0.48	0.62
xipho	0.42	0.57	0.27
Rmedial	2.70	3.22	1.60
Rlateral	2.02	3.05	5.37
Rprom	2.78	1.98	1.31
Rinfer	2.70	2.65	2.20
Lmedial	1.37	2.43	0.88
Llateral	0.97	3.11	1.92
Lprom	1.21	1.34	0.30
Linfer	1.50	1.19	0.96

TABLE 2.5: TEM(mm) of the landmarks across all subjects when placed twice on the same day. The average TEM was 1.72mm.

Landmark	x-coordinate	y-coordinate	z-coordinate
ssn	0.44	0.40	0.56
xipho	0.68	0.70	0.35
Rmedial	1.87	2.40	1.19
Rlateral	3.65	5.43	4.44
Rprom	3.01	1.78	0.76
Rinfer	2.90	1.94	9.73
Lmedial	1.54	2.31	1.01
Llateral	2.33	4.89	5.16
Lprom	2.68	2.32	0.96
Linfer	3.52	2.26	1.72

TABLE 2.6: TEM(mm) of the landmarks across all subjects when placed twice on different days. The average TEM was 2.43mm.

Landmark	x-coordinate	y-coordinate	z-coordinate
ssn	0.47	0.47	0.47
xipho	0.67	0.67	0.33
Rmedial	1.92	2.49	1.39
Rlateral	2.72	4.39	3.93
Rprom	2.58	1.66	0.94
Rinfer	2.61	2.11	7.92
Lmedial	1.88	2.34	1.08
Llateral	1.88	4.06	3.47
Lprom	1.86	1.65	0.70
Linfer	2.62	2.19	3.74

TABLE 2.7: TEM(mm) of the landmarks across all subjects when placed twice on each of two different days. The average TEM was 2.17mm.

This multiple landmarking procedure was therefore carried out for all 43 of the patient images, with the average of the four sets of landmarks being used in the following analyses. The reproducibility of each landmark can be seen in Table 2.8, and was calculated by taking the mean distance from the centroid of the four marked points to each point individually, and averaging this over all 44 cases. This average distance was found to be within 5mm in all cases, comparable to the degree of accuracy obtained by Brown et al. (1999). The average value over all landmarks was 2.7mm.

Landmark	Average Distance(mm)	Standard Deviation
ssn	0.610	0.704
xipho	0.794	0.869
Lprom	1.753	2.234
Rprom	1.984	2.782
Lmedial	2.393	2.761
Rmedial	2.488	2.980
Linfer	3.249	4.385
Llateral	4.352	4.908
Rlateral	4.685	5.619
Rinfer	4.850	7.450

TABLE 2.8: Reproducibility of final landmarks.

2.1.3 Reproducibility of Subjective Scores

There was one final area where validation was required before the data analysis could be conducted. As a means of validating the methods and results, subjective scores for the symmetry of the patients had been obtained from 6 expert observers. These were consultants in plastic surgery who were routinely undertaking breast reconstructive procedures. The subjective evaluation was carried out on 2-d photographs and these were scored according to the Harris scale, a four point grading system where 1 represents a poor result and 4 an excellent result. Symmetry was judged for appearance of the reconstructed breast in relation to the opposite side. However, in order to ensure that these scores were a valid comparison for our results, it was necessary to check the agreement that the reviewers had between each other, as well as how consistent they were individually. To investigate this, each observer scored

all 43 patients, as well as 6 control images. Furthermore, 10 repeat images were also added in order that the intra-rater agreement could be assessed.

Inter-rater Agreement

To measure the agreement between two or more observers, a statistic must be used that takes into account the fact that observers will sometimes agree or disagree simply by chance, as discussed in Fleiss (1981). The kappa statistic can be used for this purpose, with a kappa of 1 indicating perfect agreement and a kappa of 0 indicating agreement equivalent to chance. Using the notation of Fleiss, if we have m ratings on n subjects and they are rated into k categories (where $k > 2$), the overall kappa statistic κ is calculated as

$$\kappa = \frac{1 - nm^2 - \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2}{nm(m-1)\sum_{j=1}^{k} \bar{p_j}\bar{q_j}}, \tag{2.8}$$

where x_{ij} is the number of ratings on subject i into category j, $\bar{p_j}$ is the overall proportion of ratings in category j and $\bar{q_j} = 1 - \bar{p_j}$. Standard errors can then be calculated as

$$s.e.(\kappa) = \frac{\sqrt{2}}{\sum_{j=1}^{k} \bar{p_j}\bar{q_j}\sqrt{nm(m-1)}} \times \sqrt{\left(\sum_{j=1}^{k} \bar{p_j}\bar{q_j}\right)^2 - \sum_{j=1}^{k} \bar{p_j}\bar{q_j}(\bar{q_j} - \bar{p_j})} \tag{2.9}$$

Equation 2.8 results in a κ value between 0 and 1, and a guideline for interpreting this value, given by Landis and Koch (1977), is shown in Table 2.9.

The kappa score calculated for the agreement between all size raters was $\kappa = 0.646$, with a 95% C.I of (0.599, 0.694), therefore corresponding to substantial agreement beyond chance. A z-score can then also be calculated as

$$z = \frac{\kappa}{s.e.(\kappa)}, \tag{2.10}$$

Kappa	Agreement
< 0	*Less than chance agreement*
0.01-0.20	*Slight agreement*
0.21- 0.40	*Fair agreement*
0.41-0.60	*Moderate agreement*
0.61-0.80	*Substantial agreement*
0.81-0.99	*Almost perfect agreement*

TABLE 2.9: Guideline for interpretation of the kappa statistic.

leading to a p-value which can be used to determine whether the κ statistic is significant. The z-score was here found to be 26.89 with corresponding p-value < 0.0001, indicating that the statistic is highly significant. However, the null hypothesis for this test is that the sets of ratings are completely independent. Given that the ratings are on the same subject, this hypothesis of independence is highly unlikely, and thus the p-value is perhaps not overly informative. More can be determined from the magnitude of the κ statistic.

Intra-rater Agreement

Using the 10 repeats that were included, kappa statistics were also calculated for how strongly each of the raters agreed with themselves. For the comparison of only two sets of ratings, κ is calculated in a slightly different manner. A table of proportions is first calculated, where each entry p_{ij} $(i = 1, ..., k, j = 1, ..., k)$ is the proportion of ratings that are in category i for the first set of ratings and category j for the second. The observed proportion of agreement is then $p_o = \sum_{i=1}^{k} p_{ii}$, and the proportion of chance-expected agreement can be calculated as $p_e = \sum_{i=1}^{k} p_{i.}p_{.i}$, where $p_{i.}$, $p_{.i}$ are the sums of the i-th row and column respectively. The statistic is then calculated as

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{2.11}$$

with a standard error given by

$$s.e.(\kappa) = \frac{1}{(1 - p_e)\sqrt{n}} \sqrt{p_e + p_e^2 - \sum_{i=1}^{k} p_{i.}p_{.i}(p_{i.} + p_{.i})}. \tag{2.12}$$

Calculating κ for each of our six reviewers individually, the values represented a fair to substantial level of agreement, although not all of the statistics were found to be significant at p = 0.05 or less. This could perhaps be due to the small sample size however, as only 2 sets of 10 ratings were being compared, which naturally leads to less certainty in the calculated statistics. The values and their significance can be seen in Table 2.10.

Rater	Kappa	Level of agreement	p-value
1	0.355	Fair	0.134
2	0.545	Moderate	0.011
3	0.322	Fair	0.099
4	0.531	Moderate	0.013
5	0.286	Fair	0.159
6	0.643	Substantial	0.005

TABLE 2.10: Intra-rater agreement for each of our 6 observers.

Through these various validation procedures, the reliability of the subjective scores and landmarks were shown to be satisfactory, as was the accuracy of the image capture system for volume measurement. As such, confidence could be placed in the data and any results that would be derived from them. The data from this study is first seen in Chapter 4 in an illustration of the feature identification methodology. The main goal of the study, namely assessing the asymmetry present in the patients, is then tackled in Chapter 5.

2.2 Cleft Lip and Palate Data

The second dataset explored in this thesis will now be introduced, and is from a study comparing the facial shapes of children born with a cleft lip and/or palate to those of a group of control children. Three-dimensional images were captured at the Glasgow Dental Hospital, again by a stereophotogrammetry system, at the ages of 3, 6, 12, 24, 36, 48 and 60 months. Data were available on 92 patients and 198 control children, all of whom had been captured at several timepoints. Each of the children in the patient group had undergone corrective surgery between the ages of 3 and

6 months, and interest therefore lay in determining the differences that remained between these children and the control group as time went on. Validation studies have previously been conducted for this data in order to examine both the reliability of the imaging system as well as the reproducibility of the landmarks. These studies will not be discussed here, however full details can be found in Ayoub et al. (2003).

As previously mentioned, the driving motivation behind this study was an examination of the differences in facial shape between the cleft and control groups, and this has previously been assessed in terms of both landmarks points and curves by Hood et al. (2004), Barry and Bowman (2008) and Miller (2009). This question will briefly be revisited here, with an analysis of the differences in terms of the surface data being presented in Chapter 5. For the main purposes of this thesis, however, the analysis of this dataset will focus on the control data alone. While taking advantage of the longitudinal nature of the data, the aim is to build a model which describes the average facial growth seen in children between the ages of 3 and 60 months, in terms of the entire facial surface. The first step in doing so involves the creation of a set of corresponding surfaces, the methods for which are described in Chapter 3. Principal components are then used as a means of dimension reduction before a linear mixed effects model is fitted to the resulting principal component scores. The model and subsequent findings are discussed in detail in Chapter 7.

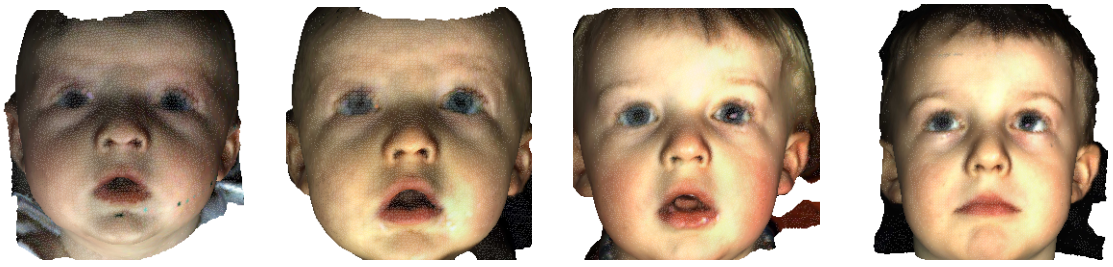

FIGURE 2.8: A control child from the dataset captured at 3, 12, 24 and 60 months respectively.

2.3 Orthognathic Data

The third and final dataset used in this thesis was again collected at the Glasgow Dental Hospital, and consists of patients who have undergone an orthognathic surgical procedure. This is a type of surgery that aims to correct deformities of the jaw, for example where a patient has either an overly recessive or overly prominent jaw bone, and is normally carried out for a combination of both medical and aesthetic reasons. It involves the cutting and realigning of either the maxilla (upper jaw bone), mandible (lower jaw bone) or both, without which the patient may have problems with breathing and mastication, as well as an unusual appearance. The aim of this study is to assess how well the post-operative outcome can be predicted, based on knowledge of both the pre-operative appearance of the patient and the planned surgical procedure.

In order to investigate this, data was available from 13 patients who had undergone this type of procedure. For each patient both a pre- and post-surgical image had been collected by means of a CT scan. However, unlike stereophotogrammetry which only captures the topology of the outer surface of an object, CT scans also give information on the internal shape and structures, in this case the bone and muscle tissues of the head. It was therefore necessary to extract both the bone and soft-tissue surfaces individually, in order that the soft tissue reaction could be modelled independently. This extraction process was performed in *MeVisLab*® (MeVis Medical Solutions AG, Bremen), a freely available software for medical image processing and visualisation. The output files from CT scanners are in DICOM format and are composed of pixels, each of which contains information about the density of the tissue at that particular point. Using the *WEMIsoSurface* module in this package, a threshold can be altered that controls the density of tissues that are displayed. Due to the very different material properties of the two surfaces, this therefore allows them to be displayed and saved independently. An example of the extracted bone and soft tissue surfaces for one of the patients can be seen in Figure 2.9.

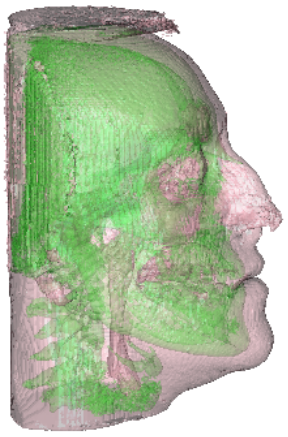

FIGURE 2.9: The pre-operative bone (green) and soft tissue (pink) surfaces for one of the patients.

As a result of this segmentation process, four images were available for each patient - namely the pre- and post-operative bone and soft tissue surfaces. The aim is then to use these images to model the deformation occurring in the soft tissue due to the surgery, and hence the post-surgical appearance, based on the changes that have occurred in the underlying bone. This information could then be used to build a prediction model for the post-operative outcome in future cases, something which is desirable both as a surgical planning tool and as a means of preparing the patient mentally for the change.

The orthognathic data is used illustratively in Chapter 4, before the key issue of surface prediction is tackled in Chapter 7.

Chapter 3

Landmark Methods and Surface Representation

Having introduced the three datasets, it is now important to consider some standard methodology and notation prior to commencing the analyses.

As discussed in Chapter 1, the field of shape analysis is an active and growing area of research with many applications in medicine, biology and beyond. As such, a wide variety of techniques have been proposed for the analysis of this type of data. However, despite the wide-ranging applications, there are some common features that are present with all shape data, and a standard set of tools are therefore available which prove useful in many of these applications. Several of these standard techniques are utilised throughout this thesis, and are therefore discussed in Section 3.1 in order to provide a basis for much of the work that follows. A more comprehensive description of the existing methods in shape analysis is given by Dryden and Mardia (1998) however, in a book which should be taken as the reference for the following section.

As in Dryden and Mardia (1998), the methods are described here in terms of the analysis of landmark points - a generally sparse set of points that are used to represent the shape of an object. However, provided an appropriate set of correspondences exists across cases, the techniques can be easily extended to the case of anatomical

curves and surface data. As the main focus of this thesis is on statistical analysis over surfaces, a discussion of the creation of these correspondences is then given in Section 3.2.

3.1 Statistical Shape Analysis

As defined by Dryden and Mardia (1998), shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. Shape is therefore unchanged by any of the Euclidean similarity transformations - those being translation, scaling and rotation.

Shapes have traditionally been defined by a finite set of points on an object. These points correspond between and within populations and are known as landmarks. More often than not, these landmarks correspond to points of anatomical significance on the object, for example the tip of the nose or the corner of an eye. However, they can also be in the form of mathematical landmarks or pseudo-landmarks. The former are points derived from the geometrical information of the object, such as a point of maximal curvature, while the latter are obtained by the interpolation of either anatomical or mathematical landmarks.

A set of landmarks is generally considered as a $k \times m$ matrix X, known as the landmark configuration. In this notation, k represents the number of landmarks while m is the number of dimensions, with each row of X therefore corresponding to a different landmark point.

3.1.1 Procrustes Methods

As the shape of an object is invariant under the Euclidean similarity transformations, the location, orientation and size of the object are not of significance. However, when comparing shape across a variety of objects $X_1, \cdots, X_n$ say, it is obviously necessary to remove these effects in order that a valid comparison of their shapes can be conducted. This can be achieved through a Procrustes alignment of the

configurations, for which two standard methods exist. In the case where only two objects are to be aligned, i.e. when $n = 2$, the configurations can be matched by a full Ordinary Procrustes Analysis, while a full Generalised Procrustes Analysis can be applied more generally for any $n \geq 2$.

To explain these methods it is first necessary to define some useful terms. Firstly, as shape has been defined as being invariant under scaling it is useful to be able to remove any size effects when comparing the shape of distinct objects. This requires a measure of size $g(X)$ such that

$$g(aX) = ag(X)$$

for any positive, real value a. Several such size measures have been presented, however the most common measure (and the one used throughout this thesis) is the *centroid size*, $S(X)$. This is calculated as

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2}, \tag{3.1}$$

where $\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}$, $\|X\| = \sqrt{trace(X^T X)}$ and C is the centering matrix given by $C = 1_k - \frac{1}{k}1_k 1_k^T$. This size measure can therefore be interpreted as the square root of the sum of squared distances from each of the landmarks to their centroid, $\bar{X}_j$.

We can then define the *set of similarity transformations* for a configuration X as the set of all configurations that have the same shape as X. This is given by

$$\{\beta X\Gamma + 1_k\gamma^T : \beta \in \mathbb{R}^+, \Gamma \in SO(M), \gamma \in \mathbb{R}^M\}, \tag{3.2}$$

where β is a scale parameter, γ is a translation vector of dimension m and Γ is an $m \times m$ rotation matrix satisfying $\Gamma^T\Gamma = \Gamma\Gamma^T = I_m$ and $|\Gamma| = \pm 1$. The set of all such rotation matrices is known as the *special orthogonal group*, and is denoted by $SO(m)$.

In order to explain the Procrustes methods for alignment, it is finally useful to define Z, the *pre-shape* of X . This is the configuration calculated as

$$Z = \frac{HX}{\|HX\|}. \tag{3.3}$$

Here, H is the Helmert sub-matrix, a $(k-1) \times k$ matrix that is used to remove location. As Z is scaled to unit size, the pre-shape of an object is invariant under both scaling and translation.

3.1.1.1 Full Ordinary Procrustes Analysis

Let us first consider the case where it is required to align two configurations, say X_1 and X_2, specifically by the matching one configuration onto the other. This alignment is then achieved through a full ordinary Procrustes analysis (OPA), a method which provides a means of obtaining the optimal least squares match between the two configurations. The first step in deriving this match involves the centering of both configurations on the origin, which can be achieved by

$$X_{1C} = CX_1, \qquad X_{2C} = CX_2,$$

where C is the centering matrix defined previously.

The matching process for matching X_1 onto X_2 is then achieved by determination of the parameters $\hat{\beta}$, $\hat{\Gamma}$ and $\hat{\gamma}$ that are required to minimise the squared Euclidean distance

$$D^2_{OPA}(X_1, X_2) = \|X_2 - \beta X_1 \Gamma - 1_k \gamma^T\|^2. \tag{3.4}$$

The parameters ($\hat{\beta}$, $\hat{\Gamma}$, $\hat{\gamma}$) therefore describe the similarity transformation of X_1 that minimises the distance to X_2, and are determined as follows. As X_1 and X_2 have been centred, clearly $\hat{\gamma} = 0$. It can then be shown that the optimal rotation matrix $\hat{\Gamma}$ is found by

$$\hat{\Gamma} = UV^T, \tag{3.5}$$

where $X_2^T X_1 = \|X_1\|\|X_2\| V \Lambda U^T$, Λ is an $m \times m$ matrix of positive elements, and $U, V \in SO(m)$. $\hat{\beta}$ is then found by

$$\hat{\beta} = \frac{trace(X_2^T X_1 \hat{\Gamma})}{trace(X_1^T X_1)}. \tag{3.6}$$

These parameters provide the minimised value of Equation 3.4, which is known as the Ordinary Procrustes Sum of Squares (OSS) and is calculated by

$$OSS(X_1, X_2) = \|X_2\|^2 \sin^2 \rho(X_1, X_2). \tag{3.7}$$

Here ρ is a measure of the distance between two configurations known as the Procrustes distance. For two configurations X_1 and X_2 with corresponding pre-shapes Z_1, Z_2, this distance is calculated as $\rho(X_1, X_2) = \arccos(\sum_{i=1}^{m} \lambda_i)$, where $\lambda_1, \cdots, \lambda_m$ are the square roots of the eigenvalues of $Z_1^T Z_2 Z_2^T Z_1$.

Due to the fact that this alignment involves the matching of one configuration onto another, namely by treating the latter as a 'target' matrix, the superimposition will generally not be the same when the order of X_1 and X_2 is reversed. Unless the two configurations are of the same size the matching process is not reversible, i.e. $OSS(X_1, X_2) \neq OSS(X_2, X_1)$, and therefore $\sqrt{OSS(X_1, X_2)}$ is not a suitable measure of distance between the two configurations. However, by scaling each configuration to be of unit size, the value

$$OSS\left(\frac{X_1}{\|X_1\|}, \frac{X_2}{\|X_2\|}\right)$$

is invariant under a reordering of the configurations, and therefore provides a suitable measure of distance between the two configurations.

3.1.1.2 Full Generalised Procrustes Analysis

Let us now consider the case where it is required to align $n \geq 2$ configurations, say $X_1, \cdots, X_n$. The method of full generalised Procrustes analysis is then implemented and involves the translation, rotation and scaling of configurations in order

to minimise the sum of squared pairwise distances between them, i.e. to minimise the equation

$$G(X_1, \cdots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - (\beta_j X_j \Gamma_j + 1_k \gamma_j^T)\|^2 \tag{3.8}$$

This minimum is called the generalised Procrustes sum of squares, and it is found subject to the constraint that the average aligned configuration $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} (\beta_i X_i \Gamma_i + 1_k \gamma_i^T)$ is of unit size, i.e. that $S(\bar{X}) = 1$.

The minimum value of Equation 3.8 can then be found through a pairwise comparison of all sets of similarity transformations for $X_1, \cdots, X_n$, i.e.

$$\begin{aligned} G(X_1, \cdots, X_n) &= \inf_{\beta_i, \Gamma_i, \gamma_i} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - (\beta_j X_j \Gamma_j + 1_k \gamma_j^T)\|^2 \\ &= \inf_{\beta_i, \Gamma_i, \gamma_i} \sum_{i=1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \frac{1}{n} \sum_{j=1}^{n} (\beta_j X_j \Gamma_j + 1_k \gamma_j^T)\|^2. \end{aligned} \tag{3.9}$$

On the other hand however, the problem can instead be viewed as an estimation of the mean shape μ, which leads to the alternative definition of

$$G(X_1, \cdots, X_n) = \inf_{\mu : S(\mu) = 1} \sum_{i=1}^{n} OSS(X_i, \mu). \tag{3.10}$$

A description of the algorithm for this alignment will not be given here, however a detailed explanation can be found in Dryden and Mardia (1998). The resulting minimising parameters for a given X_i are denoted $(\hat{\beta}_i, \hat{\Gamma}_i, \hat{\gamma}_i)$, and provide the full Procrustes fit of X_i as

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i{}^T. \tag{3.11}$$

The full Procrustes mean fit $\hat{\mu}$ has the same shape as the arithmetic mean of the set of Procrustes fits, $X_1^P, \cdots, X_n^P$, and can be calculated by

$$\hat{\mu} = \arg \inf_{\mu : S(\mu) = 1} \sum_{i=1}^{m} \sin^2 \rho(X_i, \mu), \tag{3.12}$$

where ρ is again the Procrustes distance seen in Equation 3.7.

3.1.1.3 Partial Procrustes Methods

There are many applications where the scale of an object is of interest in addition to its shape, and for this case there exist variants of both the OPA and GPA algorithms that incorporate this added size effect. These algorithms are known as ordinary partial Procrustes analysis and generalised partial Procrustes analysis respectively, and they involve the matching of configurations over rotation and translation alone, thereby retaining all differences in scale. The algorithms for these partial Procrustes methods follow intuitively from the methods of full alignment described here, however the details of these standard techniques can be found in Dryden and Mardia (1998).

3.1.1.4 Tangent Coordinates

As has been seen in the previous sections, Procrustes methods can be used in order to determine the mean of a given set of shapes. A useful approach to the statistical analysis of shapes can then be to examine the variation that exists about this average shape, as opposed to a direct analysis of the variation in the original coordinates. By examining the deviations from this mean, the data is projected to a linearised version of the shape space, known as the tangent space. The resulting projections are known as tangent coordinates, and a good approximation to these are given by the Procrustes residuals. These are the point-wise displacements between the Procrustes fit of a given configuration and the Procrustes mean configuration, and are calculated as $X_i^P - \hat{\mu}$.

An advantage of this approach is that, provided the data are not too highly dispersed, multivariate methods in this tangent space that are based on distances from $\hat{\mu}$ will be equivalent to non-Euclidean methods in shape space involving shape distances, such as the Procrustes distance ρ for example. As will be seen in later chapters, these tangent coordinates are therefore very useful and allow for the application

of standard multivariate techniques, e.g. principal components analysis, within this linearised space. The results of these multivariate analyses can subsequently be projected back into the shape space to allow for their straightforward visualisation and interpretation.

3.2 Creation Of Corresponding Surfaces

A description of several landmark-based techniques has now been given, from which it was seen that these points are very useful in terms or removing the effects of location, orientation and scale from an object, as well as providing a straightforward method for comparison across shapes. A drawback of this type of shape representation however, is that it generally provides a very sparse description of the surface. For a complex surface, such as the human face for example, it would be implausible to place enough landmarks on the image in order to describe the low-level details of the shape accurately. Additionally, images that are captured by modern techniques such as stereophotogrammetry and CT scanning are generally of a very high resolution - often consisting of many tens of thousands of points - and the chosen set of landmarks tends to be a very small subset of these. As such there is a lot of shape information that is lost by this approach.

While landmarks are generally points of anatomical significance on an object, the key feature that makes their analysis attractive is that they correspond across different objects from a population, and therefore act as points of reference for comparison. Landmark-based methods are therefore easily extendable to the analysis of curves or surfaces, provided a similar correspondence can be established between the various objects.

As discussed in Wang et al. (2000) and Lipman and Funkhouser (2009) amongst others, this creation of such point correspondences has wide-reaching applications in the fields of computer graphics and medical imaging, and as such a variety of methods have been presented. While a full discussion of this subject is outwith the scope of this thesis, the methods available are continually advancing. For example,

a recent approach presented by Lipman and Funkhouser (2009) has been shown to be capable of finding correspondences between input objects which are in distinct poses, and even those that are of slightly different forms, such as a cat and dog. Theirs is a fully automated method and involves the repeated sampling of a triplet of points from each of the two objects of interest, in order to provide a mapping for both through the Möbius transformation. For each pair of triplets, the extent of correspondence can then be examined and the optimal mappings can be deduced.

However, by far the most common approach to the determination of point correspondences is through the registration, or spatial alignment, of the images of interest. The points on each aligned surface that relate to one another can then be subsequently determined. This registration can be performed in a variety of ways, either by means of a global or local transformation, or alternatively a combination of the two. Global transformations are generally rigid transformations that deal with large scale differences in the surfaces, e.g. an affine transformation, while local transformations deal with any differences present at a smaller, more localised scale. Summaries of the literature in this area can be found in Zitov and Flusser (2003) and Brown and Rusinkiewicz (2007), and demonstrate a wide variety of the techniques available for image registration.

For the purposes of this thesis, however, the technique of Thin-Plate Spline warping has been implemented. These splines are standard mathematical tools that were introduced by Duchon (1976) and first proposed in a shape analysis setting by Bookstein (1989). They have become a very common method of image registration and, as discussed in Zitov and Flusser (2003), are particulary appropriate for the registration of medical images, where changes are mainly expected to occur at a local level. The details behind this warping technique are discussed in Section 3.2.1, while Section 3.2.2 details the specific application to the cleft lip and palate data.

3.2.1 Thin Plate Spline Warping

The thin plate spline is a specific type of radial basis function that has been used extensively in the registration of medical images (Brown and Rusinkiewicz 2007).

It is a non-rigid mapping of one surface onto another that is driven by a set of landmark points on each surface. Each landmark is mapped to its counterpart on the target surface exactly, while an interpolation of the resulting mapping functions provides a smooth transform for the remainder of the surface points. The warp is composed of affine and non-affine elements, and the optimal non-affine warp is selected by minimising the 'bending' of the surface, thereby retaining the original shape as much as possible.

As described in Bookstein (1989), although here extended to the case of three dimensions, the mapping function required to warp a surface with landmark points $P_i = (x_i, y_i, z_i)$, $i = 1, \cdots, k$, to a target surface can be described by the equation

$$f(x, y, z) = a_1 + a_x x + a_y y + a_z z + \sum_{i=1}^{k} w_i U(|P_i - (x, y, z)|), \tag{3.13}$$

where the parameters (a_1, a_x, a_y, a_z) represent the affine (or global) component of the transformation and the sum of the functions U represent the non-affine component. The function U is the basis function for the warp and is given by Bookstein (1989) as $U(r) = r^2 \log r$ when working in two dimensions, or alternatively $U(r) = |r|$ for a warp in three dimensions.

To calculate the warp, the required elements of Equation 3.13 can be determined through the creation of a series of matrices, which are defined as follows. Firstly, by taking r_{ij} as the distance between the points P_i and P_j, the matrix K is defined such that

$$K = \begin{bmatrix} U(r_{11}) & U(r_{12}) & \cdots & U(r_{1k}) \\ U(r_{21}) & U(r_{22}) & \cdots & U(r_{2k}) \\ \cdots & \cdots & \cdots & \cdots \\ U(r_{k1}) & U(r_{k2}) & \cdots & U(r_{kk}) \end{bmatrix}, k \times k. \tag{3.14}$$

As all distances r_{ii} will naturally be zero, the diagonal elements of K are also found to be zero. The matrix P is then defined from the landmark coordinates as

$$P = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_k & y_k & z_k \end{bmatrix}, k \times 4, \tag{3.15}$$

and these two matrices combine to give

$$L = \begin{bmatrix} K & P \\ P^T & O \end{bmatrix}, (k+4) \times (k+4), \tag{3.16}$$

where O is a 4×4 matrix of zeros.

For a target surface with landmark points $V_i = (x'_i, y'_i, z'_i)$, $i = 1, \cdots, k$, the affine parameters (a_1, a_x, a_y, a_z) and non-affine weights $W = (w_1, \cdots, w_k)$ can then be found through the equation

$$L^{-1}Y = (W \mid a_1 \; a_x \; a_y \; a_z)^T, \tag{3.17}$$

where $Y = (V \mid 0\ 0\ 0\ 0)^T$. The resulting parameters define three interpolations, one for each of x, y and z, which transform P onto the points V such that I_f, the bending occuring on the mapped surface, is minimised. In terms of Equation 3.13, this occurs when the sum of the squared second-order partial derivatives of f is minimised, i.e. at the minimum value of

$$I_f = \int \frac{\delta^2 f}{\delta x^2} + \frac{\delta^2 f}{\delta y^2} + \frac{\delta^2 f}{\delta z^2} + 2\left(\frac{\delta^2 f}{\delta x \delta y} + \frac{\delta^2 f}{\delta x \delta z} + \frac{\delta^2 f}{\delta y \delta z}\right) dxdydz \tag{3.18}$$

The resulting warp thus ensures a smooth surface that is mapped directly onto the target landmarks, yet has retained its original shape as much as possible.

3.2.2 Application to the Cleft Data

To allow for an analysis of the facial surfaces in the cleft palate dataset, a set of corresponding surfaces was created. Following the method of Hutton et al. (2001), the approach involved the mapping of each image in the dataset to a pre-defined 'base mesh' - a dense surface that is of a similar form to the data and from which the correspondences can be derived. As defined by Hutton (2004), the term dense here refers to the fact that the surface can be represented by a set of points where the inter-point distances are less than the size of any of the surface features, thus allowing for an accurate representation of the topology of the surface. The previously mentioned study by Hutton et al. (2001) used this technique in order to assess the patterns of variability seen in the adult human face, although a further example of their approach can be seen in Hennessy and Moss (2001) in an investigation into facial abnormalities resulting from schizophrenia.

While the thin plate spline provides a transformation for the entire surface, it is derived based solely on a selected set of landmark points, the transformations of which are smoothly interpolated. As such, it is important that a reasonable and reproducible set of landmarks are chosen to drive the procedure. As discussed in Hutton (2004), a larger number of landmarks will in general lead to a more reliable transformation and therefore improved correspondences between objects. However, the computational costs will of course be increased for a greater number of landmarks, as will the time required for landmarking of the data, and therefore a suitable balance should be found.

With this in mind, a set of 22 standard facial landmarks were selected to formulate the warps for the cleft data, and these can be seen in Figure 3.1. As interest in this study lies in examining the effects of cleft lip and palate deformities on facial appearance, a comprehensive set of landmarks was chosen around the nose and lip area. This is clearly where the majority of the effects of the deformity would be seen, and thus serves to ensure that this region is adequately described. It is also essential, however, that the landmarks selected give a reasonable and comprehensive coverage of the remainder of the surface, thereby ensuring that the face is also

warped accurately away from this central region. To this end, four landmarks were also selected around the outer edge of the face, which were in this case based on the position of the ears.

Upon inspection of the images in the dataset however, it was found that the ears had not been adequately captured by the stereophotogrammetry system in all cases, leading to a lack of available landmarks in this region. As would be expected due to the fact that all remaining landmarks were clustered around the centre of the face, registering the meshes without these points was found to have a detrimental effect on the warping process, particularly around the extremities of the face. Any such cases were therefore removed from the dataset. Retaining only those images with at least one landmark on each ear, 87 of the original 877 images had to be excluded from the study.

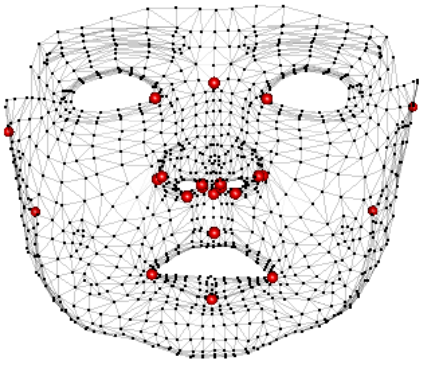

FIGURE 3.1: The base mesh and landmarks used to establish point correspondence.

The remaining surfaces were then individually warped to a pre-defined base mesh in order to determine the correspondences. The base mesh used was composed of $p = 874$ points in regularised positions, and can be seen in Figure 3.1, along with its corresponding 22 landmark points. The mesh was provided by the Department of Computing Science at the University of Glasgow, where it had been created in the package *POSER*® (Smith Micro Software, Inc., CA) and used in the study presented by Mao et al. (2004). The choice of base mesh is not overly important however, provided that it adequately describes the surface of interest. An alternative approach, and the one implemented by Hutton et al. (2001), is simply to treat one

of the existing images in the dataset as the base mesh. However, this will generally be a far denser surface and therefore leads to larger calculations. As can be seen in Figure 3.1 the details of the shape can be very adequately described by a smaller set of points such as those of our base mesh.

For any given image X, the correspondence is then achieved by determining the point on its warped counterpart X' that lies closest to each point of the base mesh Y. Although the bending of X' is minimised during the warping process, the representative points should be taken as the corresponding locations on X instead of those found on X', as this will ensure that the original shape of the object is fully preserved. The necessary set of points, R say, can therefore be denoted by

$$R = \{X : \min_s |Y_t - X'_s|, \;\; t = 1, \cdots, p\}, \tag{3.19}$$

where X_s and Y_t denote the t-th and s-th points of X and Y respectively.

FIGURE 3.2: *Left:* The original image and landmarks (yellow) along with the landmarks from the base mesh (red). *Right:* The same image after warping to the base mesh landmarks.

Each X_i is then represented by a configuration R_i, this being a set of p points whose positions correspond across all cases. These representations can therefore be treated as landmark configurations, although with a much larger set of 874 points each. An example of the resulting points can be seen in the right hand panel of Figure 3.3. Clearly these derived surface representations provide a much more detailed description of the facial shape than the landmarks alone and utilise far more of the

information inherent in the image, therefore providing a more reliable and informative description of facial shape. Additionally, due to their corresponding nature they allow for the use of standard shape analysis techniques, such as those discussed in Section 3.1.

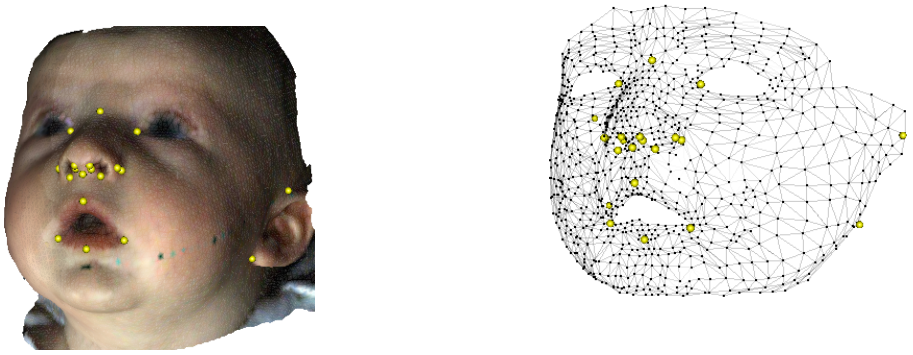

FIGURE 3.3: *Left:* The original image and landmarks. *Right:* The representative surface points found for that image.

The surface representations created here are later analysed in Chapter 5, where a comparison of the facial symmetry present in the cleft patients and control children is conducted. Chapter 7 then sees a further analysis of the control data, consisting of an investigation into the facial growth patterns seen in young children.

Chapter 4

Image Reduction and Feature Identification

The previous chapter saw a description of several useful landmark-based techniques for the analysis of shape, as well as a means of extending these methods to the analysis of a more global representation of the surface. However, although the issue of point correspondence has been dealt with, there remain several issues that can arise with surface data that are not present with a more simple data structure. When working with data captured by stereophotogrammetry, for example, areas which are not a part of the anatomical surface can sometimes be included in the final image. An example of this can be seen in the breast reconstruction data, where parts of the rig in which the women were positioned during image capture (Fig. 2.1) were at times also visible in the output image. This can be seen for one of the patients in Figure 4.1. Clearly this issue would not have an effect when analysing the landmark points alone, but when working with the surface as a whole any external areas such as this must be removed to ensure that they do not impact on the results. Furthermore, the ability to extract a particular area of the surface can also be useful for highlighting a given area as well as in the removal of excess areas, for example when interest lies solely in a particular section of the surface or alternatively a feature within it. In this instance, extracting the area in order to analyse it independently can serve to both reduce the size of the dataset and give a more focused analysis.

This chapter therefore discusses several approaches to the selection of subsections of three-dimensional surfaces, the appropriate method for which is dependent on the form of the area to be extracted and its complexity.

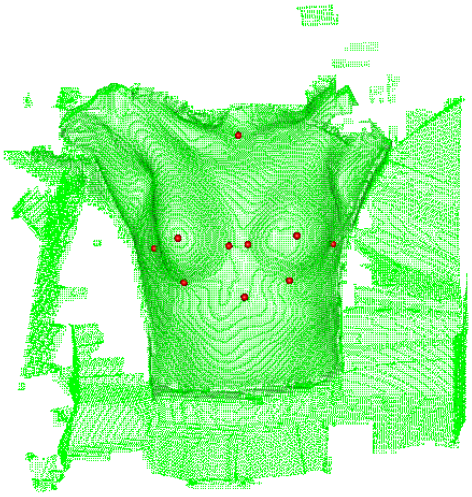

FIGURE 4.1: A breast image including areas not on the anatomical surface.

4.1 Extraction by Projection onto Two Dimensions

An initial and straightforward approach to selecting a given area of an image can be to define a set of boundary points for the desired section and subsequently determine the coordinates that lie within these points. This can be achieved by means of straight plane cuts through the image, however an intuitive way of simplifying the problem is through an appropriate projection of the surface onto two dimensions. The technique required to do so is now illustrated for the image of Figure 4.1.

Firstly, in order to ensure a suitable projection, the image should be orientated in a suitable manner. This can be achieved through the selection of an origin O that is central to the area of interest, upon which the set of surface points must be centred. For the image seen in Figure 4.1, this origin was taken to be the midpoint of L_L and L_R, the landmarks representing the most lateral points on the left and right breasts respectively. It is then required to determine a local set of axes at this

origin. By selecting one direction of the axes system (say $\overrightarrow{a_3}$) to be approximately normal to the surface in the area to be extracted, this allows the coordinates to be projected onto the remaining two axes, $\overrightarrow{a_1}$ and $\overrightarrow{a_2}$. In this instance, $\overrightarrow{a_1}$ was taken to be $\overrightarrow{a_1} = L_R - O$, with a first estimate of $\overrightarrow{a_2}$ being $\overrightarrow{a_2} = ssn - O$. This is the vector running from the origin to the suprasternal notch, which is the uppermost landmark seen in Figure 4.1. By standard vector properties, the vector that is perpendicular to these two axes is given by their cross product, i.e. $\overrightarrow{a_3} = \overrightarrow{a_1} \times \overrightarrow{a_2}$, and therefore gives an approximation to the normal direction of the surface. By then refining $\overrightarrow{a_2}$ to be $\overrightarrow{a_2} = \overrightarrow{a_1} \times \overrightarrow{a_3}$, the orthogonality of all three axes is ensured.

It is then necessary to determine the projections of the surface points onto these first two axes, which are found through an evaluation of the dot product of each point with the corresponding axis. Denoting the set of all 3-dimensional surface points by P, the two-dimensional coordinates $P' = (P'_1, P'_2)$ are therefore calculated as

$$P'_1 = P \cdot \overrightarrow{a_1}, \qquad P'_2 = P \cdot \overrightarrow{a_2}.$$

A plot of the projected coordinate points can be seen in the left panel of Figure 4.2.

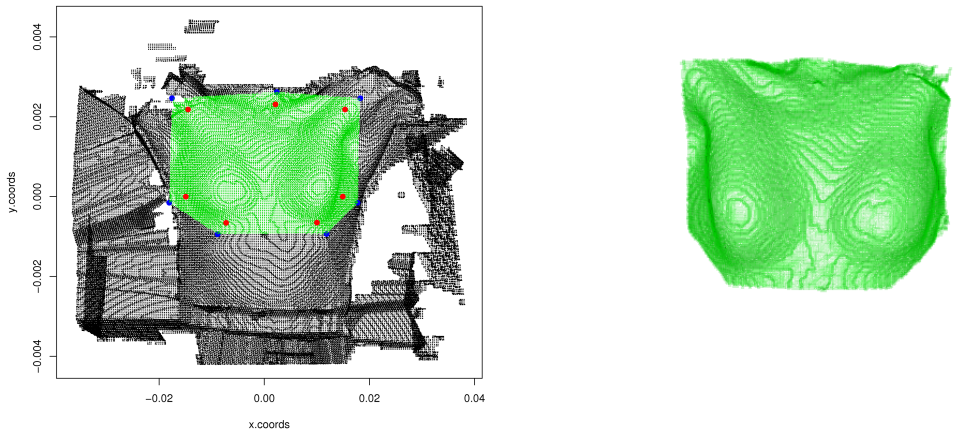

FIGURE 4.2: *Left:* The coordinates after projection onto 2-d, with the boundary points shown in red and the expanded boundaries shown in blue. *Right:* The extracted area of the image.

The set of boundary points for the area of interest must now be determined. In this case, it is necessary to reduce the surface to only the area of the upper chest. The point ssn is therefore taken as an upper boundary, leading to a selection of the boundary points as

$$B = \{I_L, L_L, L_L + ssn, ssn, L_R + ssn, L_R, I_R\},$$

where I_L and I_R are the landmarks representing the most inferior points on the left and right breast respectively. The 2-d projection of these points, B', is shown in red in the left-hand panel of Figure 4.2. However, as anatomical landmarks are often placed directly on the edges of a particular feature, the boundary must often be slightly extended to ensure the inclusion of all necessary points. Taking the centroid $\bar{B}'$ of B', each boundary point B'_i can be expanded by a factor of α in order to obtain the points

$$\mathbf{B'}_i = \bar{B}' + (1 + \alpha)(B'_i - \bar{B}'), \;\; i = 1, \cdots, n.$$

These points can be seen in blue in this same image.

Finally, the subset of P' that lies lie within this expanded set of boundary points can be found. To do so, the vectors that define the perimeter of the bounded region are calculated as

$$V_i = \mathbf{B'}_i - \mathbf{B'}_{i-1}, \qquad i = 1, \cdots, n$$

where $\mathbf{B'}_0 = \mathbf{B'}_n$. Provided the vectors V circle the boundary in a clockwise direction, the direction W_i that is perpendicular to a given vector $V_i = (v_1, v_2)$, and lies in the direction of the origin, can be found through its clockwise rotation through $\frac{\pi}{2}$. This rotation is achieved through

$$W_i = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

which, substituting $\theta = \frac{\pi}{2}$, simplifies to give $W_i = (v_2, -v_1)$. The location of any point $\mathbf{x_j} = (x_j, y_j)$ relative to V_i can then be found through its projection ω_j onto

W_i, again calculated by the dot product $\omega_j = \mathbf{x_j} \cdot W_1$. Comparing a given ω_j to the projection of any point that is known to lie on V_i, for example the boundary point $\mathbf{B}'_i$, it can therefore be determined on which side of the boundary vector the corresponding point $\mathbf{x_j}$ lies. The subset of P' given by $x_j : \omega_j > \omega_{\mathbf{B}'_i}$ therefore provides the set of all points that lie on the interior of V_i, and by repeating this process for all n boundary vectors of V, the set of points found to be on the interior for all vectors will then lie within the boundary region.

The final extracted area is shown in the right panel of Figure 4.2, where it can be seen that the surface has successfully been reduced to the area of interest alone, with all excess areas of the surface removed. This has also significantly reduced the number of surface points, leading to a decrease in the size of any calculations.

As has been demonstrated here, this method can clearly be useful for the selection or removal of certain areas of the surface, or when there is external 'noise' in the image. However, it can only be applied when the area to be selected can be adequately represented by straight cuts through the surface. The desired area is often more problematic than this, with the selection of a more complex shape being required. Two approaches to this situation are therefore discussed in the following sections.

4.2 Point Removal through Connectivity

Similarly to the breast data, the images of the orthognathic patients also included areas of the surface that were not required for the analysis. As can be seen in the left panel of Figure 4.3, the soft tissue images often contained sections of the patient's spinal chord and other internal structures of the head. This is down to the segmentation stage of the data preparation process, when the bone and soft tissue images are extracted separately from the original file. The surfaces are separated based on the density of the tissues, meaning that these other low density tissues still remain in the soft tissue image. These points could clearly affect the analysis when predicting the appearance of the soft tissue after craniofacial surgery, however, as

they would not be affected by the bone movement in the same manner as the outer surface. It was therefore necessary to find a way to remove these points.

As the points were on the interior of the image, the previous method of point selection could clearly not be applied. Doing so would result in a partial loss of the outer surface. A different approach was therefore devised, utilising the triangulation information of the surfaces. As each image is composed of a triangulated mesh, the connectivity of the points is known. It was therefore hoped that the points of connection of the facial surface with these internal structures could be located and used to separate the two areas.

On inspection of the images, it can be seen that these points of connection occur mainly around the nostrils and base of the neck, and these areas should therefore be flagged as cut-off points between the two sets of tissues. By beginning at a point known to be on the exterior of the surface, the connectivity information can then be used to trace along the surface by finding the neighbours at each point, i.e. the set of points that lie in a triangle with that point in the mesh. When a cut-off point is encountered however, no neighbours are searched for that point. Provided all points of connection are flagged as such, this will result in the selection of solely the outer mesh points.

As a first step, the set of cut-off points must be determined. Taking y as the coordinate component determining the vertical position of the points, the cut-off points C_1 that lie at the base of the neck can be found by taking those points that are within a small vertical distance of the lowest point in the mesh i.e. the points given by $\{P : P_{yi} - min_i(P_y) < \epsilon\}$, for some small ϵ. To find the cut-off points on the nose, C_2, the five landmarks of each nostril can be utilised. Fitting a smooth curve to each of these sets of landmarks points, all points within a given Euclidean distance of these curves can be defined as cut-off points. The full set of cut-off points is then $C = C_1 \cup C_2$, an example of which can be seen in red in the centre panel of Figure 4.3. The algorithm to determine all points on the outer surface then proceeds as follows:

1. Find a starting point p known to be on the exterior of the surface e.g. the surface point closest to one of the soft tissue landmarks.

2. Find all triangles of the mesh that contain p, and therefore all neighbouring points N_p of p. Select only those that do not belong to the set of cut-off points i.e. $n_p = \{N_p : N_p \notin C\}$. Set $\Delta_1 = n_p$.

3. Beginning with $j = 1$, repeat step two for each point q_i in Δ_j to find its neighbourhood n_{q_i}. Set $\Delta_{j+1} = \Delta_j \cup n_Q$, where n_Q is the union of all i neighbourhoods n_{q_i}.

4. Repeat step 3 until no non-flagged neighbours can be found for any q in n_Q.

Provided all points of connection have been specified, this algorithm will find all mesh points that lie on the outer surface. An example of the final surface found can be seen in the right panel of Figure 4.3.

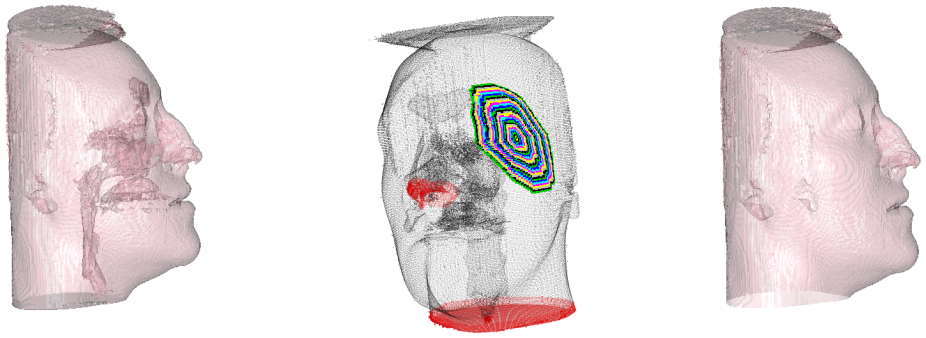

FIGURE 4.3: *Left:* The original soft-tissue image. *Centre:* The same surface showing the set of cut-off points C (red) and the first few iterations of the algorithm. *Right:* The final image after removal of the internal structures.

4.3 Feature Identification

As was shown previously, a specific area of a given surface can be extracted by means of a suitable projection of the surface onto 2-d, provided that the area can be described by a series of plane cuts. When the shape is more complex but has boundary points that are easily identifiable, the connectivity of the surface can then

be made use of to find the desired area. However, it is sometimes necessary to analyse data solely on a particular feature that has neither of these characteristics. Specific features often cannot be well defined by straight plane cuts, nor are the location of their boundaries known. These areas are therefore more problematic to extract, requiring a more complex approach to track where the boundaries of the features lie.

Taking the human body as an example, it is clear that many anatomical features are defined or outlined by areas of high curvature, such as the nose or lips for example. An example of this can be seen in Figure 4.4, where the curvature of a child's face is illustrated. The two plots show the values of maximum and minimum curvature at each point, from which it is apparent that many of the facial features are well described by the curvature present in the surface. It was thought that this information could therefore provide a detailed way in which to determine the form of a given feature.

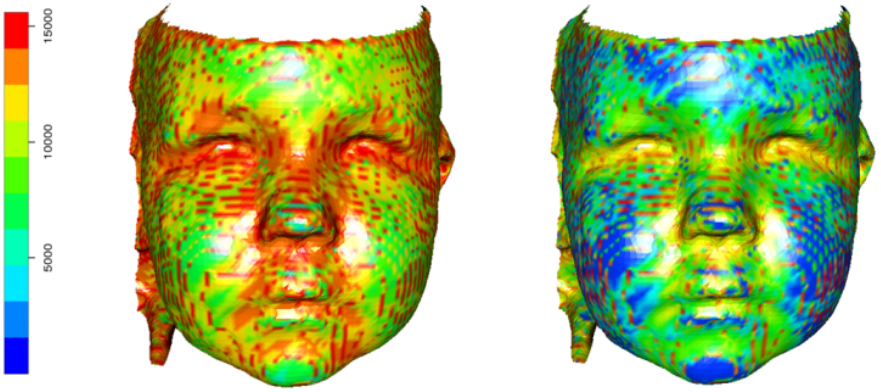

FIGURE 4.4: The maximum (left) and minimum (right) principal curvature scores for the face of a young child.

4.3.1 Surface Curvature

An obvious approach to deducing feature boundaries in this way would be to calculate the surface curvature that was present at each point of the mesh, i.e. the principal curvature scores at each point, as has been done for the facial surface in Figure 4.4. For every point p, a plane can be taken which contains the normal to the

surface at this point and that will cut the surface in a plane curve. The curvature of this curve will be different for different normal planes, and the maximum and minimum values of this curvature, κ_1 and κ_2, are known as the principal curvatures. A description of the method can be found in Spivak (1979) amongst others. Concave areas of the surface will give rise to large values of maximum and minimum curvature, while large negative values of maximum and minimum curvature will correspond to convex areas.

However, this calculation of surface curvature at each individual point can be time consuming, especially over a dense surface consisting of many thousands of vertices such as these. It would therefore be preferable to determine a less computationally intensive approach to the evaluation of these feature boundaries. Although the surfaces are two-dimensional manifolds, it is possible to reduce the complexity of the problem and work in lower dimensions. By taking transects in various directions across a surface, a series of plane curves can be obtained. It is then a far simpler calculation to build up a picture of the surface through the evaluation of these curves as opposed to through the curvature of the surface as a whole. Additionally, as is seen in Figure 4.4, images captured by stereophotogrammetry can often have a slight 'orange-peel' effect, with small ridges present on the surface. A further advantage of the use of smooth curves across the surface, in place of locally calculated curvatures at each point, is therefore that these small areas of artificial high curvature are generally disregarded, leading to a more stable representation of the curvature.

The evaluation of curvature in the transects can be achieved through the use of principal curves, the details of which will now be discussed.

4.3.2 Principal Curves

Principal curves can be thought of as smooth curves which pass through the 'middle' of a dataset. While the principal component line is a linear summary of the data, the principal curve is a non-linear summary in the form of a smooth curve. The technique was introduced by Hastie and Stuetzle (1989), in a paper which should be taken as the reference for this section. In the calculation of the curve, the orthogonal

distances to all data points are minimised subject to certain smoothing constraints. Principal curves also have the property of self-consistency, meaning that each point $\mathbf{x}$ on the curve is the average of all data points that project there, i.e. the average of all data points for which $\mathbf{x}$ is the closest point on the curve.

To provide an ordering along a given curve, it can be useful to paramaterise the curve in terms of its arc length. For an arbitrarily parameterised two-dimensional curve $\mathbf{r}(p) = [f_1(p), f_2(p)]$ which can be represented by n points $(p_1, ..., p_n)$, the length of the curve can be approximated by taking the summation of the Euclidean distances between all consecutive points. This approximation approaches the true length of the curve as n increases. The arc length of a point (p_q say) is the length of the curve until that point, and is therefore approximated by

$$s_q = \sum_{i=2}^{q} \sqrt{\Big[f_1(p_i) - f_1(p_{i-1})\Big]^2 + \Big[f_2(p_i) - f_2(p_{i-1})\Big]^2}.$$

Paramaterising the m-dimensional principal curve by arc length s, it can then be represented by m functions of s, i.e. for $m = 2$, the curve can be given by $\mathbf{f}(s) = \Big(f_1(s), f_2(s)\Big)$. This representation of the curve then proves useful in the calculation of a principal curve for the data.

In order to determine whether a given curve is self-consistent and therefore a principal curve for the data $\mathbf{X}$, the projection $s_{\mathbf{f}}(\mathbf{x})$ of each point $\mathbf{x}$ to the curve must be found. This is simply the value of s for which $\mathbf{f}(s)$ is closest to $\mathbf{x}$, and is given by

$$s_{\mathbf{f}}(\mathbf{x}) = \{s_i : \|\mathbf{x} - \mathbf{f}(s_i)\| = min_i(\|\mathbf{x} - \mathbf{f}(\mathbf{s})\|)\}$$

The curve $\mathbf{f}(s)$ is therefore self-consistent provided

$$E\Big(\mathbf{X}|s_{\mathbf{f}}(\mathbf{X}) = s\Big) = \mathbf{f}(s),$$

When deriving a principal curve, it is necessary to assess whether the curve has converged satisfactorily to the data. To do so, a measure of the average squared

distance from the data points to their projections can be calculated as

$$D^2(\mathbf{X}, \mathbf{f}) = E_s\Big(\|\mathbf{X} - \mathbf{f}(s(\mathbf{X}))\|^2\Big).$$

A suitable first approximation to the principal curve for a given set of data $\mathbf{X}$ can be given by the largest principal component of the data. By averaging the data points which project to each s, it can then be determined whether this approximation is self-consistent. If not, these averaged values can be taken as a new approximation to the curve. If this process converges to self-consistency, the resulting curve is then a principal curve for the data. The algorithm to derive a principal curve can therefore logically be defined as follows:

1. Taking $\boldsymbol{\rho}$ as the largest principal component, obtain a first approximation to the curve by $\mathbf{f}_0(s) = \bar{\mathbf{x}} + \boldsymbol{\rho}s$.
2. For the subsequent j iterations , fix s and set $\mathbf{f}_j(s) = E\Big(\mathbf{X}|s_{\mathbf{f}_{j-1}}(\mathbf{X}) = s\Big)$ in order to minimise $D^2(\mathbf{X}, \mathbf{f}_j(s))$.
3. Set $s = s_{\mathbf{f}_j}(\mathbf{X})$ and return to step 2.
4. Repeat steps 2 and 3 until a curve $\mathbf{f}_j(s)$ is found such that $D^2(\mathbf{X}, \mathbf{f}_j(s)) < \epsilon$, for some desired threshold value ϵ.

Provided this algorithm converges to a suitable distance from the data, the resulting curve $\mathbf{f}_j(s)$ is then a principal curve for the data $\mathbf{X}$ and provides a smooth, non-linear summary of the data. These are well established statistical methods for which software is readily available, such as the `princurve` package in *R* (Hastie and Weingessel, 2009).

4.3.3 Surface Transects as a Means of Dimension Reduction

As was discussed in Section 4.3.1, these principal curves can be used to examine the curvature seen in a series of transects through a given surface. This is a useful

and simplified alternative to an evaluation of the surface curvature itself. When attempting to define a given feature in this manner, the transects should ideally be taken in a direction roughly perpendicular to the edge of the feature, in order to ensure that multiple locations on the boundary can be found. For a transect in a given direction, this requires a way of determining the locations in that direction which lie on the surface, while locations not on the surface are dismissed.

This can be achieved through the creation of a set of orthogonal axes that are centred on the location where the transect will originate. One component of the axes, say $\overrightarrow{a_3}$, must lie in the normal direction to the surface at this point while a further component, say $\overrightarrow{a_2}$, should lie in the desired direction of the transect. By projecting the set of all surface points $p_i : i = 1, ..., n$ onto the remaining axis $\overrightarrow{a_1}$, the distance of each point from the vector $\overrightarrow{a_2}$ is determined. For some small threshold value δ, the points $\{p_i : p_i^T \overrightarrow{a_1} \leq \delta\}$ therefore describe the location of the surface in this direction. When only the points in the positive direction of $\overrightarrow{a_2}$ are required, the final set of transect points is given by $\{p_i : p_i^T \overrightarrow{a_1} \leq \delta \;\; p_i^T \overrightarrow{a_2} \geq 0\}$.

An illustration of this process can be seen in Figure 4.5. The bottom left panel shows the lower facial surface of a patient from the orthognathic dataset, along with the corresponding transects taken across it. In this case, interest lies in determining the jawline of the patient, and so vertical transects have been taken in order to cross this edge in multiple locations. The set of points that lie along the upper edge of the image were found and a subset b of these, taken at regular intervals along the boundary, were taken to be the points of origin for the various transects. At each $b_i \in b$, the normal direction $\overrightarrow{a_3}$ was determined by taking the third component in a PCA of a small neighbourhood of points at that position. The images are oriented such that the vertical direction is known, and this was therefore taken to be $\overrightarrow{a_2}$. As before, the orthogonality of all three axes was ensured by taking $\overrightarrow{a_1} = \overrightarrow{a_2} \times \overrightarrow{a_3}$, and subsequently refining the second component to be $\overrightarrow{a_2} = \overrightarrow{a_1} \times \overrightarrow{a_3}$. Through the appropriate projections onto $\overrightarrow{a_1}$ and $\overrightarrow{a_2}$, a set of transect points were therefore found for each origin point b_i. These points can be seen in blue in the bottom left image of Figure 4.5.

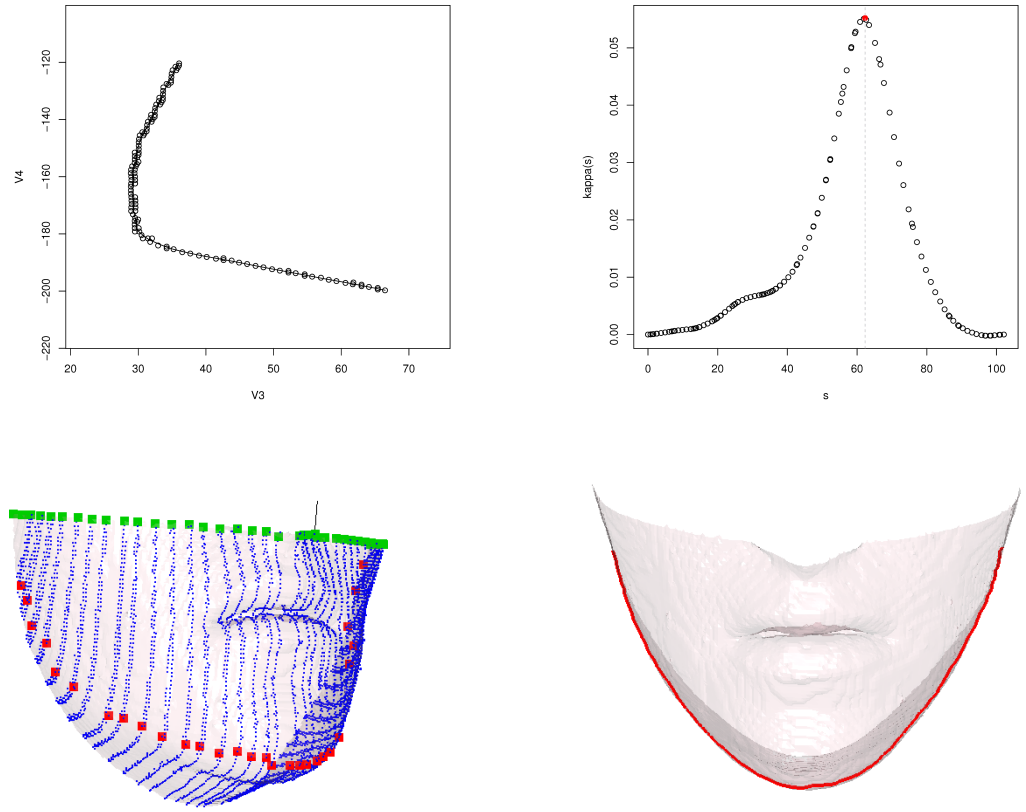

FIGURE 4.5: *Top Row - Left:* The 2-d representation of the points in a particular transect. *Right:* The corresponding curvature function for that transect, with the peak in curvature shown in red. *Bottom Row - Left:* The set of all transects. The origin point of each transect is shown in green with the peaks in curvature shown in red. Points found around the lips have been excluded. *Right:* The final jawline curve.

The curvature of each of these transects must then be examined, which requires each set of points to be represented as a smooth curve. However, as there is very little variation in the points in the direction of $\overrightarrow{a_1}$ due to the way in which the points have been selected, each set of points can be considered in two dimensions without any great loss of information. The principal curve was therefore fitted to the two-dimensional representation of each set of points, further simplifying the curvature calculation. The 2-d representation of a particular transect can be seen in the top left panel of Figure 4.5, along with the corresponding principal curve for those points.

As each principal curve $\mathbf{f}(s) = \Big(f_1(s), f_2(s)\Big)$ is parameterised in terms of arc length, the location of the points along the curve can easily be regularised. By the fitting of splines, say $x(s)$ and $y(s)$, to both $f_1(s)$ and $f_2(s)$ independently, each spline

can then be interpolated to find a set of equally spaced points along the surface in the required direction. The curvature can then be easily computed at each point through the standard formula

$$\kappa = \frac{x'y'' - x''y'}{(x'^2 + y'^2)^{3/2}}, \tag{4.1}$$

which is again described in Spivak (1979). Here, x' and x'' represent the first and second derivatives of x with respect to s, and correspondingly for y' and y''. The curvature function $\kappa(s)$ for the transect of Figure 4.5 is shown in the right hand panel of this figure.

In attempting to define feature boundaries, it is generally the stationary points of the curvature function that are of interest, specifically the local maxima and minima. These correspond to highly curved areas of the surface where the shape of the surface is changing rapidly. These stationary points can easily be found by examining the derivatives of the spline representation of $\kappa(s)$. While their location can be found by setting $\kappa'(s) = 0$, they can be characterised by their value of $\kappa''(s)$, with $\kappa''(s) < 0$ indicating a local maximum and $\kappa''(s) > 0$ a local minimum.

The local maxima were then found in this way for each of the transects, resulting in the red points seen in the bottom left image of Figure 4.5. In this application, any points found due to the curvature in the lips were excluded, resulting in a set of points that lay exclusively along the jawline. The final step then involved fitting a principal curve to these points in order to obtain a smooth representation of the jawline, an illustration of which is seen in the bottom right image of this same figure.

4.3.4 Application to Breast Reconstruction Data

With the breast reconstruction dataset, it is desired to assess the breast symmetry that has been achieved through the surgery, thereby requiring a comparison of the natural and reconstructed breast of each patient. To assess this over the surface of the image, it was therefore necessary to find the boundary of each breast, i.e. where each breast met the chest wall, in order that it could be extracted from the rest of

the image and analysed independently. As can be seen from the surface curvature of the chest in Figure 4.6, it seemed plausible that this could also be done through an analysis of the curvature as in the previous section. The figure shows the minimum curvature values of the surface, hence the presence of extremal values in the concave area around the breast boundaries.

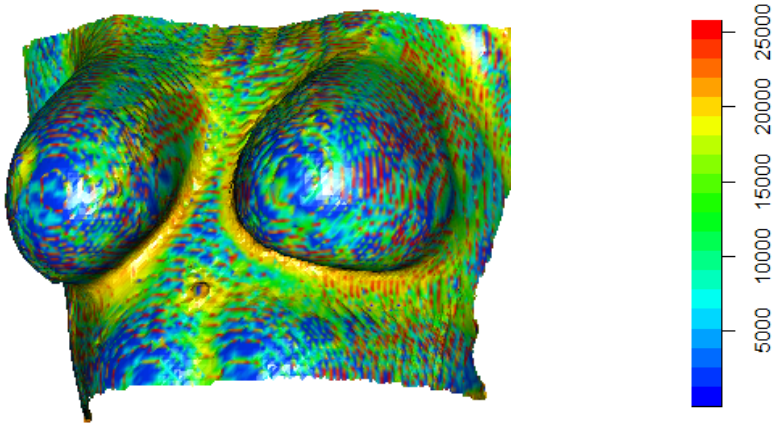

FIGURE 4.6: The minimum curvature values for the chest image of a particular patient.

As before, a local axis system can be determined in order to derive the plane cuts through the surface. In this case, however, the transects for a particular breast all originate from the same surface point, namely the landmark *prom*. This is the landmark representing the most prominent point of the breast, and therefore a transect from this point in any direction should cross the boundary of the breast. For an initial set of axes with normal direction $\overrightarrow{a_3}$ that is calibrated to find the transect in the direction of $\overrightarrow{a_2}$, the sets of points lying in any other direction can be found through a rotation of these axes about $\overrightarrow{a_3}$. The necessary rotation for a given angle α can be calculated by the matrix

$$R = \begin{bmatrix} tx^2 + c & txy - zs & txz + ys \\ txy + zs & ty^2 + c & tyz - xs \\ txz - ys & tyz + xs & tz^2 + c \end{bmatrix}, \tag{4.2}$$

where $c = cos(\alpha)$, $s = sin(\alpha)$ and $t = 1 - cos(\alpha)$, and x, y and z are the respective components of $\overrightarrow{a_3}$. By taking a sufficient number of incremental rotations through 2π radians, the curvature of the entire breast area can then be estimated through the resulting strips.

Again, it is the locations of the maxima in the curvature function that are of interest here. As there is clearly curvature present around the opposite breast, a constraint was included that the strips should not cross the midline of the chest in order to avoid confusion arising from this curvature. This midline was taken to be the line running from the landmarks *ssn* to *xipho*, the locations of which can be seen in red in the bottom left image of Figure 4.7.

Figure 4.7 illustrates the various stages of the boundary identification procedure. The original set of axes and corresponding surface transect can be seen in the bottom left image, while the resulting principal curve and curvature function $\kappa(s)$ are shown in the upper left and right panels respectively. The corresponding two peaks seen in the curvature function are shown in blue in the bottom left image. The set of all strips found for the breast are then seen in the bottom right image, along with the points corresponding to the maxima of $\kappa(s)$ for each of those strips.

As can be seen from the plots of Figure 4.7, it is possible that more than one peak in curvature is found for a given function $\kappa(s)$. While this had occasionally been an issue in detecting the jawline, a simple exclusion of points close to the lip landmarks was found to leave a suitable representation of the required curve in each case. When working with the breast images, however, this issue of multiple peaks proves to be more problematic. Firstly, there are often small bumps or pits on the surface, particularly for the reconstructed breast, which naturally lead to areas of high curvature. In addition, issues are also induced through the position taken by the patients during image capture. While the pose selected was necessary to allow for a full visualisation of the breast surface, the way in which the arms are raised above the head clearly affects the shape of the surface in the upper chest in several cases, making it difficult to detect the boundary in this area. While the curvature is well defined for the patient of Figure 4.7 and therefore only a handful of strips

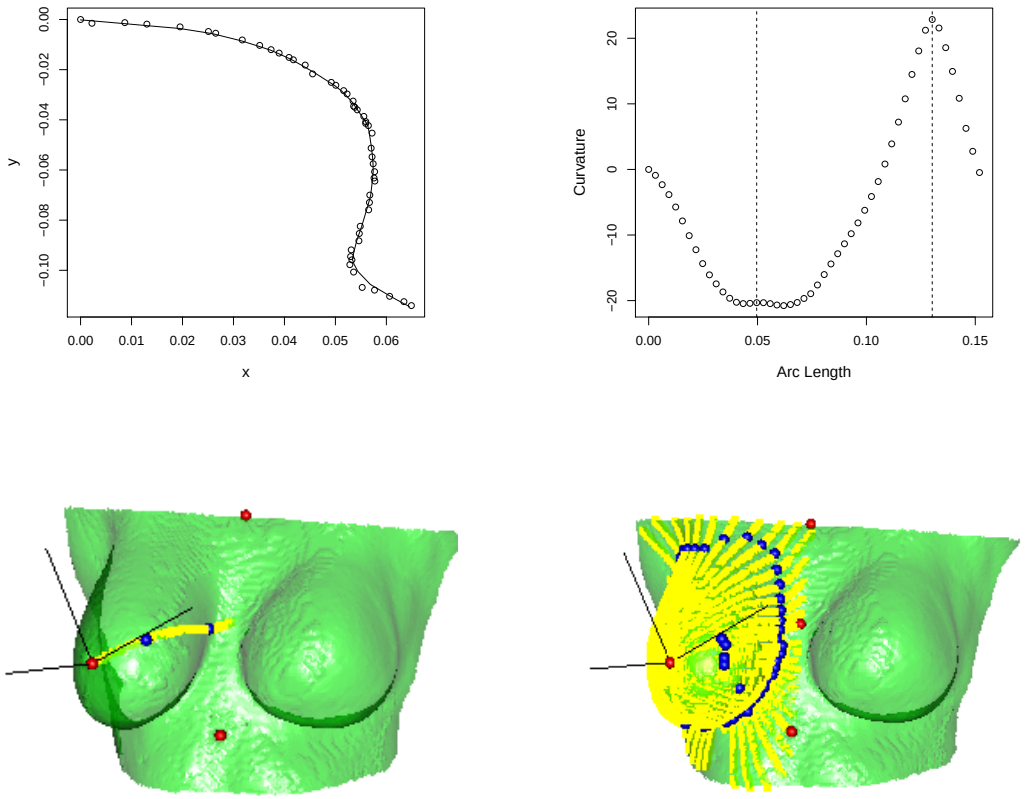

FIGURE 4.7: *Top Row - Left:* The 2-d representation of the points in a particular transect. *Right:* The corresponding curvature function for that transect, showing two peaks in the curvature function. *Bottom Row - Left:* The transect corresponding to the graphs of the top row. *Right:* The set of all transects found for the breast, along with their corresponding candidate boundary points.

have multiple points flagged as candidate boundary points, there are many other cases where the issue is a far more prominent one. An example of this can be seen in Figure 4.8, where multiple points have been found on a large number of the transects for the left breast. It is therefore necessary to determine a way of dealing with these extraneous points.

To this end, several exclusion criteria can usefully be introduced for the boundary points. Denoting the set of all candidate boundary points as b, points which do not belong to the true boundary can be determined and removed as follows:

- Firstly, exclude all points within a small Euclidean distance d_α of the most prominent point i.e. $b \rightarrow b_i : \|b_i - prom\| < d_\alpha$. This removes any problems

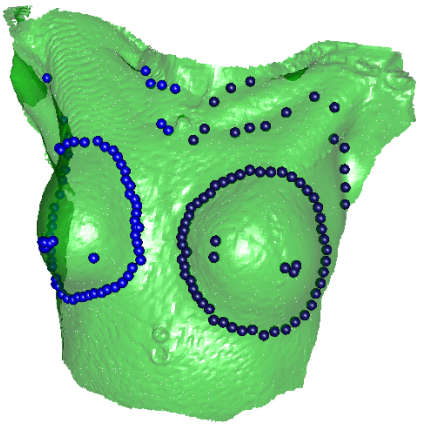

FIGURE 4.8: The candidate boundary points found for a particular patient. The different shades of blue distinguish between the points corresponding to the left and right breast.

that can occur from the curvature present at the nipple.

- Additionally, exclude any points which lie too far above or below π, the plane of the chest wall. This can be achieved by finding the projections P_M and P_L of the most medial and lateral landmarks of the breast, M and L, onto the normal direction of the chest wall. This direction is taken to be the average of the normal directions at *ssn* and *xipho*. Points whose projections lie more than some distance d_β outwith this range are then excluded, i.e. $b \rightarrow b_i : \{P_{b_i} > (\min(P_M, P_L) - d_\beta) \ \wedge \ P_{b_i} < (\max(P_M, P_L) + d_\beta) \ \}$.

- Finally taking b_{yi} as the vertical position of the point b_i, exclude any points that lie too far above the landmark *ssn* or too far below the most inferior landmark I of the breast i.e. $b \rightarrow b_i : \{(b_{yi} - ssn_y) < d_{\gamma_1} \ \wedge \ (I_y - b_{yi}) < d_{\gamma_2}\}$. This is done to remove points that are erroneously selected due to the overhead position of the arms or curvature in the stomach.

Following an application of these criteria, all candidate points that remain should be in a plausible position for the boundary. As a final check however, a principal curve $f(s)$ can be fitted to the set of remaining points, but solely to those candidate points which lie next to a series of other candidate points (in order to remove any

individual points that lie away from the others). An individual point b_i is then only kept provided that it is sufficiently close to a smooth curve through the other points, i.e. provided that $\|b_i - f(s)\|^2 < d_\delta$.

The final step in determining the boundary then requires an interpolation between the existing boundary points, in order to determine where the boundary should lie at any locations where all candidate points have been excluded. To ensure a smooth curve in areas where several points were missing, it was found necessary to fit a spline to the set of points on either side of the gap. The position of each of these splines can then be predicted for this area by means of a linear extrapolation, with a weighted average of the two predictions providing a best approximation of the curve in this location. For a gap with n missing positions and corresponding predicted locations $p_i; i = 1, \cdots, n$ and $q_i; i = 1, \cdots, n$, the weighted average at the position i can be calculated as

$$\frac{N_i p_i + (n - N_i) q_i}{n},$$

where $N = (1, \cdots, n)$.

An illustration of the predicted points for a particular curve can be seen in Figure 4.9. Here, the blue points are the fixed points on which the predictions were based, while the final predictions are shown in red. Taking this combined set of fixed and predicted points, an estimate of the boundary location should then be available at a set number of locations around the perimeter of each breast. The final boundary is again defined by a principal curve through these points, thereby ensuring a smooth description of the feature outline.

Due to the fact that the shapes of the surfaces are very complex and varied, it proved difficult to build a fully automated system for detecting the breast boundary. As previously mentioned, while necessary to allow a visualisation of the entire breast surface, the position undertaken during image capture can often create problems by introducing curvature in areas where it would not normally be present. As a result of this, an increased number of predictions are required for the boundary in this upper region, as compared to the rest of the feature boundary. The reliability of these predictions should therefore perhaps be looked at it greater detail. A particular

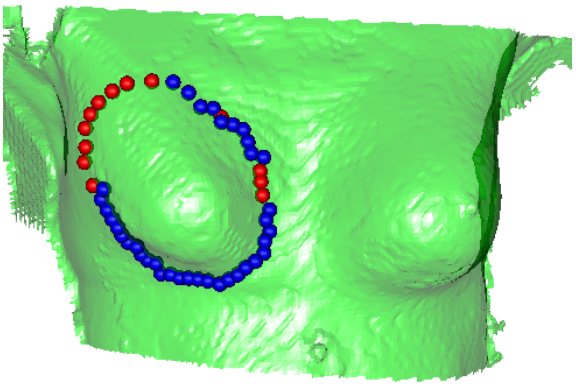

FIGURE 4.9: An image of a chest showing the boundary points through through the curvature of the chest (blue) and the predicted locations that were interpolated between them (red).

difficulty was also found for the cases with smaller breasts, where the curvature was not as well-defined. These various issues meant that it was only possible to build a fully automated system for around half of the 43 patients in the dataset. However, the various thresholds $d_\alpha, d_\beta, d_{\gamma 1}$ and d_{γ_2} provided a standardised way to tailor the process to problem cases, and through their adjustment, a reasonable set of boundaries was obtained for all patients through this method. Several examples of these boundaries can be seen in Figure 4.10.

Creation of Corresponding Surfaces

While these boundaries define the areas of the surface that represent each breast, it remains necessary to extract these areas in order that they can be analysed independently from the chest wall. As discussed in the previous chapter, the creation of a set of corresponding points across all breasts will serve to make them more easily comparable as well as allowing for the use of standard shape analysis techniques. While this could be achieved through the warping methods of Section 3.2.2 and the use of an appropriate base mesh, there are only four landmark points available for each breast and they do not provide a decent coverage of the entire breast area. For example, there are no landmarks in the upper section of each breast. This could

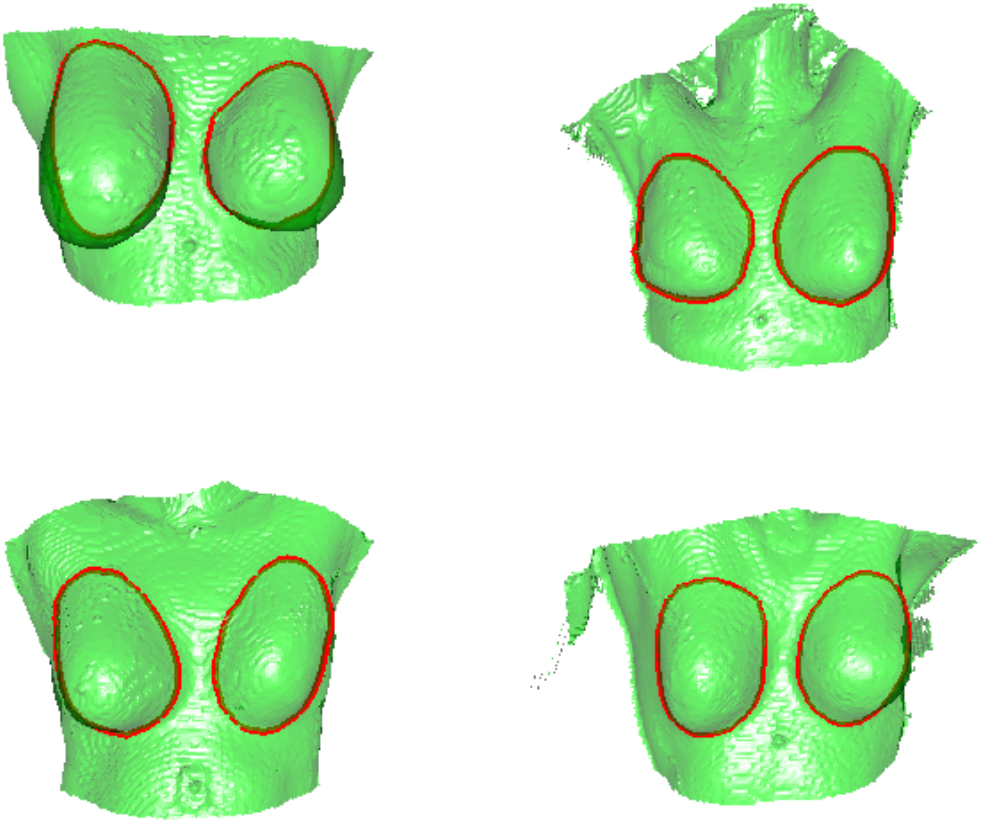

FIGURE 4.10: The final boundaries found for 4 patients from the dataset.

result in warps that are unreliable and therefore cause problems in the construction of the corresponding surfaces. Ideally a different approach should therefore be taken. A novel approach to the creation of these corresponding points was therefore devised, making use of the surface transects found for boundary identification.

By taking the first transect in a fixed direction for all surfaces (e.g. by setting $\overrightarrow{a_2} = ssn - prom$), a further t transects can be found at regularised angles $\hat{\alpha}$ from this original direction, where $\hat{\alpha} = \frac{2\pi}{t}$. These transects will then be in corresponding positions across all images and, provided the direction of rotation is reversed on one side of the chest, will also correspond between the two breasts of an individual patient.

Through the fitting of a principal curve and subsequent cubic spline, the positions of each transect can then be found at regularised intervals along the surface. As

each of the breasts vary in size and shape, the points should be taken at proportionally equal distances along the curves, thereby ensuring correspondence across the surfaces. This requires that the arc length s_i at each point of a given curve is scaled so that $\hat{s_i} = \frac{s_i}{max_i(s)}$. Denoting the principal curve of a given transect as $f(\hat{s}) = (f_1(\hat{s}), f_2(\hat{s}), f_3(\hat{s}))$, it can then be represented in terms of three cubic splines $g_1(\hat{s})$, $g_2(\hat{s})$ and $g_3(\hat{s})$.

As all transects converge at the most prominent point, it is beneficial to select points more sparsely towards this end of the strip, in order to obtain a more regular spacing across the surface. To achieve this, a sequence $\tau_i, (i = 1, \cdots, n)$ can be constructed from equally spaced quantiles of the exponential distribution and normalised to fit in the range (0, 1), thus ensuring a decreasing distance between points as the strip is traversed. It is then necessary to find the location of each transect at each

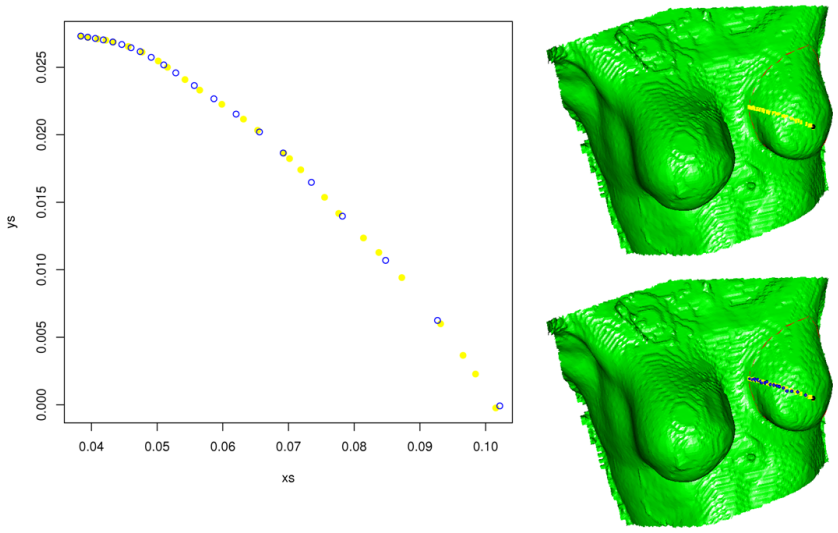

FIGURE 4.11: The original transect points (yellow) and estimated surface locations at τ (blue).

proportional distance τ_i, given by $(g_1(\tau_i), g_2(\tau_i), g_3(\tau_i))$. This process is illustrated for a particular transect in Figure 4.11, where the original transect points can be seen in yellow and the estimated spline locations at each τ_i are shown in blue.

By repeating this process across all strips, a representation is obtained of the entire breast surface, consisting of nt points at set locations. For the breast data, 20 points were found on each of 51 transects, resulting in a surface representation of

1020 points for each breast. An example of this can be seen in Figure 4.12. In

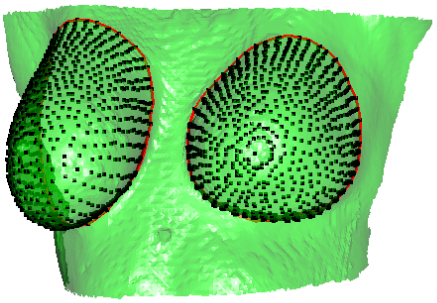

FIGURE 4.12: The set of corresponding points determined for a given patient.

comparison to the four landmark points that were previously available for each breast, this representation is clearly more informative and provides a much more detailed description of the shape. While a small number of landmarks are utilised (e.g. in the calculation of the axes and in the exclusion criteria of the boundary points), an advantage of this use of curvature to define shape is that less emphasis is placed on these human-defined landmark points as the method is driven more by the information inherent in the surface itself. For surfaces such as this, where there is a dearth of anatomically well-defined landmarks, this type of approach to surface representation could prove to be more reliable than the traditional landmark-based warping methods of Chapter 3.

Chapter 5

Asymmetry

Through the various methods discussed in the previous chapters, it was possible to create corresponding surface representations for both the cleft palate and the breast reconstruction datasets. These representations can now be analysed by means of the many landmark-based techniques. In order to address the aims for both studies, this will begin with an investigation into the assessment of the symmetry of the surfaces.

Symmetry is an aspect of shape which is important in a variety of applications, many of which are in the fields of biology and medicine. It is widely accepted that symmetry is an aesthetically desirable trait and there are therefore several kinds of corrective surgery where a symmetrical outcome is the main objective, for example the repair of a cleft lip and palate in young children. In these cases it is of interest to be able to quantify the asymmetry present in the post-surgical surface, in order to examine any abnormalities still remaining. Furthermore, even when not the direct focus of a study, the topic of symmetry can be relevant. As discussed in Bookstein (1996*a*) and Klingenberg et al. (2002), strong symmetry can lead to statistical problems when it is not taken into account. For example, when an object is perfectly symmetric, the position of certain points can be determined by that of another, which will give rise to a singular covariance matrix.

Symmetry is defined as an exact correspondence in position or form about a given

point, line, or plane [1]. The most common type of symmetry in a biological setting is bilateral symmetry, which generally refers to a similarity in the left and right hand side of an object. There are other types of naturally occurring symmetry, for example rotational symmetry, and a discussion of the analysis of these can be found in Savriama and Klingenberg (2011). For the purposes of the following chapter, however, the focus will lie purely on the analysis of bilateral symmetry.

A simple approach to the analysis of symmetry can be the comparison of either landmark distances or sizes from either side of an object of interest. For example, in a study into schizophrenia Brignell et al. (2010) used the difference in volumes between the left and right sides of the brain as an indicator of asymmetry, while both Brown et al. (1999) and Ercan et al. (2008) used the distances between selected landmarks in order to quantify the asymmetry present in the human breasts and face respectively. However, these are one dimensional measures of size rather than descriptors of shape and are thus not very informative of differences in the topography of the two sides.

More informative methods for the analysis of asymmetry have been presented by Mardia et al. (2000) and Kent and Mardia (2001), who analyse the asymmetry present in terms of the full geometry of the object. Their approach distinguishes two different types of bilateral symmetry, namely matching symmetry and object symmetry. Object symmetry occurs when the left and right halves of an individual object are symmetric i.e. when the axis of symmetry passes through the object itself. An example of this type of symmetry can be seen in the human face. Matching symmetry, on the other hand, is when a symmetric structure is present in two separate features that lie on different sides of the figure, such as the left and right hands of the human body. In this case the axis of symmetry passes midway between the two objects.

Due to the fact that a perfectly symmetric object will match exactly with its mirror image, the method is based on a comparison of the original object with its reflection in an arbitrary line or plane. The original and reflected configurations are aligned through Procrustes registration, before the distances between their corresponding points are taken as a measure of asymmetry. A formal test for significant directional

[1] Collins English Dictionary

asymmetry can then also be conducted, an extension of which was presented by Klingenberg et al. (2002).

In addition to determining the degree of asymmetry present in an object through these methods, it can be possible to extract further information on the nature of the asymmetry. Bock and Bowman (2006) built on the previously mentioned techniques in order to decompose the asymmetry of an object into its various constituent parts. They showed that, through using a comparison of the object with its reflection at each stage of the Procrustes matching procedure, it was possible to examine the proportion of asymmetry that was attributable to the size, position and orientation of the various features independently. This provides a far more informative assessment of the differences present in the two sides of an object.

This chapter illustrates the assessment of asymmetry on two distinct applications. Firstly, Section 5.1 looks at the case of object symmetry, studied in terms of the faces of young children born both with and without a cleft lip/palate. Section 5.2 then deals with the issue of matching symmetry, based on a comparison of the natural and reconstructed breasts of a group of mastectomy patients. For this data an analysis of the landmarks is first explored, before the methods are extended to examine the asymmetry over the breast surfaces. A novel approach is presented for examining the sources of asymmetry in this case.

5.1 Object Symmetry: An Application to the Cleft Palate Data.

5.1.1 Introduction

Many children are born every year with either a cleft lip or cleft lip and palate, a form of facial disfigurement that occurs during gestation. Reconstructive surgery can be used to attempt to repair the damage, with the surgical aims being two-fold - namely improved functionality and a more 'normal' appearance. While clefts can present as either uni- or bi-lateral, the dataset of this study consisted of unilateral

cases alone, all of whom had undergone a reconstructive procedure between the ages of 3 and 6 months. Data was available on 194 control children in addition to 76 patients - of whom 34 had a cleft lip only (UCL), and 42 had a cleft lip and palate (UCLP). The data had been collected longitudinally over a 5 year period, resulting in multiple images for each child.

As the abnormality appears solely on one side of the face in this cohort of patients, the aim is to match this side to the unaffected side of the face as closely as possible during the procedure. This will result in a more symmetric, and hence pleasing, appearance. Interest therefore lies in determining how effectively the surgery had brought the patient group into line with the control group in terms of facial symmetry. Following the methodology of Mardia et al. (2000), the extent of asymmetry present was evaluated for all cases. The asymmetry in this dataset has previously been examined in terms of the landmark points up to the age of 2 years as was presented in Bock and Bowman (2006), however this is now extended to the case of asymmetry over the surface for all timepoints.

5.1.2 Surface Representation

In order to assess the shape differences between the two sides of a given object, it is necessary to know which points correspond across the sides and can therefore be appropriately compared to one another. To standardise the number of points and their positions across the cases, a set of corresponding surfaces was therefore created for all images by the methods described in Chapter 3. As the symmetry of the corresponding surfaces is to be examined in this case, a symmetric base mesh was created for determining the necessary correspondences. This served to ensure that each point on the right hand side of a particular surface had a known corresponding point on the left hand side. The base mesh that was used and its (also symmetric) landmarks can be seen in Figure 5.1. Following this procedure, each image in the dataset was then represented by a set of 1140 corresponding surface points. An example of this is shown in the left hand panel of Figure 5.2, where the image has been overlaid with the representative surface points found for that case.

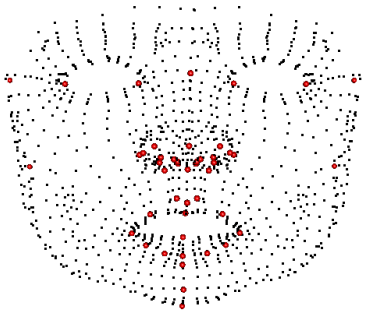

FIGURE 5.1: The symmetric base mesh and landmarks.

While the methods derived by Mardia et al. (2000) are for application to a set of landmark points, the corresponding nature of our surfaces allows for a simple extension of the technique to this data - simply by treating this set of representative points as a large landmark configuration. The standard techniques can then be applied in the usual manner.

5.1.3 Measurement of Asymmetry

As discussed by Mardia et al. (2000), any perfectly symmetric object can be matched exactly with its mirror image. The extent to which a configuration differs from its own reflection therefore provides a measure of asymmetry. To illustrate this, we will consider a configuration X, consisting of p points in m dimensions.

Naturally, any symmetric object will be composed of two distinct types of points - namely paired points and midline points. Paired points are points that have a natural partner on the opposite side of the object, while midline points lie individually along the midline/plane of the object and have no natural counterpart. In terms of the human face, the outer corners of the eye would be considered as paired points, while the tip of the nose provides an example of a midline point. When comparing a configuration of points to its reflection, it is therefore necessary to relate the paired points to their reflected counterpart, while the midline points will be related to their own reflection. It can therefore be useful to write X as $X = (X_l, X_m, X_r)$, where

X_l, X_r represent the paired points on the left and right sides respectively, and X_m represents the set of midline points.

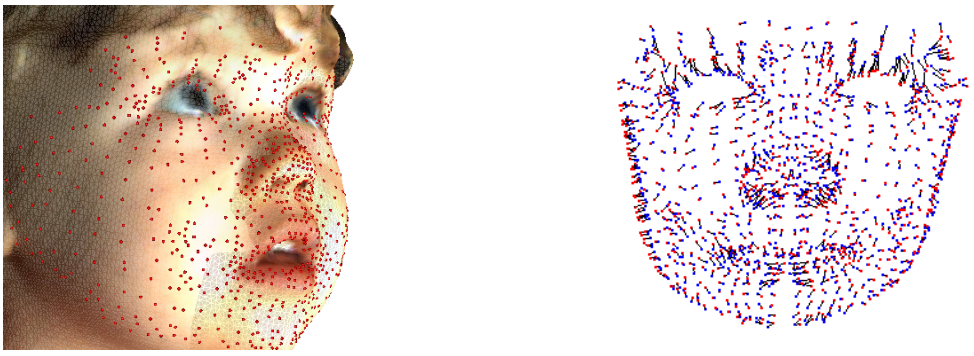

FIGURE 5.2: *Left:* An image from the dataset with the representative points found for that surface displayed in red. *Right:* The same configuration of points (again in red) and its matched reflection (shown in blue) with a line joining the various sets of paired points.

As the location of a surface in space is not of interest here, the ease of calculation can be increased by first centering X on 0, i.e. by subtracting the centroid $\bar{X}$ from each point X_i, where $\bar{X} = \sum_{i=1}^{p} \frac{X_i}{p}$ and $X_i = (x_i, y_i, z_i)$ is the i-th row of X. Similarly, as the size of the configuration is not of interest, X can be scaled to be of unit size. This is achieved through the division of X by its centroid size $\|X\|$, where $\|X\| = \sqrt{tr(X^T X)}$. This removal of size ensures that the asymmetry values are purely related to shape, and are therefore comparable across the different surfaces.

A reflection X^R of X can then be obtained by reflection in an arbitrary line (when $m = 2$) or plane (when $m = 3$), achieved for example by reversing the sign of the coordinates in the first dimension. When the reflection is carried out in this way, X^R will also be centred on 0, ensuring that the locations of the two configurations are matched.

It is then necessary to align X and X^R. As the configurations are already of the same scale, this can be achieved through the ordinary partial Procustes algorithm, the details of which were discussed in Chapter 3. However, the labels of the paired points must first be reversed for X^R in order to ensure that each point of X is being matched to its corresponding point in X^R. The appropriate reordering of the points is given by $X^R = (X_r^R, X_m^R, X_l^R)$. As the two configurations are already centred on

0 and of the same size, orientation is the only similarity transformation that need be considered during the Procrustes registration. It is therefore simply required to identify the $m \times m$ rotation matrix Γ that minimises $\|X - X^R\Gamma\|^2$, thereby optimising the match between X and X^R.

Upon this alignment of the two configurations, the asymmetry can then be quantified through the distances between the original configuration and its optimally matched reflection. Specifically this is done by examining the point-to-point displacement between the two configurations, leading to an asymmetry score given by

$$A = \frac{\|X - X^R\Gamma\|^2}{p}.$$

This score is standardised by the number of points on the surface, leading to an intuitive interpretation of each score as the average squared distance per point.

An example of the matching process can be seen in Figure 5.2 for one of the control cases. The original set of representative points can be seen in red in both images of this figure, while the right hand image also shows its matched reflection in blue. The distance between each set of paired points is illustrated by a line connecting the two points. As would be expected for a control child, the differences between the original configuration and its reflection appear to be relatively small.

5.1.4 Results

This reflection and matching process was applied to both the cleft and control data, resulting in an asymmetry score for each child at a series of timepoints. For each of the three subject groups, i.e. UCL, UCLP and control, the calculated average asymmetry score at each timepoint can be seen in Figure 5.3. As the scores arise from a sum of squares, a square root transformation has been applied to reduce skewness and transform the values back to their original scale.

The results seen here are as would naturally be expected. At the 3 month timepoint, prior to the reconstructive surgery, there is a marked difference in the asymmetry scores of the two cleft groups in comparison to the control group, with the higher

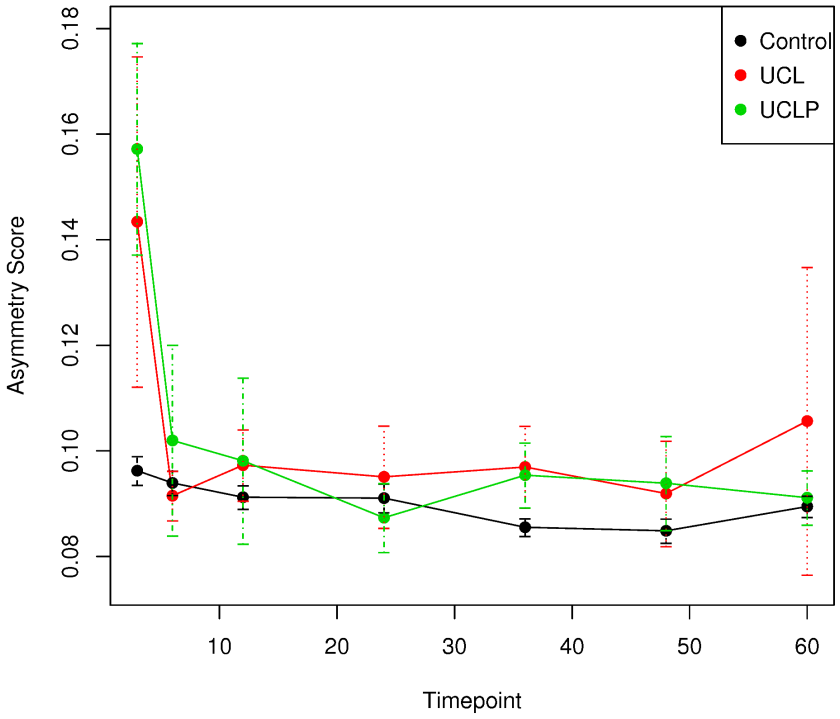

FIGURE 5.3: The average asymmetry score at each timepoint for each of the three subject groups. Confidence intervals are shown for the estimates at each point.

scores for the UCLP group reflecting the more severe nature of this cleft type. By the 6 months timepoint however, when all surgical procedures have been conducted, the average scores in both patient groups have decreased considerably. While generally still slightly higher than the average for the control group from this timepoint onwards, the differences are much less pronounced. There are a couple of timepoints where it appears that the asymmetry is actually lower in one or other of the patient groups (i.e. the UCL group at 6 months and UCLP group at 24 months), however this is likely to be due to variability associated with the relatively small sample size. As can be seen in Table 5.1, the number of images available is much smaller in the two patient groups than in the control group across all timepoints. This leads to less stable estimates of the mean, as is seen by the wider range of the confidence intervals for these groups.

At an individual level, the highest asymmetry score was unsurprisingly found for a 3 month old child from the UCLP group, who is shown in the left-hand image of Figure 5.4. The child found to have the lowest score is also shown in the right-hand

	Timepoint (months)						
Group	3	6	12	24	36	58	60
Control	90	105	111	104	86	41	108
UCL	13	11	13	5	10	7	3
UCLP	9	6	6	11	28	13	10

TABLE 5.1: The number of images available at each time point in the three separate subject groups.

image of this same figure.

FIGURE 5.4: The images found to have the highest and lowest asymmetry scores respectively.

5.2 Matching Symmetry: An Application to the Breast Reconstruction Data

In contrast to the object symmetry displayed by the face, the case of matching symmetry is now considered in terms of the breast reconstruction data. This dataset consists of 43 patients, all of whom had undergone a unilateral mastectomy and reconstruction procedure, specifically a reconstruction by *Latissimus dorsi* flap. The aim of such a procedure is to rebuild the breast in such a way that it is as similar to the existing breast as possible, thereby maximising breast symmetry. To date, the vast majority of evaluations of these reconstructive breast surgeries have been

subjective, and are therefore open to issues with inter-observer agreement and reproducibility, as discussed in Cardoso et al. (2006) and Cardoso et al. (2007). An objective method of assessment would therefore prove to be very useful.

While it is widely accepted that there is a natural asymmetry present in the wider population, as discussed in Brown et al. (1999) amongst others, interest lies in whether there is a more pronounced asymmetry within this patient group than would normally be seen. Unfortunately, no control data is currently available for this study, however it was hoped that reproducible methods could be derived for the analysis of asymmetry that would prove useful in any future comparison.

As discussed earlier in the chapter, matching symmetry applies when two physically separated structures are present on a figure in the form of mirror image copies, say X and Y. Generally speaking this is a simpler situation than that of object symmetry, as the reordering of paired points is not required, nor are there any midline points to be considered. However, there are some different issues to be considered with this type of symmetry and these will be discussed throughout this section.

Given an appropriate set of landmarks or representative surface points, all locations on X will correspond to a particular location on Y. The matched configurations X and Y^R can therefore be used to assess the asymmetry present, where Y^R again denotes the reflection of Y in an arbitrary line or plane. The previously discussed case of object symmetry involved the matching of a given object with its own reflection, meaning that the original and reflected structures were of the same size and thus scale was not an issue. With matching symmetry, however, the issue of scaling is immediately raised. In this case, X and Y are separate structures, and it must be decided whether any size differences in the two objects are to be taken into account. When purely the shape differences are of interest, each configuration can be scaled to be of unit size through a division by its centroid size, i.e. $\frac{X}{\|X\|}$, $\frac{Y}{\|Y\|}$, thereby removing any size differences present. However, in many applications it would be informative to retain these differences in scale of the configurations, in order to allow for a more complete description of the differences between the two structures.

In the following application to the breast reconstruction data, any size differences between the two sides are of great importance, as the aim is to match the two breasts not only in shape but in size also. The extent of this size effect was therefore investigated prior to the analysis of asymmetry.

5.2.1 Comparison of Breast Size

5.2.1.1 Centroid Size

As a first step in examining the disparity between the two breasts, the difference in size between the natural and reconstructed breast was investigated through a comparison of their respective centroid sizes, i.e. $\|X\|$ and $\|Y\|$. For each patient, these were calculated using the four landmarks that had been placed on each breast - namely the most prominent, inferior, medial and lateral points of the breast. An example of the landmarks can be seen in Figure 2.7.

The resulting values were plotted and can be seen in the left-hand panel of Figure 5.5. The line shown is the line of equality, and therefore deviations from this line indicate the extent of the difference between the two sides. As many of the points are clustered around the line, this suggests that the size differences present are small. This is also backed up by the mean values, with the average centroid size for the reconstructed breast being only slightly smaller than the average size of the unreconstructed breast, at 0.113m and 0.115m respectively. These differences were then assessed more formally by means of a paired-sample t-test, where the null hypothesis was that the mean difference between the two sets of sizes is equal to zero. The resulting p-value of 0.51 indicates that the differences in centroid size between the two breasts are not significant when assessed through the centroid size of the landmarks.

On inspection of the graph, however, there appears to be a trend in the data. It seems that in cases with a larger natural breast size there was a greater difference between the two sides, with the reconstructed breast generally being considerably smaller. To investigate this, the difference in centroid sizes was calculated for each

case and compared to the size of the natural breast. The resulting plot can be seen in the right panel of Figure 5.5 and appears to confirm this, with a clear positive relationship between the two variables. Upon fitting a regression, the slope was found to be 0.54 with 95% confidence interval (0.41, 0.67), thus confirming that, in terms of centroid size, as the natural breast size increases the post-surgical discrepancy between the two breasts tends to be greater.

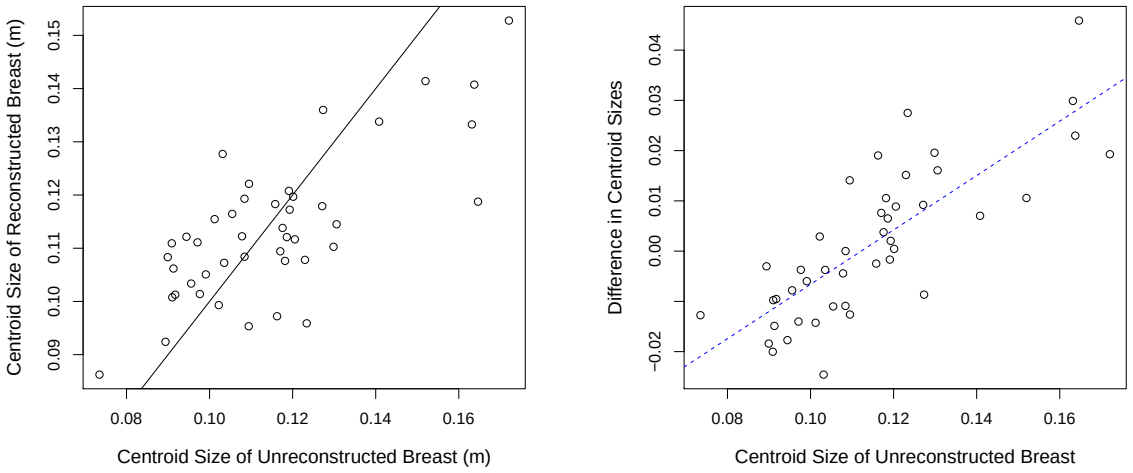

FIGURE 5.5: *Left:* Centroid sizes of the reconstructed and natural breast along with the line of equality. *Right:* Centroid size of the natural breast against the difference in centroid sizes. The fitted regression line is shown in blue.

5.2.1.2 Breast Volume

From the three-dimensional image obtained, it is also possible to calculate the volume of the breast. This was done using the Breast Analysis Tool (BAT), software developed at the University of Glasgow and described in Oehler (2006). In order to calculate breast volume in this software, four landmarks must be placed at the extremities of the breast segment, from which a rectangular surface patch is created. The software then creates an artificial chest wall, through which the volume of the enclosed area can be calculated. An illustration of the procedure can be seen in Figure 5.6. The method had been previously validated against the water displacement method, which was taken to be the gold standard for volume determination. Details of the validation study are given in Chapter 2.

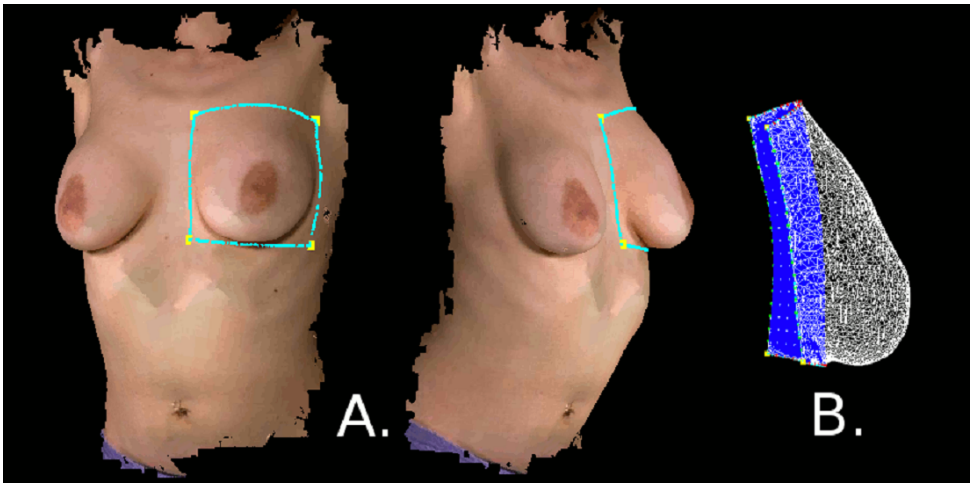

FIGURE 5.6: Calculation of volume in BAT software for a 3-d breast model.

For each patient, the volume of both the natural and reconstructed breast was calculated four times and the average of these was values taken as the final volume. The within-person standard deviation for volume measurement was 21 cc. The values for the reconstructed and natural breast can be seen in Fig. 5.7, along with the line of equality. It seems from the plot that there is a marked difference in volumes of the two breasts, with the natural breast tending to have a considerably higher volume in most cases. These differences were again formally tested through a paired-sample t-test, and in this case the differences were found to be highly significant at $p < 0.0001$, with a mean difference of 176.8cc (95% CI = 103.5, 250.0).

Through this volume assessment, it was again found that there was a tendency for the differences to be greater in cases where the natural breast size was larger, as seen in the right hand panel of Figure 5.7. The slope of the regression line was slightly greater this time at 0.65 (0.54, 0.75). As this trend was seen by both methods of comparison, this suggests that it is perhaps more difficult to create a reconstructed breast similar to the existing breast in the case where the natural breast size is larger.

Since these volume measurements are calculated from thousands of surface points, as opposed to the four used for centroid size calculation, it is likely to be a more accurate measure of the size of the breasts. It would therefore appear that there is a

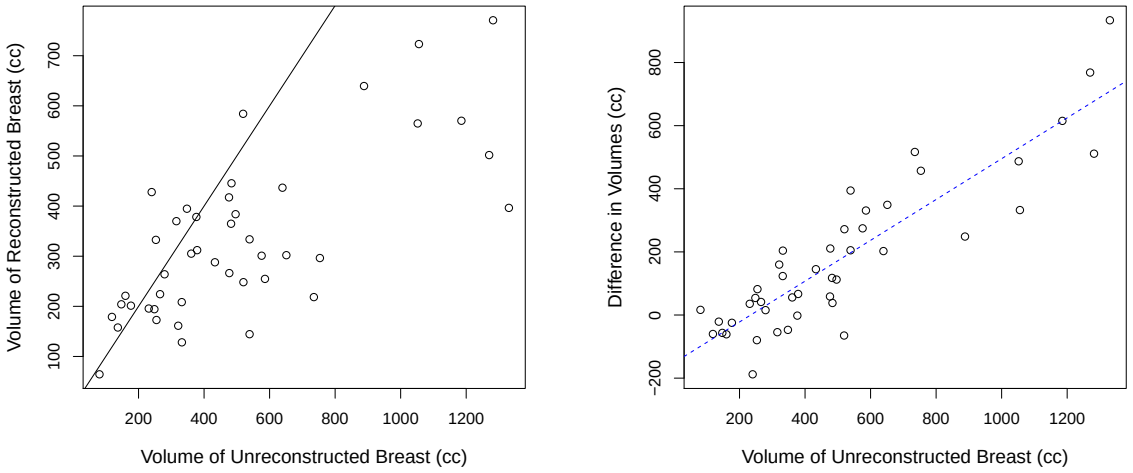

FIGURE 5.7: *Left:* Calculated volumes of the reconstructed and natural breast showing the line of equality. *Right:* Volume of the unreconstructed breast against the difference in volumes, along with the fitted regression line.

significant difference in the size of the reconstructed and natural breasts, something which should be taken into account when conducting an analysis of the asymmetry.

5.2.2 Landmark Asymmetry

The asymmetry of the breasts was primarily assessed for the set of landmark points only. Taking X to represent the four landmarks on the natural breast and Y the corresponding landmarks on the reconstructed breast, the two configurations can both be scaled by $\|X\|$, the centroid size of the natural breast. This scaling serves to retain the relative size differences between the two configurations, while standardising the scale across all cases. The configurations of the natural breasts are then all of unit size, while the configurations of the reconstructed breast are scaled relative to this.

Taking the standard approach to matching symmetry, the asymmetry scores would then be evaluated through the distances between X and the Procrustes matched reflection of Y. As the locations of the configurations are different in this case, the matching process would involve the estimation of the translation parameter γ in addition to that of the rotation matrix Γ, leading to a score given by

$$A = \frac{\|X - (\Gamma Y^R - \mathbf{1}_p \gamma^T)\|^2}{p}$$

where $\mathbf{1}_p$ a vector of 1's of length p.

By this approach the breasts are treated as separate configurations and fully registered prior to the calculation of the asymmetry score, meaning that the scores solely reflect any differences that exist in the shape or size of the two breasts. However, it may be that there are also differences in the position or orientation of the breast that are not be taken into account by this method, but that would likely be influencing the overall symmetry of the patient's appearance. In order to deal with this, this standard approach was extended to allow for an inclusion of these factors.

In place of using an arbitrary plane, the midline plane of the chest can instead be used as the plane of reflection, allowing the relative position of the breast on the other side of the chest to be found. This reflection then allows for a more informed analysis of the asymmetry to be performed, as the positional and rotational differences in the breasts can be examined. Taking the landmarks *ssn* and *xipho* as the midline of the chest, an approximation of the midline plane $\boldsymbol{\pi}$ would pass through these points in the normal direction to the surface, and should roughly divide the surface into the left and right hand side. By reflecting in this plane, the configuration Y would then be shifted to the same side as X, while its position relative to the midline is maintained.

However, the plane $\boldsymbol{\pi}$ is clearly very dependent on the normal direction, say $\mathbf{z}$, that is found for the surface. If $\mathbf{z}$ is not an acceptable approximation of the normal direction of the chest wall, this will result in an inappropriate estimate of $\boldsymbol{\pi}$ and the reflected Y^R will lie too far either above or below the chest wall. This normal direction would generally be found through a principal components analysis of a localised set of points on the chest wall. However, as the chest is a curved surface $\mathbf{z}$ was found to vary greatly depending on the points that were chosen. In order to deal with this, a more robust method for the determination of Y^R was devised.

In place of a direct reflection onto the opposite side of the chest, the points were first reflected in the anterior-posterior direction. This was achieved through reflection in

the plane containing the points *ssn*, *xipho* and $\bar{Y}$, the centroid of the landmark configuration of the reconstructed breast. Taking the midline vector to be $\mathbf{v}_m = ssn - xipho$, this reflected configuration of Y can then be rotated around the midline by using $\mathbf{v}_m$ as an axis of rotation. The matrix required for a rotation of this type was previously seen in Equation 4.2. In this case, the x, y and z of this equation would represent the respective components of $\mathbf{v}_m$. The angle of rotation can again be given by α, and a rotation by any α will ensure that the distance of each of the points from the midline is maintained. Selecting the α that minimises the distances between X and the reflected Y, i.e. $\hat{\alpha}$, it will not only be ensured that the points are reflected onto the opposite side of the chest, but that they are also lying flush with the chest wall. This method therefore leads to a more confident estimate of the relative location of the reconstructed breast on the opposite side of the chest. The configuration corresponding to this reflected and optimally rotated Y will now be denoted by Y^R.

The global asymmetry score arising from the two configurations X and Y^R, is then given by

$$A_G = \frac{\|X - \hat{Y}^R\|^2}{p}.$$

As this score includes all size, orientation and position differences, it is therefore a more accurate description of the differences between the two breasts.

The distribution of these scores can be seen in Figure 5.8, again shown on the square root scale in order to reduce skewness. The scores range from 0.007 to 0.062, with a mean score of 0.021. As can be seen from the histogram, there are two cases with markedly higher asymmetry scores than the rest of the group. These patients can be seen in Figure 5.9, and unsurprisingly demonstrate a pronounced difference between the two breasts.

Comparison to Subjective Scores

In order to validate the methodology, subjective evaluations of symmetry were obtained from 6 expert reviewers - consultants in plastic surgery who were routinely

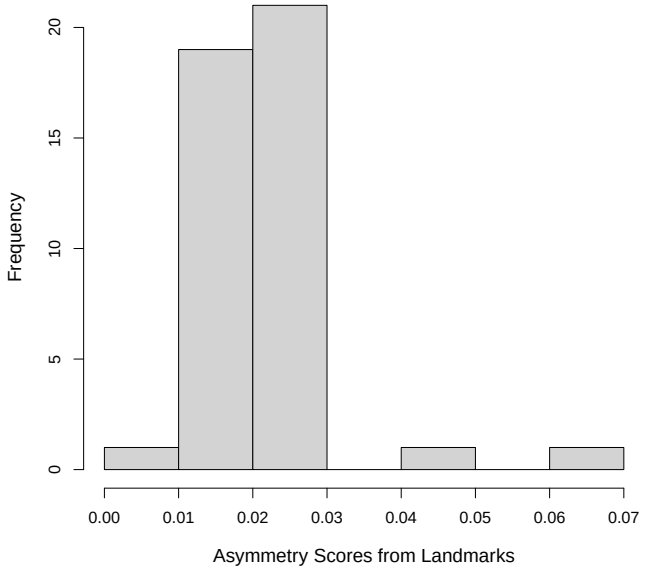

FIGURE 5.8: Distribution of asymmetry scores when calculated from the landmark points.

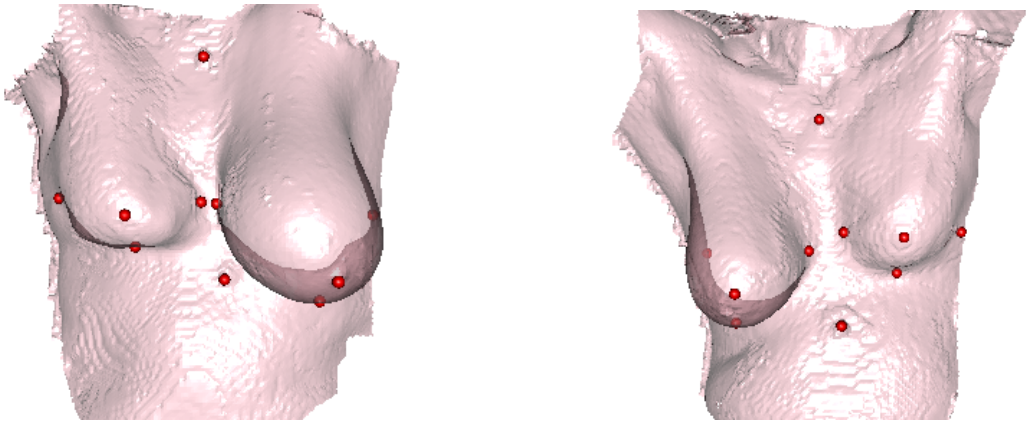

FIGURE 5.9: The two patients found to give the highest asymmetry scores.

undertaking breast reconstructive procedures. The subjective evaluation was carried out on 2-d photographs of the patients that were scored according to the Harris scale, a four point grading system where 1 represents a poor result and 4 an excellent result. Symmetry was judged for the appearance of the reconstructed breast in relation to the opposite side. For each patient, the average of the 6 evaluations was taken as the subjective score.

The relationship between these subjective scores and the calculated asymmetry scores can be seen in Figure 5.10. As would be expected the relationship is a negative

one, as low values in the subjective score represent a high degree of asymmetry. As the subjective scores arise from the average of a set of ordinal variables and are therefore not strictly continuous, the nonparametric Spearman correlation coefficient was used to evaluate the strength of the relationship. The correlation coefficient for the two sets of scores was found to be $\rho = -0.54$. This correlation is not as high as would perhaps be hoped, however correlation is sensitive to extreme values and therefore the two cases seen to have markedly higher asymmetry scores may be affecting this. As can be seen in the plot, the majority of the other scores lie reasonably close to the line of best fit.

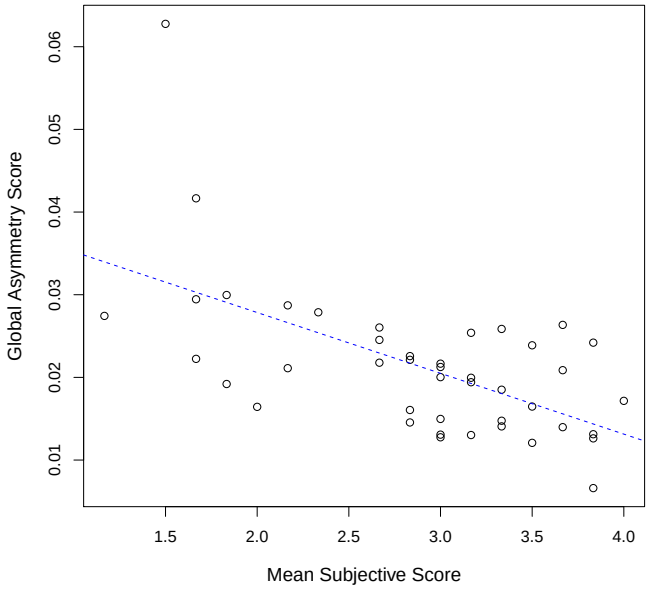

FIGURE 5.10: The average subjective scores plotted against the square root of the calculated asymmetry scores. The line of best fit is also shown.

5.2.3 Decomposition of Asymmetry Scores

An advantage of obtaining the reflection in the previously discussed manner is that, in addition to more accurately capturing the overall asymmetry present in the chest, it allows for the components that are contributing to this asymmetry to be investigated. As discussed in Bock and Bowman (2006), in the case of object symmetry a given feature can contribute to the overall asymmetry of an object in a variety of

ways. This study demonstrated how, when holding the remainder of an object fixed, rotating or translating a particular feature will alter the overall asymmetry score. This therefore allows the effect of that feature's position and/or orientation on the overall asymmetry of the object to be investigated.

Bock and Bowman (2006) illustrated their technique on the cleft lip data, where the feature investigated, e.g. the nose, was represented by a subset of the original configuration X. The position, scale and orientation of the feature relative to the rest of the surface was therefore known. Although we are here interested in examining two distinct configurations, their relative positions and orientations have been determined through the method of reflection and rotation that has been implemented. They can therefore be used to evaluate the contributions to asymmetry of the orientation and location of the reconstructed breast. In this setting, the asymmetry arising from the size differences of the breasts can also be assessed. To evaluate the contributions of these various effects, the matching procedure can be implemented in a staggered fashion, with an analysis of the mismatch performed at each stage.

Firstly, in order to investigate the difference in position between the two breasts, Y^R can be shifted to be in the same position as X. The translation required to achieve this is described by $\boldsymbol{\delta} = (\delta_x, \delta_y, \delta_z) = \bar{X} - \bar{Y}^R$, where $\bar{X}$, $\bar{Y}^R$ are the centroids of the unreconstructed and reconstructed breast landmarks respectively. The shifted configuration $\hat{Y}^R$ is then found by taking the first m columns of $[T_{\boldsymbol{\delta}}] \left[Y^R | \ \mathbf{1}_p\right]^T$, where T is the translation matrix

$$T_\delta = \begin{bmatrix} 1 & 0 & 0 & \delta_x \\ 0 & 1 & 0 & \delta_y \\ 0 & 0 & 1 & \delta_z \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By matching the locations of the two configurations in this way, the overall distance between the sets of corresponding points will clearly be decreased. The magnitude of this decrease can therefore be taken as the extent of asymmetry that is due to the position of the reconstructed breast. The contribution to asymmetry from the

location of the breast can therefore be denoted by

$$A_L = A_G - \frac{\|X - \hat{Y}^R\|^2}{p}.$$

Subsequently, the contribution to asymmetry from the orientation of the reconstructed breast can be determined. This is similarly achieved through optimally matching X and $\hat{Y}^R$ in terms of orientation, before again evaluating the improvement seen in the match of the two configurations. The optimal rotation matrix $\hat{\Gamma}$ can again be found through the ordinary partial Procrustes algorithm, and leads to a contribution of the orientation of the reconstructed breast to asymmetry that is given by

$$A_O = A_L - \frac{\|X - \hat{\Gamma}\hat{Y}^R\|^2}{p}.$$

Finally, the effect of the size difference between the two breasts can be assessed. Both configurations had previously been scaled by $\|X\|$ in order to standardise distances across all cases. However, by now scaling $\hat{Y}^R$ to be of unit size, i.e. to be of the same size as X, the extent of asymmetry occurring from the differences in the scale of the two configurations can also be evaluated. The magnitude of this size effect is given by

$$A_S = A_O - \frac{\|X - \hat{\Gamma}\frac{\hat{Y}^R}{\|\hat{Y}^R\|}\|^2}{p}.$$

Upon this removal of all location, orientation and size effects, it follows that any remaining asymmetry must be due to an intrinsic difference in the actual shape of the two breasts, and this intrinsic shape difference is therefore quantified by

$$A_I = \frac{\|X - \hat{\Gamma}\frac{\hat{Y}^R}{\|\hat{Y}^R\|}\|^2}{p}.$$

This method therefore leads to a partitioning of the overall asymmetry score into four different components, namely the asymmetry which is due to location, orientation,

scale and shape difference, i.e.

$$A_G = A_L + A_O + A_S + A_I.$$

5.2.3.1 Results

The individual contributions to asymmetry were calculated for each of these components. By dividing each of the components A_L, A_O, A_S and A_I by the overall global score A_G, their effect can be studied as a proportion of the overall asymmetry for any given case. These proportional contributions are shown in Figure 5.11, both in terms of the individual patient profiles and the overall trends. The average contribution was found to be highest for the intrinsic asymmetry A_I at 57.7%, indicating that the actual differences in the shapes of the breasts were by far the most important factor in the asymmetry. This was followed by location, with the average value for A_L being 20.4%. Orientation and size were found to account for even less still, with the average values being $A_O = 12.9\%$ and $A_S = 9.1\%$ respectively. This low value for the size effect is perhaps surprising, particularly as there was previously found to be a significant difference in the size of the two breasts. However, it is thought that this may be due to the shape changes that also tend to occur as the size of the breast increases. As the breast is attached at the chest wall, it is logical that the prominence of the surface, i.e. the position of the most prominent landmark, would change more than that of the other three landmarks with any significant size increase. A simple scaling of all points would therefore not describe this change well, leading to much of this size changed being included in the intrinsic asymmetry component. This was confirmed by comparison to a decomposition that allowed for the landmark *prom* to be scaled independently of the three remaining landmarks, as scaling in this manner led to an increase in the average value of A_S to 20%. With a few exceptions, the profiles of most individuals showed a similar pattern over the four contributing factors, as can be seen in the right-hand panel of Figure 5.11.

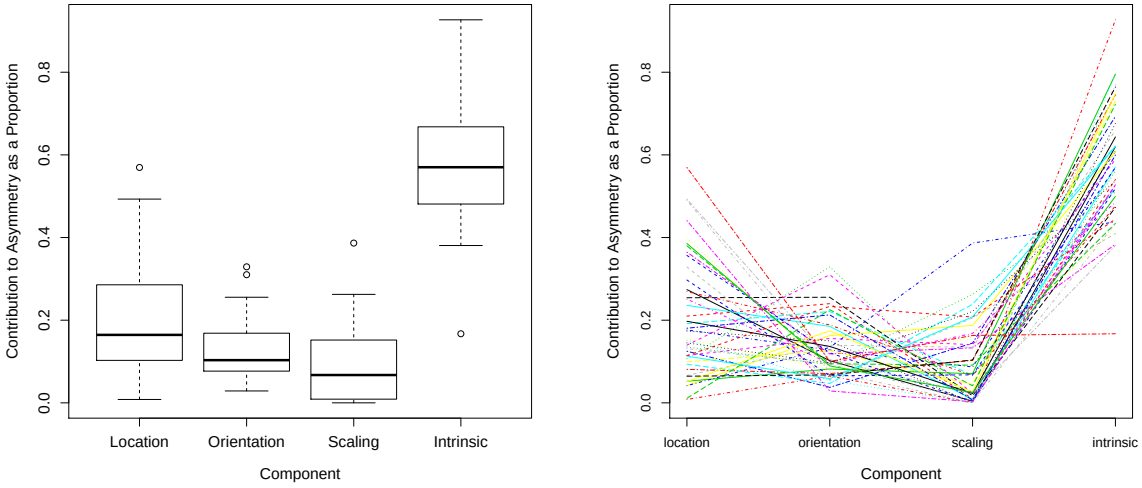

FIGURE 5.11: *Left:* Boxplot of the proportional contributions to asymmetry from the location, orientation, size and shape of the reconstructed breast. *Right:* The same proportional contributions on an individual level, with each line representing a different patient.

5.2.4 Surface Asymmetry

As was discussed in Chapter 4, through an examination of the surface curvature, it was possible to determine where the boundary of the breast lies and therefore derive a set of corresponding surfaces for the patients. Treating each of these surfaces as an extended landmark configuration, the previous methods can then also be applied to assess the asymmetry present across the surfaces. Although the number of points p has now increased from 4 to 1021 for each breast, the surface representations were found to give very similar results to those from the landmark analysis, as can be seen in Figures 5.12 and 5.13. For this data, the scores ranged from 0.010 to 0.093, with an average value of 0.024. The correlation of the scores with the subjective ratings was $\rho = -0.505$, similar to the strength of relationship between the landmarks and subjective score. As seen in the right hand plot of Figure 5.12, there were again two outliers that could be affecting the strength of this correlation. The breakdown of asymmetry also gave similar contributions for the surface data, with mean values $A_I = 55.7\%$, $A_L = 23.1\%$, $A_O = 13.1\%$ and $A_S = 8.1\%$.

The relationship between the scores calculated from the landmarks and those from the surfaces can be seen in Figure 5.14 . The line shown is the line of equality. There

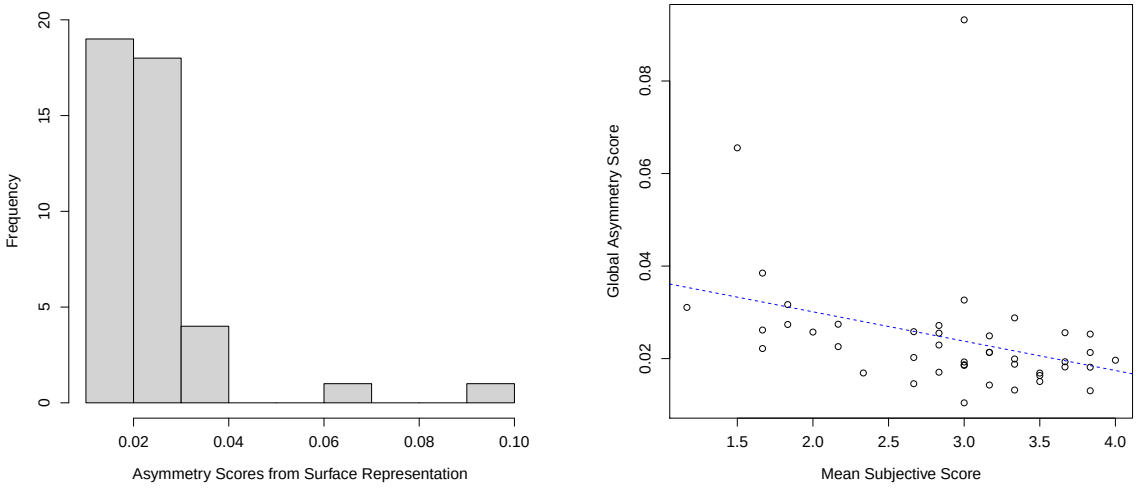

FIGURE 5.12: *Left:* Distribution of asymmetry scores when calculated from the surface representations. *Right:* The average subjective scores plotted against the square root of the asymmetry scores, as calculated from the surface representations.

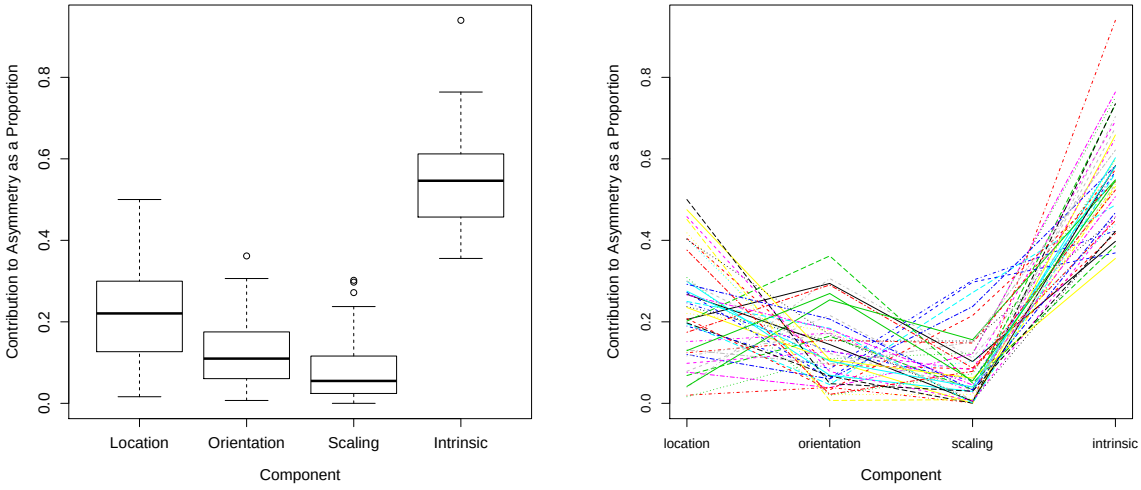

FIGURE 5.13: *Left:* Boxplot of the proportional contributions to asymmetry from the location, orientation, size and shape of the reconstructed breast, in terms of the surface representations. *Right:* The same proportional contributions on an individual level, with each line representing a different patient.

appears to be a strong correlation between two sets of scores, with the majority of points lying close to the line. However, it can be seen that there is one clear outlier. This was a case for which the surface score was extremely high while the value found from the landmarks was far smaller. The correlation coefficient between the two sets of scores was found to be $\rho = 0.50$, however this increased to $\rho = 0.77$ upon exclusion

of this outlier. The patient corresponding to this point is shown in the right hand panel of Figure 5.14, along with the representative surface points that were found for the two breasts. From the image it is clear to see that, when determining the breast boundary through the surface curvature, the upper boundaries of the two breasts were found to lie in very different positions. There are no landmarks placed on this upper region of the breast, and therefore these differences were not flagged through the landmark method of asymmetry. This highlights an advantage of using the surface representation as opposed to simply a set of landmark points, since any areas of the surfaces that are not adequately described by the landmarks will be automatically included in the surface analysis.

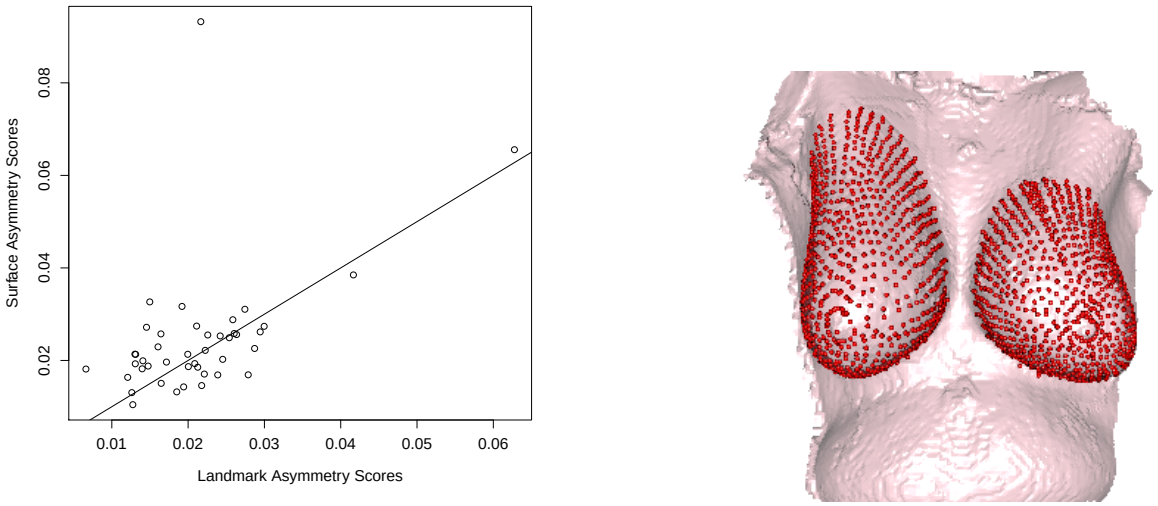

FIGURE 5.14: *Left:* The landmark scores plotted against the surface scores, along with the line of equality. *Right:* The patient corresponding to the outlier in the plot, along with representative surface points used for asymmetry calculation.

A further advantage of the use of surface representations is that they allow the results to be displayed in a more informative way than with the landmark points alone. For example, the differences between the mean reconstructed and natural breasts can be examined, as is seen in Figure 5.15. The size differences that were found can be seen here, with the average of the natural breasts being slightly larger than that of the reconstructed breasts. Clearly looking at the average configurations of the four landmark points alone would give a far less intuitive depiction of these breast shapes.

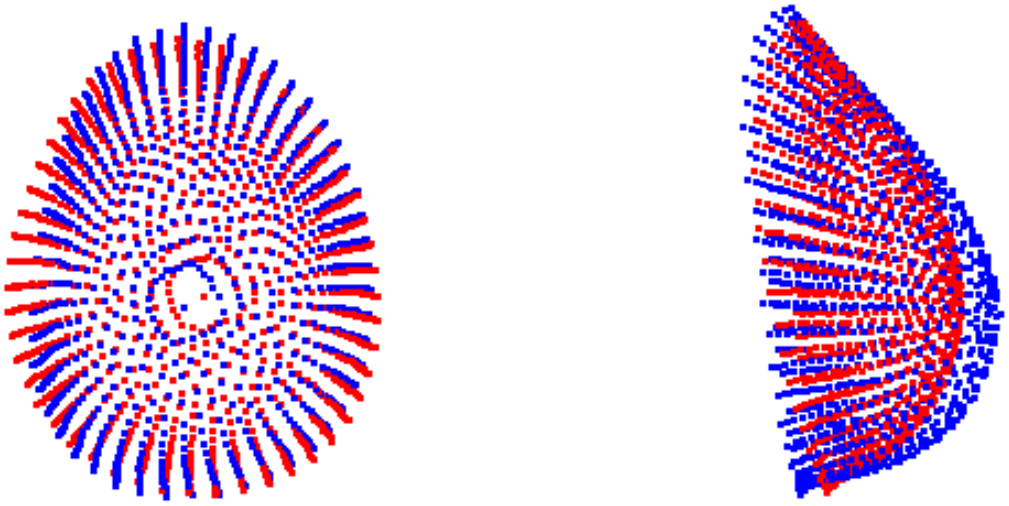

FIGURE 5.15: The average natural (blue) and reconstructed (red) breast, from both the frontal and profile views.

The asymmetry of these surfaces can be further investigated by means of a principal components analysis, a technique that allows the variability of a set of surfaces to be explored and is discussed in greater detail in Chapter 6. The analysis involves the eigendecomposition of the covariance matrix of the tangent coordinates, where the resulting eigenvectors describe the various principal components. Each of the resulting principal components (PC's) then represents a different mode of variation in the data, with the PC scores of each surface representing how related the surface is to that particular mode of variation. A difference between the PC scores of the natural and reconstructed breasts for any particular PC would therefore indicate a difference in the shape of the two breasts.

The standardised scores for each of the first 8 PC's can be seen in Figure 5.16, for both the natural and reconstructed breasts independently. These 8 components accounted for almost 90% of the variation in the dataset, with the first three explaining 46%, 13% and 9% of the variability respectively. As can be seen, the scores were found to be similar across many of the PC's. This would suggest that, while there may be differences between the natural and reconstructed breast for a given patient, there may not be an overall structure or pattern to these differences. For a small number of the PC's however, for example PC1, PC3 and PC6, there does appear to be a slight degree of separation between the two set of scores. For the

components that explain only a very small proportion of the variability, this may not be of any great consequence. However, as PC1 is accounting for almost half of all of the variation, this suggestion of a difference between the two groups for this PC should be explored.

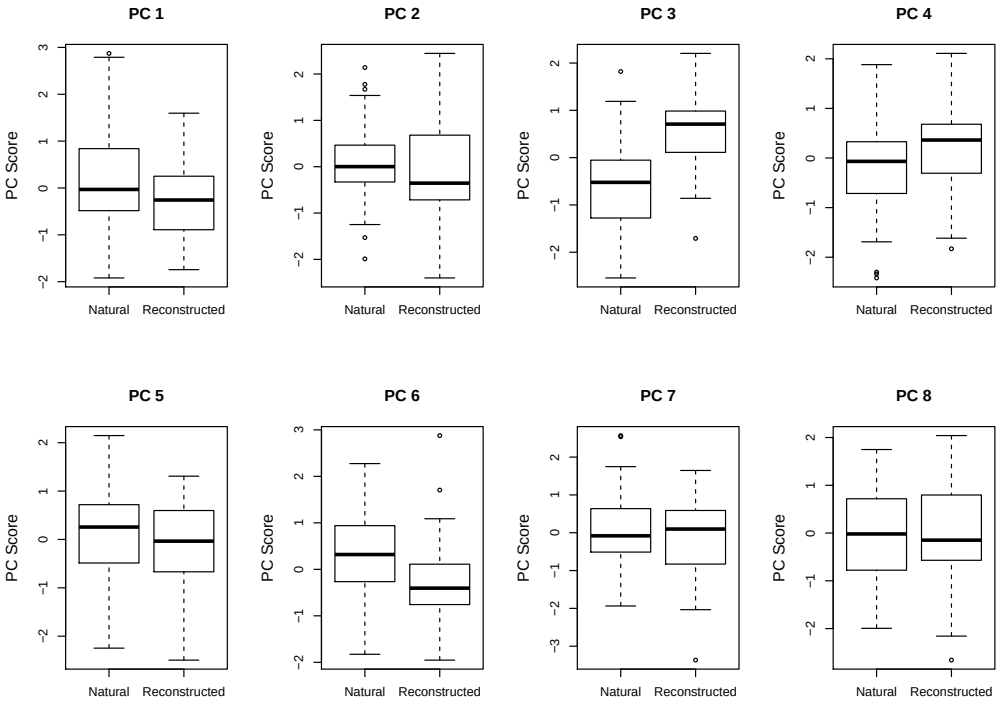

FIGURE 5.16: The standardised PC scores for the natural and reconstructed breasts separately, shown for each of the first 8 components.

As any differences in scale were retained in the data, the first principal component was unsurprisingly found to be strongly related to the centroid size of the configurations, as can be seen in the left panel of Figure 5.17. The correlation between breast size and the scores for PC1 was found to be 0.98 (0.96, 0.98). The values for the natural breast are shown in black, with those for the reconstructed breast shown in red. As was seen in the boxplot of Figure 5.16, the scores for PC1 tended to be slightly higher for the natural breast, confirming the trend seen earlier for the natural breast to be slightly larger. This is perhaps more clearly seen in the right hand panel of Figure 5.17 however.

The effect of this first component is displayed in the top row of Figure 5.18, where the mean surface is displayed in green along with ± 2 standard deviations of the

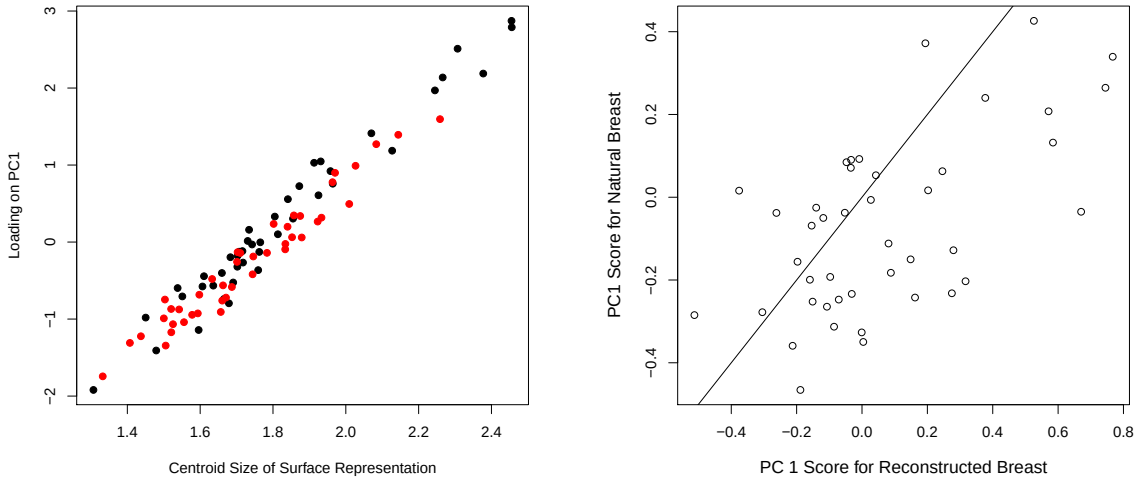

FIGURE 5.17: *Left:* The strong relationship between centroid size and the score for the first principal component. The scores for the natural and reconstructed breasts are shown in black and red respectively. *Right:* The scores for the first component for the natural and reconstructed breasts.

corresponding eigenvector. It is clear from the image that there is a fundamental change in the form of the breast as the size is increasing, with the surface changing from a reasonably flat to a more convex shape with increasing values of PC1, and as such a simple scaling of the points would therefore not capture this growth pattern well. This gives weight to the earlier hypothesis of the shape changes that occur with growth and helps to explain the low importance given to scaling in the decomposition of the asymmetry scores. The effect of the second component is shown in the bottom row of this figure, and appears to be mainly related to rotational effects.

5.3 Concluding Remarks on Asymmetry

Through the creation of sets of corresponding surfaces, the asymmetry of both the cleft palate and breast reconstruction patients was successfully assessed in terms of the surface data. As with landmark data, standard approaches to asymmetry can be applied to these surfaces in order to obtain asymmetry scores for each case. However, it was seen that the use of these surfaces also allows for a more informative visualisation of the results. For example, the application of a principal components

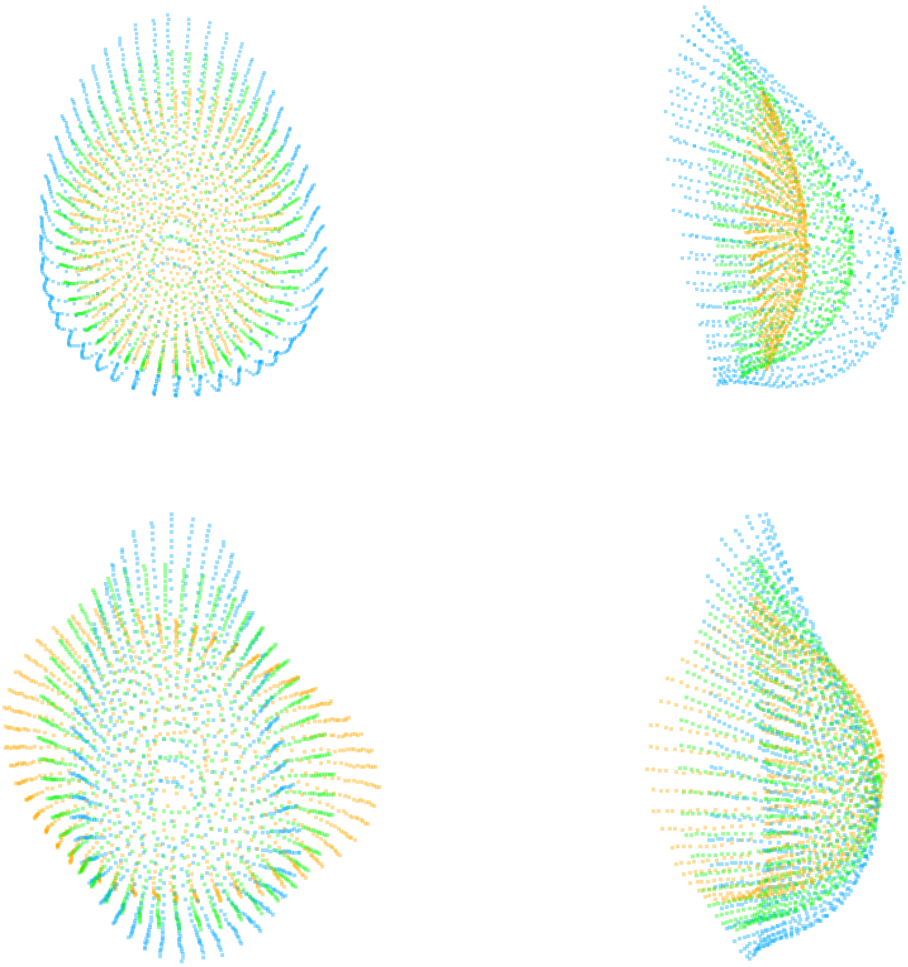

FIGURE 5.18: The top and bottom rows display the variability described by the first and second components respectively, from both the frontal and profile view. The mean breast is shown in green, with ±2 standard deviations of the corresponding eigenvector shown in blue and orange respectively.

analysis to the data can allow for the investigation of subtle shape differences between the two sides of an object.

Through visual examination and comparison to subjective scores, it was concluded that an appropriate method for determining the severity of asymmetry in the breast reconstruction patients had been devised. While it should logically be the case that the evaluation of asymmetry over the surface is more accurate, as more of the shape information is being utilised, the results were found to be very similar by both methods. There was one case, however, which was found to give very different scores when assessed in terms of its landmarks and surface representations. This was found to be due to the discrepancy in the upper boundaries found for the breasts,

which differed greatly in an area where no landmark information was available. Two important points should be noted from this:

Firstly, the importance of landmark selection is emphasised. In determining the set of landmarks to be used for an asymmetry analysis, it is essential that the entire area of interest is adequately described by the points to ensure that any differences such as this are recognised. It was seen in this case that the sole inclusion of the landmarks on the lower and lateral extremities of the breast led to an unreliable assessment of the differences in the upper area of the surface.

Secondly, the importance of accurate boundaries for surface data is also highlighted. As the boundary selected determines the area of the surface that will be included in the comparison, an inaccurate boundary can clearly adversely affect the analysis of asymmetry to a great extent. Much care should therefore be taken to ensure that this outer limit is captured well.

This chapter has also shown that, through the use of an appropriate reflection of the configuration to be matched, the asymmetry decomposition technique discussed in Bock and Bowman (2006) can be usefully extended to the application of matching symmetry. As opposed to one single value quantifying the asymmetry of a surface, this provides a much more informative description of the shape differences that are present.

Chapter 6

A Longitudinal Study of Facial Growth

While the previous chapter saw a brief investigation into the facial differences between the cleft and control children, focus will now shift to analysis of the control data alone. This will consist of an exploration of the 'normal' growth patterns that are seen in the faces of young children.

6.1 Introduction

Over the years, many studies have been presented that look at facial growth in children who present with dental or craniofacial problems, e.g. Ras et al. (1995), Kusnoto et al. (1999) and Dellavia et al. (2010) to name but a few. However, there is considerably less information available on the normal facial growth seen in the wider population. This information is of great interest, and could be used to assess whether a particular child appears to be growing at a healthy rate, as well as to examine the facial differences present in children with certain medical conditions, such as those born with a cleft lip and palate or with various kinds of syndromes.

Several studies have been presented that look at normal facial growth using two-dimensional images, often by means of assessing landmark points and the distances

and angles between them, for example Bishara et al. (1985), Chvatal et al. (2005) and Jacob and Buschang (2011). One frequently used approach involves a principal components analysis of the landmark points, which is described in Morris et al. (1999) amongst others. In this method, the landmark configurations are registered to one another before principal components are used to examine the changes in size and shape over time. Other methods for modelling growth, such as a parallel growth model for example, have been proposed in Kent et al. (2000, 2001).

However, the studies previously mentioned have all been conducted on two-dimensional data. As the face is a three-dimensional object it would of course be preferable to have information on the changes occurring in all three dimensions, as discussed in Hennessy and Moss (2001) and Kau and Richmond (2008). This is an area where considerably less work has been done to date. Ferrario et al. (1999) analysed the 3-dimensional changes in a cohort of 8-15 year olds, while Yamada et al. (2002) assessed 3-d data for a younger age group, however both studies were based solely on a set of landmark points. While this is a good starting point and can provide a general overview of facial growth, it is clearly not ideal for investigating the changes over a dense surface in detail.

An advancement to these landmark methods was presented by Barry and Bowman (2008). In a study which compared the growth of children born with a cleft lip and/or palate to that of a group of control children, the facial surfaces were described in terms of curves as well as landmark points. To reduce the dimensionality (and therefore complexity) of the problem, these curves were parameterised by B-splines, with a linear mixed effects model being applied to the resulting spline coefficients.

Recently however, further advancements have been made to the assessment of growth in terms of surface data, with Kau and Richmond (2008) and Hutton et al. (2003) both investigating changes over the entire facial surface in three dimensions. Kau and Richmond (2008) analysed the changes appearing between the ages of 12 and 14 in a group of children, all 55 of whom were captured at 5 distinct timepoints over the two year period. No modelling was conducted on the data however, with the analysis consisting solely of a calculation of the distances between the aligned

images for each patient, as well as between the average male and female faces at each timepoint.

The study by Hutton et al. (2003) examined growth over a longer time period, namely 0 to 50 years, based on a cohort of 400 subjects. However, their data was not longitudinal, consisting of only a single image for each subject. To deal with this, a kernel smoothing technique was utilised in order to compute the average growth trajectories for each sex. This technique, however, assumes that the pattern of growth is not significantly different between different individuals of the same sex. Hennessy and Moss (2001) have shown that this is untrue, concluding from their three-dimensional landmark analysis that there were significant differences in the growth trajectories of different subjects.

The dataset presented in the following study is the same as that used in Barry and Bowman (2008), and was introduced in Chapter 2. For the purposes of this study, however, only the control data is analysed. This consists of longitudinal data from 194 different children, captured at several timepoints between the ages of 3 and 60 months. In an extension of the work done by Barry and Bowman (2008), the facial changes are examined over the surface as a whole, with the hope of gaining a more informative picture of growth over time. At the time of writing, there has been no longitudinal study reported which investigates the three-dimensional facial changes in children of this age group in terms of an analysis of the entire surface.

The chapter proceeds as follows. Following the methodology of Cootes et al. (1992), Section 6.2 details the building of a Point Distribution Model, and presents findings from the resulting principal components analysis. Then, as in the previous study by Barry and Bowman (2008), a linear mixed-effects model is used to capture the complex structure of the data. These models are introduced in Section 6.3, along with a description of the model refinements that were required for this particular case. The findings of the model are presented in Sections 6.4 - 6.6, including an investigation into the effect of gender, before a general discussion is given in Section 6.7.

6.2 A Point Distribution Model

As detailed in Chapter 3, a set of corresponding meshes was created for all images in the dataset. These meshes all had an equal number of vertices, the positions of which were corresponding between and within subjects. The advantage of this setup is that each point can be treated as a landmark and compared across different surfaces, enabling the use of standard landmark analysis techniques such as Procrustes methods and Principal Components Analysis (PCA).

The use of PCA to examine the variation in a set of corresponding surfaces is known as a Point Distribution Model (PDM). This was a term first coined by Cootes et al. (1992), with further applications of the technique being presented in Andresen et al. (2000) and Hutton et al. (2001). Following the methodology of these studies, it was first necessary to align all surfaces. This was done by means of a partial Generalised Procustes Analysis (GPA), where all $k = 874$ points of each mesh were treated as the landmark points for matching. This technique is described fully in Section 3.1.1, and involves a least squares alignment of all configurations in order to remove the effects of location and orientation. As it was desired to explore both the shape and size changes over time for this study, scaling was not included in the alignment procedure. However, when solely the shape changes are of interest, it is simply a matter of scaling all configurations to be of unit size prior to the analysis being conducted.

After application of the GPA, each aligned configuration X_i can be represented by its vector of coordinates $\mathbf{x}_i = (x_1, y_1, z_1, \cdots, x_k, y_k, z_k)$. This leads to a mean shape vector of

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i,$$

where $n = 683$ is the number of images in the dataset. The Procrustes residuals for each case are then given by

$$\boldsymbol{r}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}$$

These residuals were introduced in Section 3.1.1 and express the deviation from the mean at each point, making them useful for exploring the variation in the dataset.

They are an approximation to the tangent coordinates, and working in this linear space as opposed to the original shape space allows us to perform standard multivariate analyses. From these residuals, the $k \times k$ covariance matrix $\boldsymbol{S}$ can be calculated as

$$\boldsymbol{S} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{r}_i \boldsymbol{r}_i{}^T.$$

The eigenvectors of this matrix then represent the different modes of variation, or principal components (PC's), of the data, while the eigenvalues indicate the proportion of variability that is accounted for by each of these PC's. However, as discussed in Hutton et al. (2001), the same first n eigenvalues, λ_i, can instead be obtained through the eigendecomposition of the matrix

$$\boldsymbol{T} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{r}_i{}^T \boldsymbol{r}_i.$$

As $n < k$ for our data, this matrix is smaller than S (at $n \times n$ elements), resulting in an easier calculation of the eigendecomposition. The first n eigenvectors of $\boldsymbol{S}$ can then be calculated from those of $\boldsymbol{T}$ by $\boldsymbol{\phi_i} = \boldsymbol{R}_i \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ are the eigenvectors of $\boldsymbol{T}$ and $\boldsymbol{R}_i$ is the matrix with the vectors of Procrustes residuals as its columns.

The standardised PC score of the j^{th} component for a particular image i is then calculated as

$$s_{ij} = \frac{\boldsymbol{\phi}_j \boldsymbol{r}_i}{\lambda_j^{\frac{1}{2}}}.$$

Under this type of model, it is generally found that a fairly small number of components are sufficient to adequately describe a large proportion of the variation present. In order to describe a desired amount, say $P\%$, of the variability, it is only required to retain the first p components, where p is the smallest value such that

$$\frac{\sum_{j=1}^{p} \lambda_j}{\sum_{i=1}^{n} \lambda_i} \geq 100P\%$$

From this subset of the components, any shape in the dataset can then be recreated by adding a weighted combination of the corresponding eigenvectors to the mean

face, i.e.

$$x = \bar{\boldsymbol{x}} + \boldsymbol{\Phi b}$$

where $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1 \ ... \ \boldsymbol{\phi}_1)$ is the matrix of the first p eigenvectors and $\boldsymbol{b} = (b_1 \ ... \ b_p)^T$ is the set of weights for each respective eigenvector. As discussed in Cootes et al. (1992), since the variance of b_i can be shown to be λ_i, any face that is modelled should lie within the variation seen in the dataset provided that, for all i,

$$-3\sqrt{\lambda_i} \leq b_i \leq +3\sqrt{\lambda_i}.$$

6.2.1 Results of the Principal Components Analysis

Applying these methods to the face data it was found that only 8 PC's were necessary to explain more than 90% of the variability present, with the first two PC's together capturing 79.4% of the variation. The trends of the scores for these 8 PC's over time can be seen in Figure 6.1, along with 95% confidence intervals. The profiles of the scores suggest that the effect of time is not necessarily a linear one, something which should be considered when fitting a model.

The scores for PC1 were strictly increasing over time, which suggested that this component was perhaps reflecting the increase in size as the children grew older. To confirm this, the centroid size of each configuration was calculated and compared to the corresponding score for PC1. This is a common measure of size and was previously defined in Section 3.1.1.

The centroid sizes were found to be very highly correlated with the scores for PC1, with a correlation coefficient of 0.998 (95% C.I. of (0.997, 0.998)). A strong relationship was also found between age and this component, as would be expected. Both of these relationships are illustrated in Figure 6.2. This suggested that the majority of the change in size was dealt with by this PC, while the remaining components were capturing the variation in shape. PC1 alone accounted for 66.0% of the variation in the data.

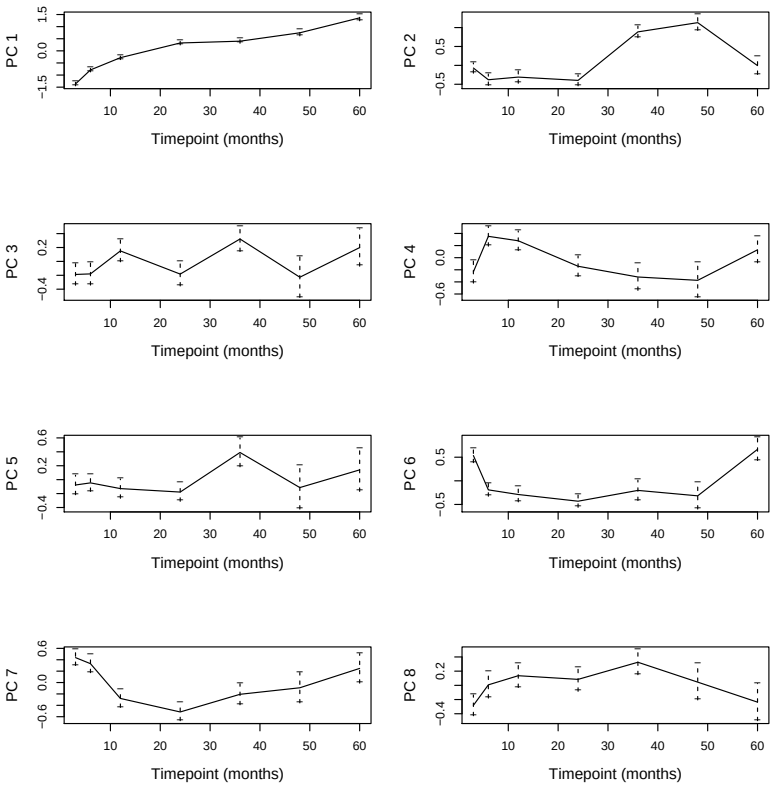

FIGURE 6.1: The mean trends over time for our first 8 principal component scores, showing 95% confidence intervals.

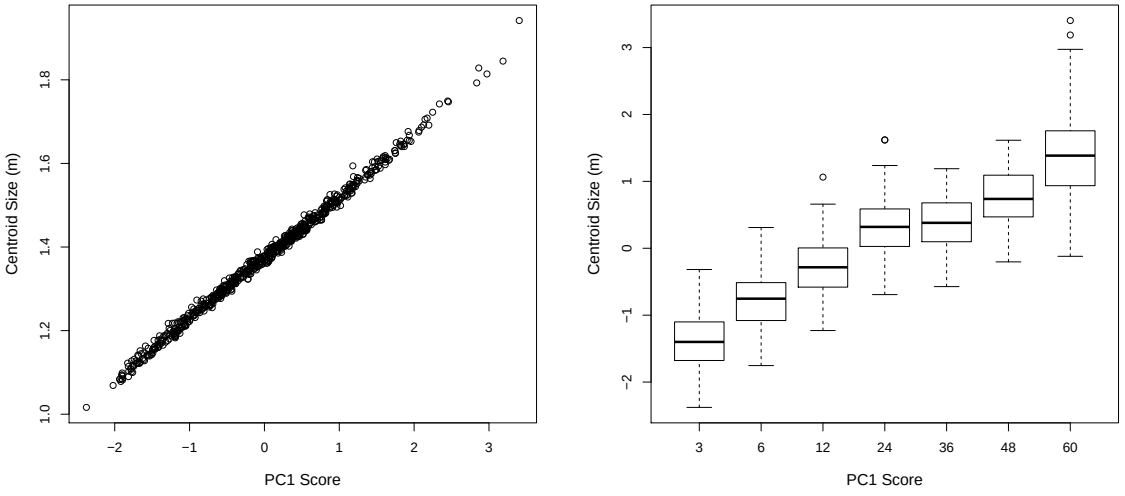

FIGURE 6.2: *Left:* The strong relationship between the first component and size. *Right:* Boxplots of the first PC score by timepoint.

By adding the corresponding eigenvector directly onto the coordinates of the mean mesh, i.e. $\bar{\boldsymbol{x}} + \gamma \boldsymbol{\phi_j}$, the individual effect of a given principal component j can be

FIGURE 6.3: The modes of variation of PCs 1-3. The mean face is seen in the centre row, with ± 1 and 3 standard deviations of the PC shown above and below.

projected back into shape space. This allows a clear visualisation of the mode of variation being described, and will show the plausible range of effect of the component provided that $-3\sqrt{\lambda_j} < \gamma < 3\sqrt{\lambda_j}$. The effects of the first three components can be seen in Figure 6.3. As has already been shown, PC1 is strongly related to the changes in size. There is also a shape effect seen for this PC, with the face appearing to lengthen from one end of the range to the other. PC2 also seems to be related to facial length, as well as changes to the roundness of the face, while PC3 seems to be describing the variation in the width of the lower face compared to the upper face.

From the scatterplot of all pairwise combinations of the PC scores seen in Figure

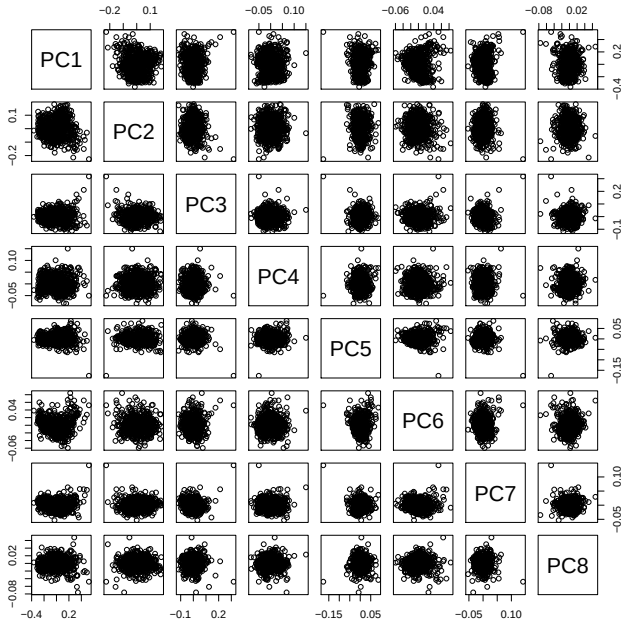

FIGURE 6.4: Scatterplot of all pairwise combinations of principal component scores.

6.4, there does not appear to be much of a significant pattern in the relationships between the various scores. This is to be expected however, due to the fact that principal components are constrained to be orthogonal to one another at a global level. While there are naturally correlations between the various points on each face, this structure can be captured by a small number of these orthogonal components, negating the need to model each of these coordinate points individually. As only the first 8 PC's are required to explain a large proportion of the variation in the data, the dimensionality of the problem can be substantially reduced by modelling the facial growth based on these PC scores directly. The following section details the approach taken to do so.

6.3 Application of the Linear Mixed Effects Model

In analysing longitudinal data, the fact that there is a relationship between repeated observations on the same individuals must be observed. We may expect there to be a particular trend over time, but this may differ slightly between different individuals. As discussed in Laird and Ware (1982), a linear mixed effects model allows for this

structure in the data by treating time as a fixed effect whilst allowing other variables, the random effects, to vary from individual to individual.

This approach is therefore fairly standard in modelling longitudinal data, and has previously been applied in a shape analysis setting by Barry and Bowman (2008). In this study, a mixed-effects model was fitted to each pairwise combination of responses (where the responses were either coordinate positions or B-spline coefficients), with the results then being aggregated across all parameter estimates. This was done following the methodology of Fieuws and Verbeke (2006). It was hoped, however, that for the present study a joint modelling of all responses would be possible. As discussed in Barry (2008), one advantage of this joint modelling approach is that the association structure between the points can be taken into account. It also allows an examination of the joint effect of time on all of the responses, i.e. the effect seen on the entire shape, rather than just on each individual PC. As the number of explanatory variables can often be large within the field of shape analysis, problems in the estimation of the covariance matrix can arise with a joint modelling approach. However, the use of PCs in place of the tangent coordinates themselves has substantially reduced the complexity of this matrix calculation, allowing this to be a more feasible approach.

As we are modelling PC scores, they are orthogonal to one another at a global level by design. This was seen by the lack of any clear pattern in their relationships in Figure 6.4. However, at a lower level it is very possible that the scores for a particular subject may be related, i.e. it is possible that the random effects, or person-specific deviations from the population fitted values, may be associated. The model should therefore allow for correlation in the random effects. Additionally, on examination of the profile plots in Figure 6.1, it is clear that the trends over time are different for each of the PCs. It is therefore necessary to allow the time variable to have a different effect on each of these scores i.e. to allow for a unique slope as well as a unique intercept for each of the components. The basic model required to jointly model our p principal components scores $\mathbf{s}$ for a given individual i can therefore be represented as

$$\mathbf{s}_i(t) = \boldsymbol{\beta}_0 + \mathbf{b}_i + \boldsymbol{\beta}_1 t + \boldsymbol{\epsilon}_i(t), \tag{6.1}$$

where the random error vector, $\boldsymbol{\epsilon}_i(t)$ follows a $N(0, \Sigma_p)$ distribution, with Σ_p being the diagonal matrix with elements $\sigma_1^2, \cdots, \sigma_p^2$. The random intercepts $\mathbf{b}_i$ are distributed as $N(\mathbf{0}, D)$, where

$$D = \begin{pmatrix} \tau_1^2 & \tau_{12} & \cdots & \tau_{1p} \\ \tau_{21} & \tau_2^2 & \cdots & \tau_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{p1} & \tau_{p2} & \cdots & \tau_p^2 \end{pmatrix}.$$

A plot of the standardised residuals against fitted values for this model can be seen in Figure 6.5, with a loess smooth fitted line overlaid. When the residual variance over timepoints is constant this line should be a horizontal cut at zero. As interest lies primarily in the modelling of the fixed effects, the fitted values are taken here as the population rather than individual fitted values, i.e. without the inclusion of the random effects. While the variance is relatively constant for some of the components, for a small number (i.e. PC2, PC4 and PC6) it seems to differ somewhat over time. This suggests that the current model is not capturing the variation in the data optimally. In order to model this complex dataset more adequately, however, there are ways in which the model can be extended, and several of these are discussed in the following sections.

6.3.1 Investigation into non-linear time effect

In the previous model time is treated as a continuous variable, forcing it to have a linear effect within the model. However, due to the non-linear appearance of the profile plots in Figure 6.1 it may be more appropriate to treat this variable as a factor, allowing it to have a more flexible effect. This can be confirmed by examination of Figure 6.6, where it can be seen that the linear trend fitted across time in Model 6.1 does not produce a particularly good fit to our data. In order to improve on this, the model can be altered to be of the form

$$\boldsymbol{y}_i(t) = \boldsymbol{\beta}_0 + \boldsymbol{b}_i + \boldsymbol{\beta}_1 t_6 + \boldsymbol{\beta}_2 t_{12} + \boldsymbol{\beta}_3 t_{24} + \boldsymbol{\beta}_4 t_{36} + \boldsymbol{\beta}_5 t_{48} + \boldsymbol{\beta}_6 t_{60} + \boldsymbol{\epsilon}_i(t), \qquad (6.2)$$

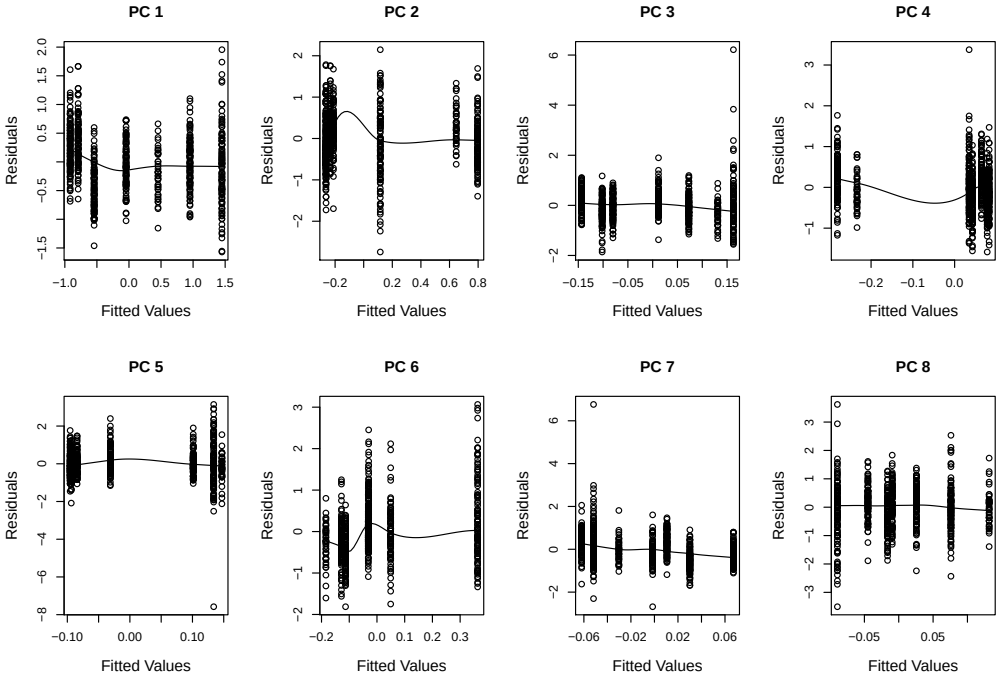

FIGURE 6.5: For each PC individually, the residuals plotted against the fitted values for the model with a continuous time effect.

Here, the variables $t_6, \cdots, t_{60}$ are indicator variables which take the value 1 at timepoints 6 to 60 months respectively, and are 0 otherwise. The baseline is therefore the 3 month timepoint, at which each of the indicator variables take a value of 0. Under this model, the random intercepts $\boldsymbol{b}_i$ and random error vector $\boldsymbol{\epsilon}_i(t)$ are distributed exactly as in Model 6.1.

The fitted values arising from this model are shown in Figure 6.7 alongside the empirical means. As can be seen the fit seems to be very good, with little deviation from the actual mean PC scores. This would be expected due to the allowance of a much more flexible effect for time, and is clearly a vast improvement in fit when compared to Figure 6.6.

These two models are nested, with Model 6.1 being a restricted version of Model 6.2, and so the two models can be compared by means of a likelihood ratio test. The χ^2 statistic for this test is computed as $\chi^2 = -2(LL_0 - LL_1)$, where LL_0 and LL_1 are the log-likelihood values for the restricted and full models respectively (seen in Table 6.1). This statistic must then be compared to a chi-squared distribution, where the degrees of freedom are equal to the difference in degrees of freedom (or difference in

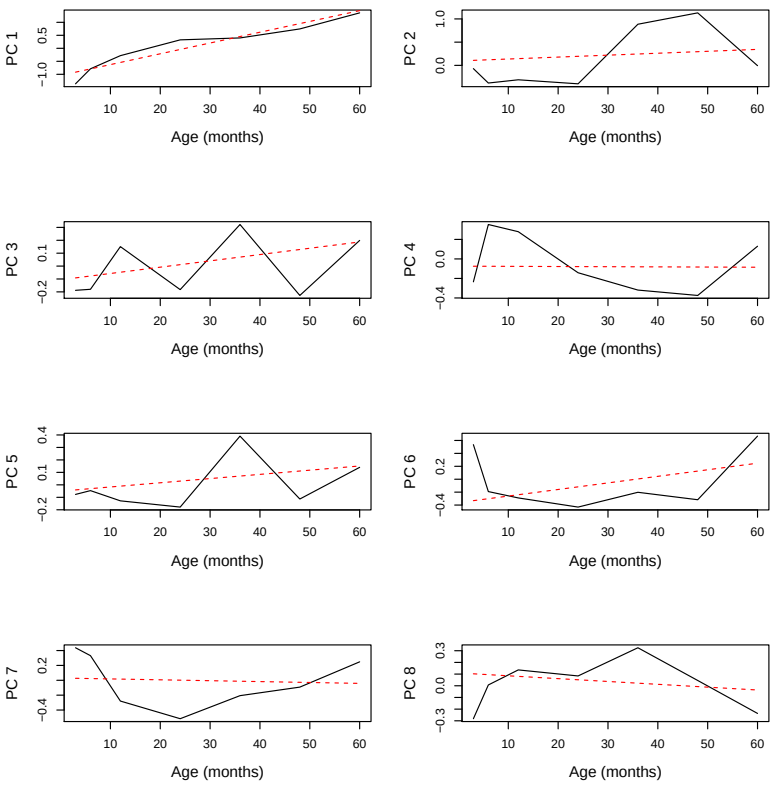

FIGURE 6.6: The empirical mean PC scores across all timepoints (black) along with the fitted values from Model 6.1 (red). This model assumes a linear time effect.

Model	Log-Likelihood	Number of Parameters
6.1	-6945.8	53
6.2	-6597.6	93

TABLE 6.1: Log-likelihoods and numbers of parameters for the models with both linear and non-linear time effects.

the numbers of parameters) for the two models. In this case the resulting value of the statistic is $\chi^2 = 696.5$, corresponding to a p-value $< 1 \times 10^{-6}$ when compared to the χ^2_{40} distribution. This confirms that the allowance of a non-linear time effect is producing a significantly better fit, and should therefore be retained in our model.

This is further confirmed by the improvement seen in the plot of residuals against fitted values for this model, shown in Figure 6.8. The variance now appears to be relatively constant across the timepoints for all PCs, suggesting that a good fit to the data has been achieved.

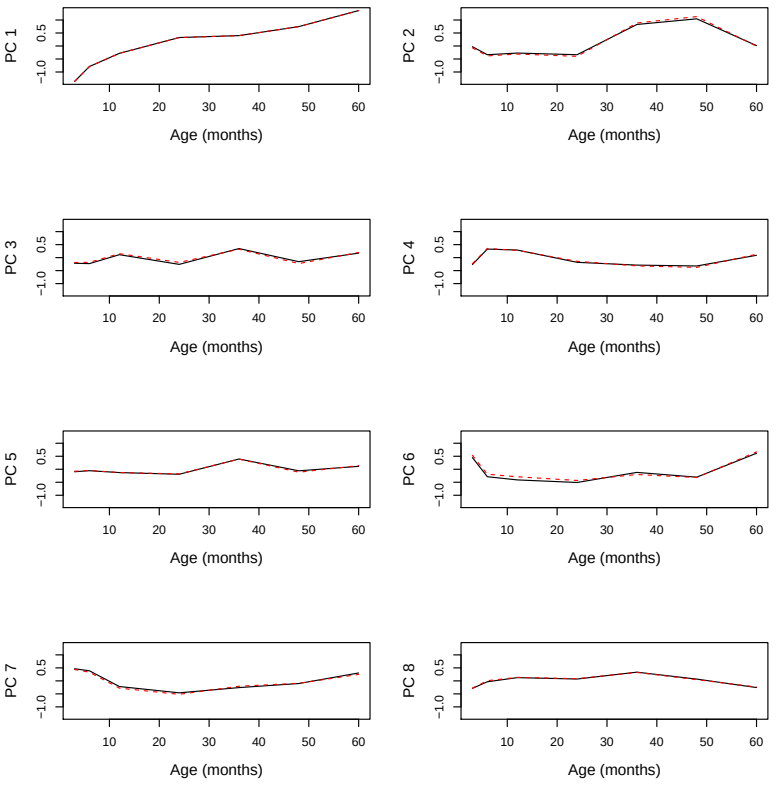

FIGURE 6.7: The empirical mean scores over time for the first 8 principal components (*black*), with the corresponding fitted values overlaid (*red*), when time is allowed to have a non-linear effect.

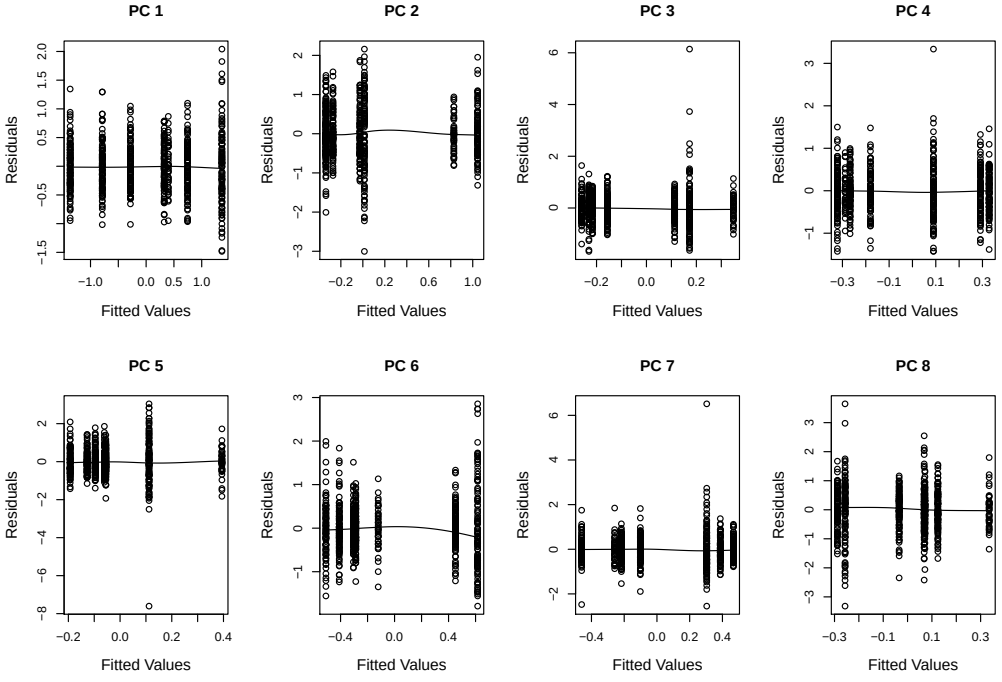

FIGURE 6.8: For each PC individually, the residuals plotted against the fitted values for the model with time as a factor.

6.4 Interpreting the Model

The fitted model is assessing the growth of the subjects as a function of the fixed effect, namely time. It can therefore be used to examine the average face that we would expect to see at each timepoint through the use of an appropriate subset of the fixed effects, $\hat{\boldsymbol{\beta}}$. However, as we are modelling the PC scores of the Procrustes residuals, the results must be projected back to the shape space before they can be visualised. To do so, the mean shape vector must be added to the estimated mean vector of tangent coordinates for that particular combination of fixed effects. The expected face for a timepoint corresponding to the fixed effects $\hat{\boldsymbol{\beta}}$ is therefore calculated as

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}} + \sum_{j=1}^{p} \lambda_j^{1/2} s_j(\hat{\boldsymbol{\beta}}) \boldsymbol{\phi_j},$$

where $s_j(\hat{\boldsymbol{\beta}})$ is the estimated average value of the j^{th} PC score for $\hat{\boldsymbol{\beta}}$.

As well as a point estimate of the mean face for a given timepoint, a measure of the uncertainty of the estimates can also be derived. Using var($\boldsymbol{\beta}$), the variance-covariance matrix of the fixed effects, confidence regions can be created around each vertex of the mesh. The first step in doing so is to extract the subset of var($\boldsymbol{\beta}$) which is relevant to the fixed effects $\hat{\boldsymbol{\beta}}$, say var($\hat{\boldsymbol{\beta}}$). Using the reduced model matrix X required for this particular set of effects, the variance of the mean vector of tangent coordinates can then be calculated as

$$var(\hat{\boldsymbol{x}}) = \boldsymbol{\phi} \boldsymbol{\lambda}^{1/2} X var(\hat{\boldsymbol{\beta}}) X^T (\boldsymbol{\lambda}^{1/2})^T \boldsymbol{\phi}^T.$$

The estimated mean face for each timepoint is displayed in Figure 6.9. As is unsurprising due to the dominance of the first PC, the most noticeable changes over time are an increase in size and a lengthening of the face. The face also appears to become less flat as the child grows older, however this is more easily visualised in three dimensions.

Examples of the 95% confidence regions can be seen in Figure 6.10, for both the 6

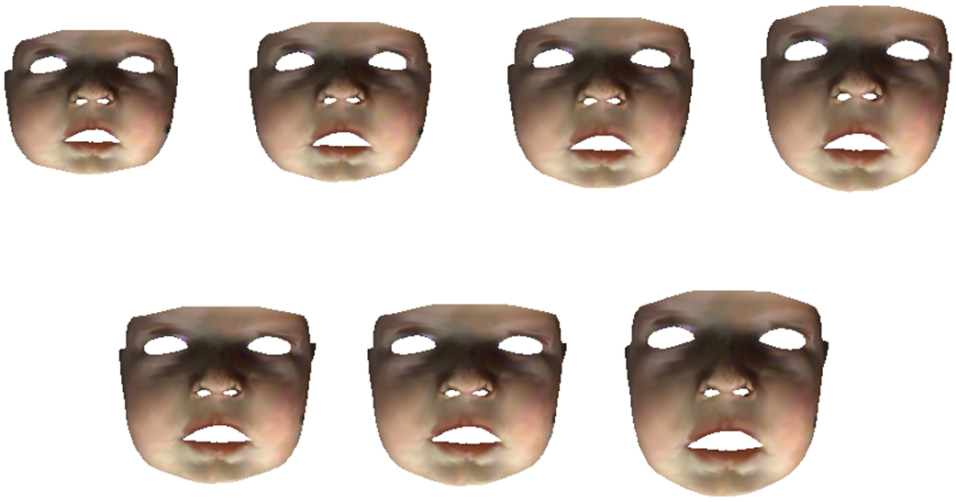

FIGURE 6.9: The estimated average face at each timepoint. Top row (L-R): 3, 6, 12 and 24 months. Bottom row (L-R): 36, 48 and 60 months.

and 36 month mean faces. The majority of the regions are ellipsoidal, with the majority of the variation being in the vertical and anterior/posterior directions. This is perhaps unsurprising since these are the main directional changes illustrated by the first and second components, which cumulatively accounted for almost 80% of the variation in the data. Across all timepoints, it was found that there is more uncertainty present in the areas towards the edge of the mesh, while the smaller confidence regions around the central face suggest that our model can more accurately explain the growth in this area. This is likely to be partially due to the variation induced during the warping procedure, when the corresponding surfaces were being created. Of the 22 landmarks used to calculate the warp (Fig 3.4), the majority lie in the central region of the face, leading to a greater accuracy of the transformation in this area. This will subsequently lead to less variation in the surface points around this area.

6.5 Testing for a Sex Effect

While the previous model describes the average change in a child's face over time, it is also of interest to examine whether there is a difference in these growth patterns between males and females. In order to investigate this, the model for an individual

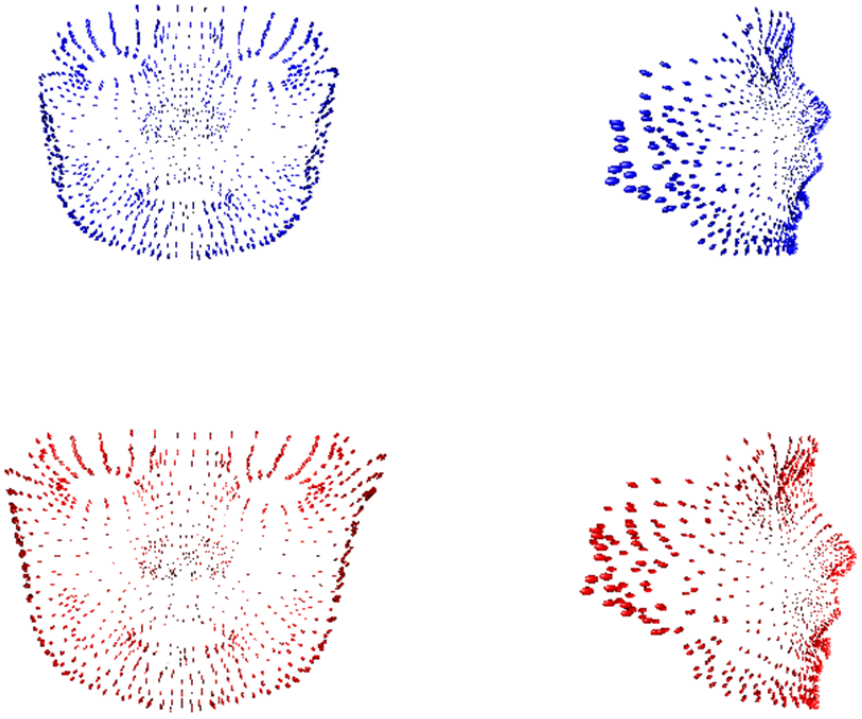

FIGURE 6.10: 95% confidence regions for a child of both 6 months (blue) and 36 months (red).

i can instead be denoted as

$$\boldsymbol{y}_{ij}(t_k) = \boldsymbol{\beta}_0 + \boldsymbol{b}_i + \boldsymbol{\beta}_{jk} + \epsilon_{ij}(t). \tag{6.3}$$

In this case, j is an indicator variable for sex, taking a value of 0 for females and 1 for males, and k is an indicator variable for the timepoint. The $\boldsymbol{\epsilon}_{ij}(t)$ and $\boldsymbol{b}_i$'s vectors are again distributed as before.

The profiles of the PC scores for both males and females can be seen in Figure 6.11, and again show a very good fit to the data. Plotting the residuals against the fitted values, this model including a sex term seems to be an improvement, as the loess smoothed lines are reasonably flat for all components (Figure 6.12). By comparison of the maximised log-likelihoods with that of Model 6.2 (shown in Table 6.2), the inclusion of the sex term was found to be significant at p-value < 0.0001, suggesting that differences exist in the male and female growth patterns at this age.

Model	Log-Likelihood	Number of Parameters
6.2	-6597.6	93
6.3	-6544.7	149

TABLE 6.2: Log-likelihoods and numbers of parameters for the models with and without a sex effect.

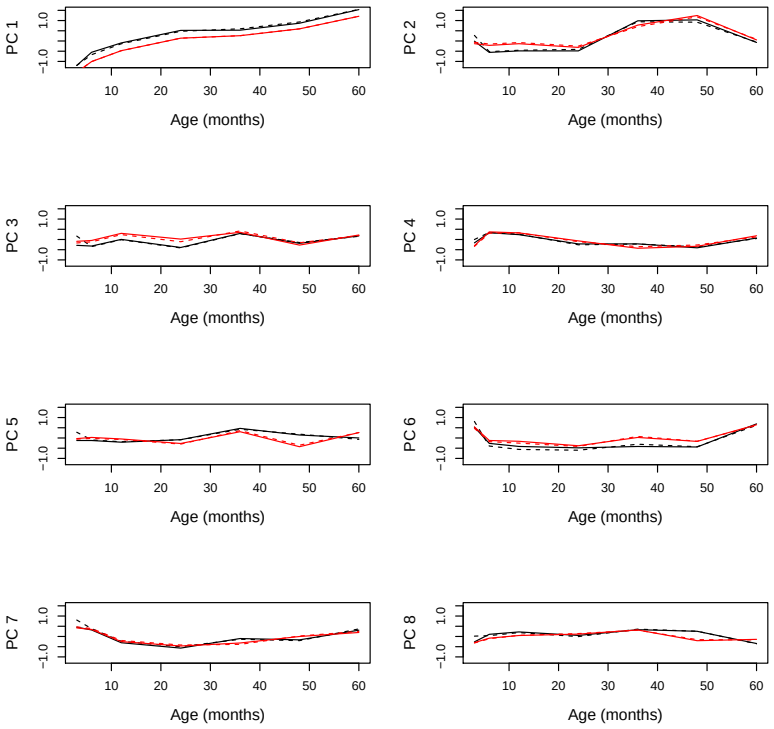

FIGURE 6.11: The trends over time for our first 8 principal component scores for both males (black) and females (red). The empirical means are shown along with the fitted values from our model (represented with dashed lines).

To confirm that this was a genuine effect, a permutation test was applied to the data. This involved an arbitrary reordering of the factor representing sex, so as to randomly assign each child as either male or female, before Model 6.3 was refitted. A new χ^2 statistic was then calculated to assess whether sex was still significant under this new labelling. Repeating this process 500 times, the distribution of the test statistic was as seen in the left panel of Figure 6.13. The statistic was found to be higher than our observed value of 105.6 in only 3 (or 0.6%) of the 500 permutations. This indicates that it is highly unlikely that the sex effect seen in our data is down to chance and that there is a genuine difference between the boys and the girls over this time period.

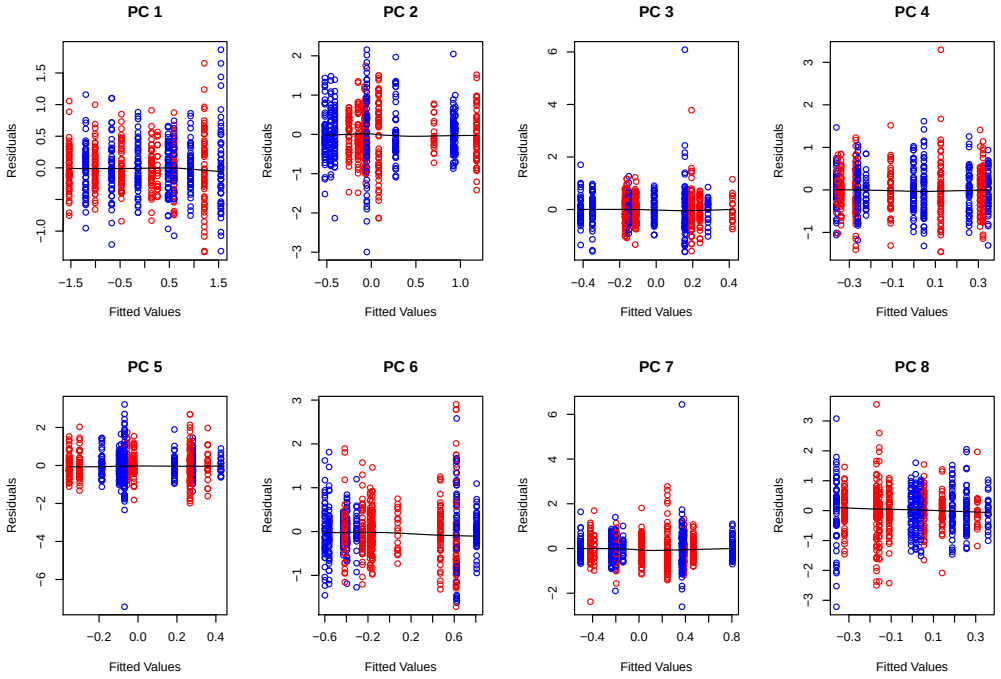

FIGURE 6.12: The residuals plotted against the fitted values for each of the PCs individually, for the model including a sex effect. Females are represented by red points while males are represented by blue.

The 95% confidence regions under this model can be seen for each sex independently in Figure 6.14, for a child of age 6 months. There does seem to be a slight separation in the two surfaces, as would be expected from the previous results. However,

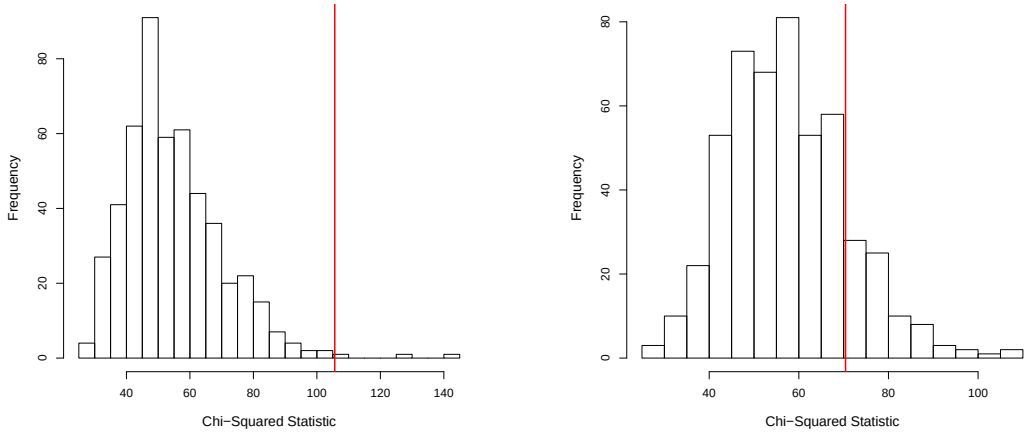

FIGURE 6.13: The distributions of the chi-squared statistic after 500 permutations of sex factor, for both the model including size (left) and the model where size has been removed (right). The red line indicates the value of the test statistic under the original ordering.

looking at the profiles in Figure 6.11, there only appears to be a clear difference

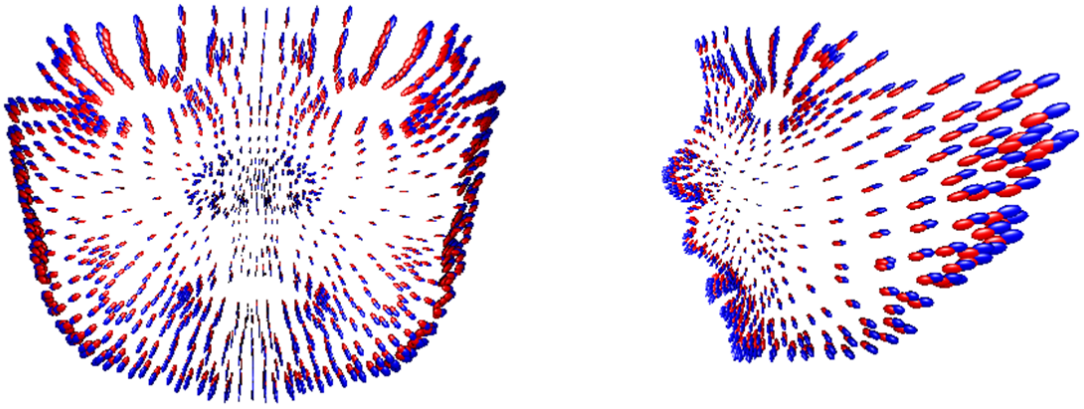

FIGURE 6.14: 95% confidence regions for both boys (blue) and girls (red) at 6 months old.

between the two sexes for the first principal component i.e. the component related to size. For all other components there is no clear separation between the two genders. The PC scores for the first component were generally higher for males, indicating that they tended to have larger faces than females over this age range. This size difference is also apparent in Figure 6.14.

To investigate whether the facial shape was changing differently for the two sexes, as opposed to just the size, the size differences were removed from the analysis. This involved scaling each corresponding mesh to be of unit size, prior to performing the principal components analysis and fitting the models. The scaling was achieved by dividing each configuration by its centroid size $S(X)$.

Under this new setup, the first PC is no longer related to size, as can be seen in Figure 6.15. Refitting the model, there are no longer any clear differences between the two sexes for any of the first 8 PC's, as illustrated in the profile plots of Figure 6.16. When applied to the scaled data, the likelihood ratio test between this Models 6.2 and 6.3 gives a p-value of 0.09, indicating that the sex term in the model is no longer significantly improving the fit. The log likelihoods for each of the models can be seen in Table 6.3.

This result was again be verified by means of a permutation test, the distribution of which can be seen in the right panel of Figure 6.13. In this case, the χ^2 statistic was found to be higher than the observed value of 70.4 in 73 (or 14.6%) of the

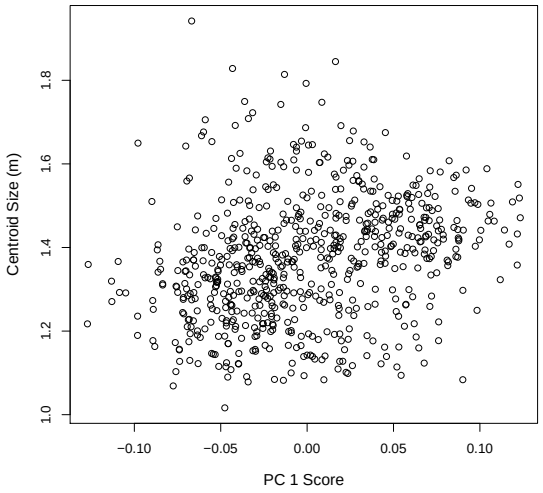

FIGURE 6.15: Scatterplot of centroid size versus scores for PC1 for the case where all meshes have been scaled to unit size.

Model	Log-Likelihood	Number of Parameters
6.2	-6719.4	93
6.3	-6684.1	149

TABLE 6.3: Log-likelihoods and numbers of parameters for the models with and without a sex effect, when applied to the scaled data.

500 permutations. As this is higher than the usual equivalence level of 5%, there is therefore no significant evidence of a sex effect when working with these scaled surfaces. This suggests that, outwith the discrepancies in size, the two genders illustrate similar patterns of growth.

6.6 Interpolation Between Timepoints

While an estimate has been calculated for the average face seen at each available timepoint, advantage can be taken of the fact that the face should grow in a smooth fashion between these timepoints. By means of a cubic spline interpolation, an estimate can therefore be obtained of the average face at any timepoint within our range of data i.e. the estimated mean face at any age between 3 and 60 months.

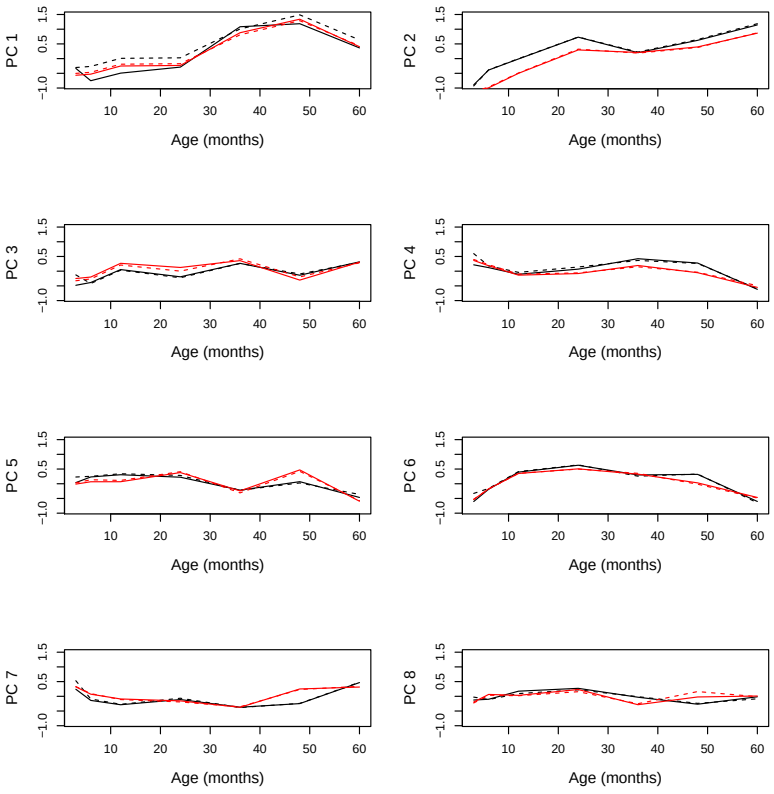

FIGURE 6.16: The trends over time for our first 8 principal component scores for both males (black) and females (red), after the removal of all size differences. The empirical means are shown along with the fitted values from our model (represented with dashed lines).

For each individual item x_i of the coordinate vector $\mathbf{x}$, this can be done by fitting a spline to the expected positions of that item at all seven timepoints. The resulting fit can then be used to determine where that point would lie at any particular time. An example of this can be seen in Figure 6.17, both for an individual point and the resulting overall interpolated surface. These cubic splines are established statistical methods for which well-defined algorithms and software are available. A description of the methodology can be found in Green and Silverman (1994).

A visualisation of the resulting growth pattern can be seen at `http://www.stats.gla.ac.uk/~jsmith/growth1.gif`. Looking at this smooth representation of growth, however, a possible issue with the data is highlighted. Between the ages of 24 and 36 months the face appears to decrease in length, something which clearly goes against our knowledge of human growth. To check that this was indeed an artefact of the data and not a problem induced by the model, the distributions of facial length and

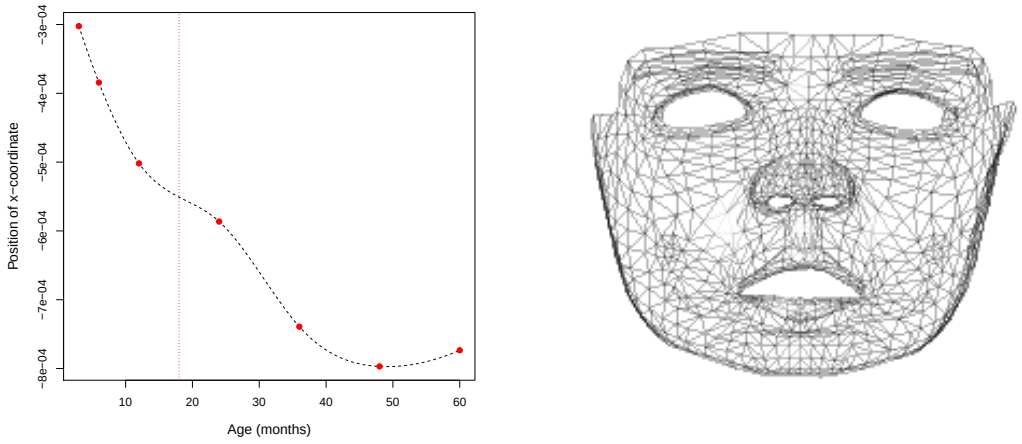

FIGURE 6.17: *Left:* The cubic spline fitted to the x-coordinate position of a point in our standardised mesh. The estimated position of the point at 18 months can be seen. *Right:* The overall predicted surface for a child of 18 months.

width were investigated over the various timepoints. Using a subset of the landmarks seen in Figure 1.1, the Euclidean distance between the points n and sl was calculated for each individual as an indicator of facial length, while the distance between exL and exR was taken to represent the facial width. The left column of Figure 6.18 shows the distribution of these measurements over time, and confirms the trend seen in our model fitting. While the width of the face is strictly increasing over time, the length appears to decrease slightly between the 24 and 36 month timepoints.

Investigating this further, it seemed possible that this issue related to a change in the natural resting position of the mouth. Up to and including the 24 months timepoint, the vast majority of the subjects had been captured open-mouthed. However from 36 months onwards, the opposite was true. This change in resting position was apparent in many of the subjects, including the child seen in Figure 2.8. The distribution of the mouth positions over time can be seen in Figure 6.19.

To confirm that this issue was indeed being caused by the change in mouth position, the landmark distances were looked at once more, however this time with all closed mouth cases excluded. As can be seen in the right column of Figure 6.18, looking solely at these cases was seen to remove this issue with the facial length, confirming

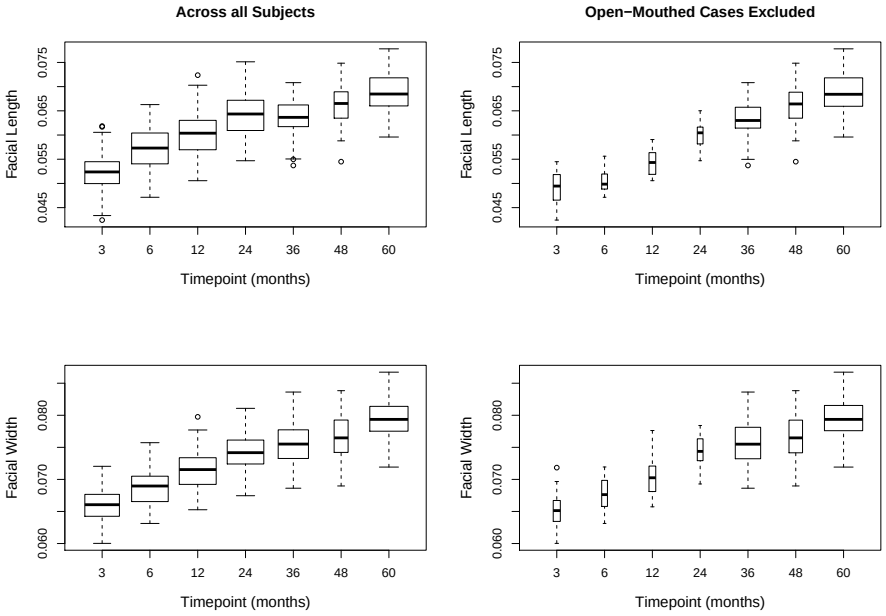

FIGURE 6.18: The distributions of facial length and width over time, with and without the open-mouthed cases. The width of each box is proportional to the sample size at that timepoint.

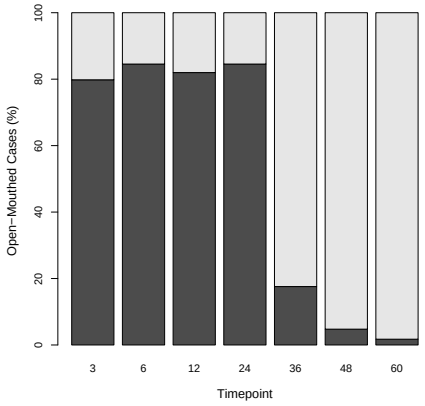

FIGURE 6.19: The proportion of images with open mouths (dark grey) and closed mouths (light grey).

that the artefact seen in our results is in fact down to the common change in mouth position.

While one possible solution to this problem would be to exclude either the open or closed mouthed cases in order to work with a homogeneous subject group, as was seen in Figure 6.19 this would substantially reduce the sample size. An alternative approach however, and the one utilised here, is to retain all subjects in our analysis but instead remove the area where the problem occurs - namely the lower part of

the face. To do so, a reduced base mesh was created which excluded all points in the lower jaw. This was done simply by taking a planar cut through the original base mesh, in order to remove a particular subset of the points. The new base mesh can be seen in Fig 6.20, and was used to create a new set of corresponding meshes that were based only on this upper section of the face. Applying the PCA to these

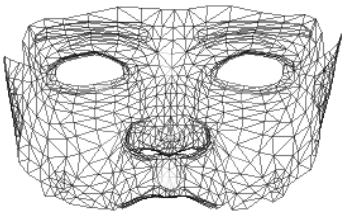

FIGURE 6.20: The reduced base mesh, with all points in the the lower jaw excluded.

meshes and modelling the principal component scores as before, the resulting facial growth trajectory was found to be strictly increasing in all directions as required. This can be seen at http://www.stats.gla.ac.uk/~jsmith/growth2.gif.

However, an obvious drawback of this method is that only the changes seen in the upper face are modelled by this method. A visualisation of the whole face can be created through the warping of the original base mesh to these upper face estimates, and this was achieved through the Thin Plate Spline method described in Chapter 3. Through this approach, however, the growth seen in the lower face is artificial and would require verification. Again, a three-dimensional representation of the growth pattern by this method can be found at http://www.stats.gla.ac.uk/~jsmith/growth3.gif.

6.7 Concluding Remarks on Facial Growth

This chapter has explored a technique for the modelling of growth in 3-d longitudinal shape data. While non-longitudinal data can be collected more quickly and generally

in greater volume, its analysis in terms of growth modelling generally makes the assumption that all individuals have the same growth trajectory. As has been shown by Hennessy and Moss (2001) amongst others, this is not necessarily true. A key advantage of longitudinal data such as this is therefore that we can allow for subject-specific variations in the growth pattern, something which is easily achieved through the use of a mixed effects model.

Another advantage to the use of a mixed effects model is the flexibility allowed for by their hierarchical structure. As discussed in Goldstein (1986), due to the fact that average as well as individual trends are being assessed, all subjects can be utilised under this model even when data is only available at one or two timepoints. This was extremely advantageous here, as the majority of the subjects in the dataset did not have images available at all timepoints.

The use of a principal components analysis substantially reduced the dimensionality of the problem. Taking advantage of the natural correlations present in the various surface points, the variance in the dataset was successfully reduced down to several orthogonal modes of variation. It was found that only 8 components were required to explain the vast majority of the surface variation across all cases, meaning that any potential difficulties in the estimation of the covariance matrix were sidestepped. This allowed for a joint modelling of all eight components, and therefore a joint modelling of the overall shape changes.

The fitted model allowed for correlation of the random effects as, even though the PCs are orthogonal at a global level, the PC scores for a given individual may be associated with one another. It was also found necessary to include time in the form of a factor in the model, thereby allowing the rate and direction of growth to occur in a non-linear fashion.

The change in prevalence from open mouthed to closed mouthed cases is of concern when interpreting the results of this model. It is clear that this shift was having a large effect on the estimated mean faces and resulting in an unrealistic decrease in facial size after 24 months. While fitting the model based only on the upper face removes this artefact, it is clearly not ideal as the true changes in the lower

jaw remain unknown. This will perhaps always be an issue when modelling facial changes from such a young age, as it would clearly be challenging to ensure that a 3 month old child closes their mouth during the image capture process.

Another possible source of error due to this change in the mouth position is the selection of the base mesh. The current method employs a base mesh with an open mouth, to which each of the cases are warped based on the landmark points. While this clearly may not be ideal for the images when the mouth is closed, changing it to a base mesh with a closed mouth would result in the same issue arising with the open mouthed cases. One way around this could perhaps be by creating the corresponding surfaces in a different manner, perhaps by interpolating between the landmark points of each individual case to create a large set of semi-landmarks across the surface as a whole. This would remove the need for the use of a base mesh.

When looking at relatively closely spaced timepoints such as these, it seems reasonable to assume a smooth and relatively constant rate of transformation between each of the measurements, something which is necessary to justify the spline interpolation used between the fitted means. However, with more sparse data this assumption may not be sound. As is widely accepted, humans undergo periods of increased growth or 'spurts', during which this simple interpolation between data points may not be appropriate. This should possibly be looked at in greater detail.

Generally, the methods demonstrated in this chapter have proven to be a useful way of investigating the facial changes that occur between the ages of 3 and 60 months in this cohort of children. In addition to providing an estimate of the mean face at each timepoint, the model can be used to calculate a 'normal' range of faces for that age and therefore determine whether a given child is growing as would be expected. Overall, the key growth patterns seen over this time period were an increase in size and a lengthening of the face. This fits well with what would be expected in the transition from baby to toddler. Upon investigating the effect of sex, it was found that while males tended to have slightly larger faces than females over this age range, the facial shapes in themselves were similar. It is very possible that this would not

be the case at a later age, however, when the differences between the two sexes become more apparent.

Although the current dataset has allowed us to model the average growth trajectory for a child of up to 5 years, there is no reason why this cannot be expanded to a larger age range. There may be scope for follow up imaging of these same children at a later date.

Chapter 7

Prediction of Soft Tissue Deformation

The final study discussed in this thesis is concerned with modelling the changes that are seen in the facial soft tissues as a result of craniofacial surgery, specifically for the purposes of predicting the post-surgical appearance of a patient. Although still in the realms of facial modelling, the approach taken here differs greatly to that of the previous study, as will be seen throughout this chapter.

7.1 Introduction

With many types of reconstructive surgery an accurate prediction of the post-surgical outcome is an essential tool. It provides the opportunity to compare different surgical options before settling on a treatment plan and can also be used to give the patient an indication of post-operative appearance. One area where this prediction is very desirable is in the field of orthognathic surgery, a type of surgery which aims to correct deformities of the jaw. It involves the cutting and repositioning of either the maxilla, mandible or both, without which the patient may have problems with breathing and mastication, as well as an unusual appearance.

Clearly this procedure can have a huge effect on the patient's appearance, and as such it would be a useful resource to have an idea of the likely outcome for a given planned procedure. This would be an invaluable asset to the patient to allow them to prepare themselves for the change as well as to ensure they wish to go ahead with the surgery. It would also be an extremely useful tool for surgical planning.

However, a problem with the prediction of post-operative appearance is that we would not necessarily expect the soft tissue surface to be transformed in the same way as the bone, as it is a much more flexible tissue and will stretch and mould under tension. The challenge is therefore to build a model that can predict how this tissue will react for a given transformation in the underlying bone.

There has been much work done previously approaching this problem from a two-dimensional point of view, generally using lateral cephalograms, for example Power et al. (2005) and Donatsky et al. (2009). However, these are a two-dimensional representation of a three-dimensional object, and are therefore not satisfactory for a full visual representation of the structure, as discussed in Papadopoulos et al. (2002), Westermark et al. (2005) and Caloss et al. (2007). A three-dimensional prediction can give us a much more realistic and complete idea of outcome.

There are now several packages available which attempt to model this soft-tissue reaction in 3-d, such as *Maxilim*® (Medicim, Belgium)and *Dolphin3D*® (Dolphin Imaging, Chatsworth, CA). Almost all of these are based on highly complex finite element models, mathematical models which describe the behaviour of the human tissue (Keeve et al. 1996, Mollemans et al. 2007). However, these models are computationally demanding and there is a shortage of studies in the literature reporting the accuracy of their prediction.

One other approach which has been presented and does not use finite elements is that of Xia et al. (2001). Their method allows for multiple-direction displacement of the soft-tissue in the chin area of the face, and for either skin advancement or setback of the tissue outwith this region, according to pre-defined bone:soft-tissue movement ratios. This method has the advantage of real-time visualisation on a conventional personal-computer system, however its accuracy has so far only been

visually examined. Clearly a quantitative validation is highly important to ensure the accuracy of prediction.

Similarly to Xia et al, our approach also allows different regions of the face to be modelled distinctly without the use of complex finite elements theory. As there can be a shift in either the maxilla, the mandible or both, the method allows different regions of the face to deform individually, based on the changes seen in the underlying bones. All patients in our study had undergone either an advancement or setback of either bone without any significant lateral movement, so we expected the biggest displacement to occur in the frontal region of the face. We therefore wished to determine a model which decreased the strength of deformation applied when moving further away from this central region.

The chapter proceeds as follows. Section 7.2 describes the model preparation, namely the way in which the different facial regions were defined and the method in which the change in bone position was determined. The use of principal curves to index the surface is also detailed. The deformation model applied for prediction across all regions is then described in Section 7.3 and some illustrative results are presented in Section 7.4, before a final discussion in Section 7.5.

7.2 Data Preparation

Determining the Bone Movement

Of course, in predicting the outcome of orthognathic surgery the aim would be to build a model that uses only the pre-surgery image (both bone and soft tissue) along with knowledge of the planned procedure in order to estimate the outcome. However, there is often a discrepancy between the change that the surgeon plans to carry out and the actual change that occurs, as discussed in Bryan and Hunt (1993) and Chabanas et al. (2004), amongst others. This is a rather unpredictable phenomenon, due to the different responses of the human body to these procedures, and it is unrealistic that this can be modelled with a small data set such as ours.

Therefore, as an exploratory method, we wish to use knowledge of what the actual change in the bone has been as opposed to the planned change in order to predict what the effect will be on the soft tissue of the patient.

To investigate the extent to which this could be predicted, data were available from 13 patients who had undergone an orthognathic procedure. For each patient there were pre- and post-surgical images of both the bone and soft-tissue surfaces, and these were initially aligned through the iterative closest point (ICP) tool in *MeshLab*® (Visual Computing Laboratory, Pisa, Italy), an open-source tool developed with the support of the 3D-CoForm project. The ICP procedure is a means of deriving the optimal rigid-body alignment of two surfaces, and is based on the method of Pulli (1999). As it was unlikely that the upper area of the face would have been affected by surgery it should therefore still match up well with the pre-surgery image. The matching was therefore performed based solely on a forehead patch from the post-surgical scan. This alignment of the images allowed for a clear visualisation of the changes that had taken place in the bone position. The aligned soft-tissue surfaces of each patient can be seen in the upper rows of Figures 7.8 - 7.12.

As mentioned previously, either the maxilla, the mandible or both can be repositioned during the surgery, depending on the abnormality present. Therefore it was necessary to look at the shift, if any, in each of these bones separately. The transformation undergone by each bone was found in a similar manner to that used for the surface alignment i.e. a small patch from the front of the pre-surgical maxilla was matched by ICP to the post-surgical image. As before, the two surfaces were roughly aligned in *MeshLab* based on four manually placed landmark points, before the match was iteratively improved until convergence. A point-to-point ordinary Procrustes registration of the original and aligned patches was then used to determine the translation, γ_u, and rotation, Γ_u, undergone by the upper jaw during surgery. This process was then repeated for the mandible in order to obtain γ_l and Γ_l, the parameters for the translation and rotation seen in the lower jaw. These transformations then provide the natural starting point in attempting to model any changes in the soft tissue.

Defining the Prediction Regions

Prior to modelling the reaction to these bone changes, however, it was necessary to define the areas of the soft tissue that would be expected to be affected by the surgery. While it would not be expected that the entire soft tissue surface is altered by a localised change in the bone, it was important to ensure that the area was large enough to include any areas that were affected. To investigate this, the differences between the pre- and post-surgical soft tissue surfaces were examined for each case in order to give a impression of the extent of the area affected. For each point on the pre-surgical surface, the distance to the closest point on the post-surgical surface was calculated. A visual representation of these distances can be seen in the second rows of Figures 7.8-7.12, from which it is clear that the majority of the changes had occurred in the chin and lower cheek regions.

Each patient image had been annotated with a set of landmark points (seen in Figure 7.1), and these were used to define two regions on the soft tissue surface - one corresponding to the upper jaw (which we would expect to be affected by the transformation of the maxilla), and one to the lower jaw (affected by the mandible). In order to ensure a smooth joining of these two areas, and therefore a plausible representation of the post-surgical face, the two regions were chosen so that there was an overlap between them.

To define these regions, a standard distance d was firstly defined for each patient as the distance between the landmarks at the tip of the chin, pg, and the base of the nose, sn, i.e.

$$d = \|pg - sn\|. \tag{7.1}$$

All points within a given multiple of this distance from the landmark ls, which was taken to be a central point on the face, were then considered as points that may be affected by the surgical changes. However, as it was seen that there was not any significant facial change above the cheek area, the landmark representing the outer corner of the left eye, namely exL, was taken as an upper bound for the prediction region. Therefore, given the set of all surface points p, the overall region R where

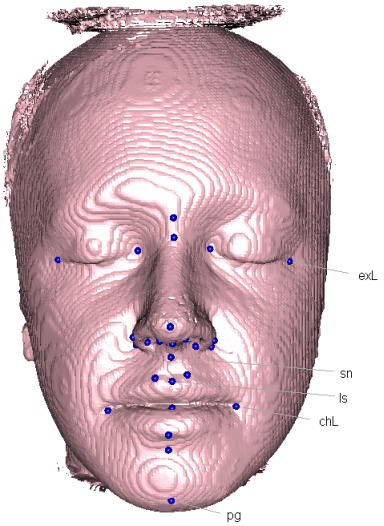

FIGURE 7.1: The thirty landmarks marked on each soft tissue image, including pogonion (pg), subnasale (sn), labiale superius (ls) and left exocanthion (exL).

an effect was expected to be seen is given by

$$R = \{p_i : \|p_i - ls\| < d \ \ \wedge \ \ p_{iy} < exL_y\}, \tag{7.2}$$

where p_y and exL_y are the coordinates denoting the vertical position of these points.

This region was then split into an upper and lower prediction region, which indicated the areas expected to be affected by the mandible and the maxilla respectively. The midline of the lips (taken to be a horizontal plane at chL) was the natural place for this region to be split, however the boundary was extended slightly beyond this limit for each region by an amount equal to $\|sn - ls\|$. This ensured an overlap between the two regions, meaning that a smooth merge between the two transformations could be achieved. An example of the prediction regions can be seen in Figure 7.2.

As it was thought that the movement of each point within these regions could be modelled as a function of the change in the underlying bone, a starting point would be to assume a movement identical to that of the bone. It could then be investigated how much this transformation should be dampened due to the flexibility of the soft tissue. To allow us to control the extent to which the transformation was being applied at each point, the upper region R_u was firstly rotated and moved in exactly the same manner as the maxilla had been to give the fully transformed surface, R_u^*

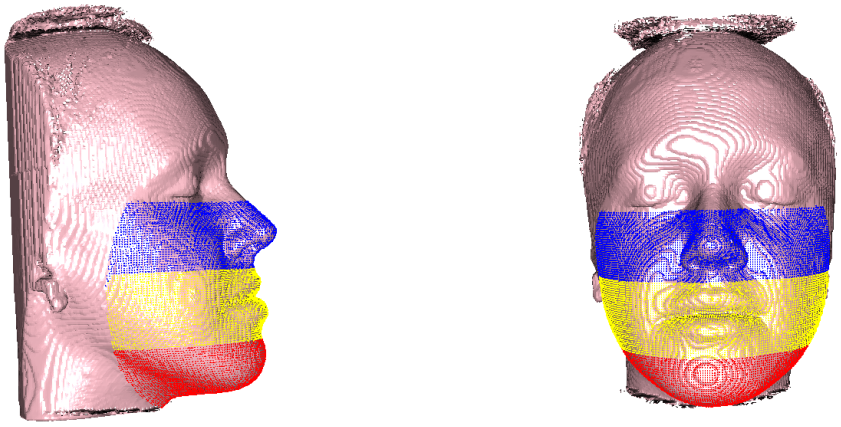

FIGURE 7.2: The three prediction regions considered - upper jaw (blue), lower jaw (red) and the overlap of the two (yellow).

i.e.

$$R_u^* = R_u \Gamma_u + \mathbf{1}\gamma_u^T. \tag{7.3}$$

The vectors V_u which carried each point from its original position to its new transformed position could then be calculated as

$$V_u = R_u^* - R_u. \tag{7.4}$$

The vectors of transformation for the lower region, V_l, were then found in the same manner by using γ_l and Γ_l, the rotation and translation that had occurred in the mandible. It was hoped that scaling parameters for these vectors could then be modelled in order to find the degree of transformation required for an optimal prediction of the post-surgical surface.

Assigning the Weights

In predicting these scale parameters, it was not expected that the effects of the surgery would be seen to the same extent over the different areas of the soft-tissue surface. As all bone movements seen in our patients were in the anterior-posterior direction, we wished to allow the deformation occurring in the soft-tissue to be strongest at the front of the face (i.e. in the direction of change). It was hoped to

do this by by assigning weights to all points, so that the weights decreased as the distance of the point from this area of maximal change increased. It was required that the weighting system used would allocate a value between 0 and 1 to each point, so that for points with weight 1 the transformation would be fully applied, while points with weight 0 would remain in their original position.

Determining these weights required the use of principal curves. As was described in Chapter 4, these are self-consistent smooth curves which pass through the 'middle' of a data cloud, as introduced by Hastie and Stuetzle (1989). A useful application of these curves is that they can be used to trace along the surface of an object, allowing us to index positions on the surface while disregarding those which are not. In order to determine the weights for the prediction region, the following procedure was carried out for both the upper and lower sections separately.

1. Firstly, from a combination of the landmark points and the distances between them, specify four corners c_i $(i = 1, \ldots, 4)$ which define a central patch on this region. (Fig. 7.3)

2. Find all points q which lie within this patch. This is done by means of a projection of r, the set of all points within the region, onto 2-d, through the technique described in Section 4.1.

3. Assign a value of 1 to all points within this patch, to indicate that maximal change is expected to occur in this central area.

4. Calculate a local axes system, say $\vec{a_1}, \vec{a_2}$ and $\vec{a_3}$, where $\vec{a_1}$ is the direction running from the patch centre to its top left corner i.e.

$$\vec{a_1} = c_1 - \sum_{i=1}^{n} \frac{q_i}{n}. \tag{7.5}$$

   Define $\vec{a_3}$ as the normal direction to the surface (i.e. the third component direction of a principal components analysis of the patch), and subsequently calculate the remaining axis as $\vec{a_2} = \vec{a_1} \times \vec{a_3}$. $\vec{a_3}$ can then be refined as $\vec{a_1} \times \vec{a_2}$, to ensure all three axes are orthogonal.

5. Find the strip of points s running from the outer edge of the central patch to the boundary of the prediction region in the direction of a_1. These points will be within a small threshold distance δ of $\vec{a_2}$, and in a positive direction when projected onto $\vec{a_1}$ i.e. $s = \{r_i : r_i^T \vec{a_2} \leq \delta \ \wedge \ r_i^T \vec{a_1} > 0 \ \wedge \ r_i \notin p_i\}$.

6. Fit a principal curve to the strip s and find the arc length l_i at each point s_i.

7. Scale this set of arc lengths so that $0 \leq l_i \leq 1 \ \forall s_i \in s$, and assign these scaled arc lengths as the weights w, i.e.

$$w_i = \frac{l_i}{max_i(l)} \tag{7.6}$$

8. Assign weights in different directions by incrementally rotating the axis system around a_3 by 360°, repeating steps 5 to 7 after each rotation.

Scaling the arc lengths to be on a (0, 1) scale allows us to see how far along the surface each point is, as a proportion of the distance to the region boundary in that particular direction. As we expect the effects to fade as the distance from the area of maximal change increases, this is therefore an indicator of how strongly we expect it to be affected by the movement of the bone. Taking a sufficient number of transects through the surface (by using small rotational increments) assigns weights to a large proportion of the points on the surface. For any point not included, we can simply find the weighted point that lies closest and assign the corresponding weight, resulting in a value for all points within the region. An example of the calculated weights can be seen in Figure 7.4.

7.3 The Model

It was then required to use the weights that been devised to calculate how quickly the effects of the surgery fade as you move along the surface, and consequently how strongly the transformation should be applied at a given point.

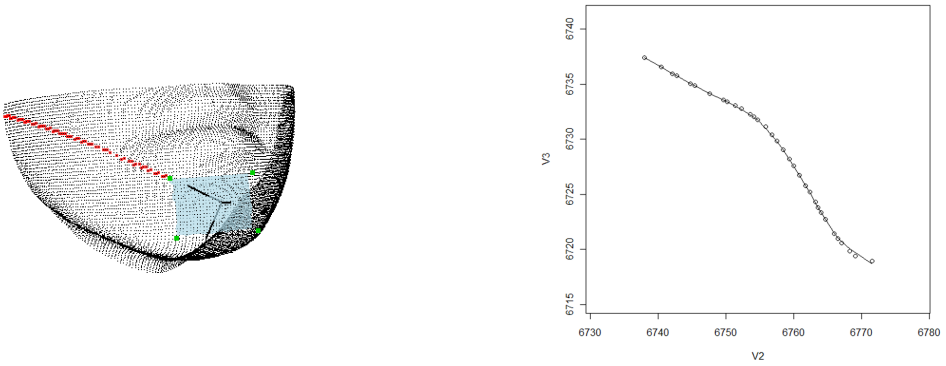

FIGURE 7.3: *Left:* The upper prediction region, with central patch shown in blue. The axes can be seen along with the strip of points found in that particular direction. *Right:* The principal curve that was fitted to this set of points, represented in two dimensions.

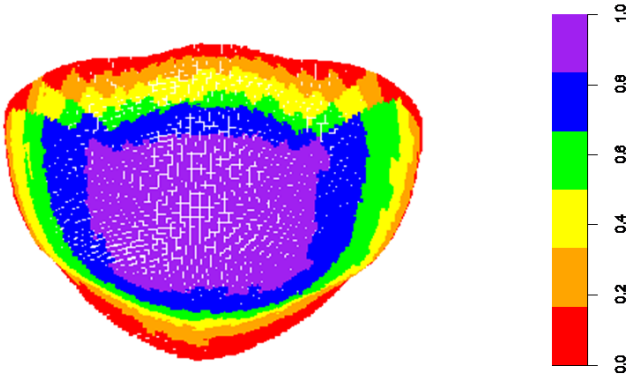

FIGURE 7.4: An example of the weights calculated for the lower prediction region.

For the upper and lower prediction regions separately an exponential decay model with two parameters was used, with the first parameter α determining the rate at which the effects fade. The function chosen ensured that the scaling applied to these transformation vectors would go from 1 to 0 as the weights decreased within the region, but the value of α would determine how quickly this occurred. The shape of this function can be seen in the left panel of Figure 7.5.

It was thought, however, that even at the most central region the soft tissue may not move as much of the bone does, as it will mould under tension. Therefore the second parameter β was included to scale the overall transformation taking place.

The equation for a point r_i in the upper region, R_u, could then be represented by

$$P_{ui} = \beta_u \left(\frac{1 - e^{\frac{-w_{ui}}{2\alpha_u^2}}}{1 - e^{\frac{-1}{2\alpha_u^2}}} \right) V_{ui}. \tag{7.7}$$

The corresponding equation for the lower region, R_l, is then given by

$$P_{li} = \beta_l \left(\frac{1 - e^{\frac{-w_{li}}{2\alpha_l^2}}}{1 - e^{\frac{-1}{2\alpha_l^2}}} \right) V_{li}, \tag{7.8}$$

where w_u, w_l represent the weights for the upper and lower regions respectively.

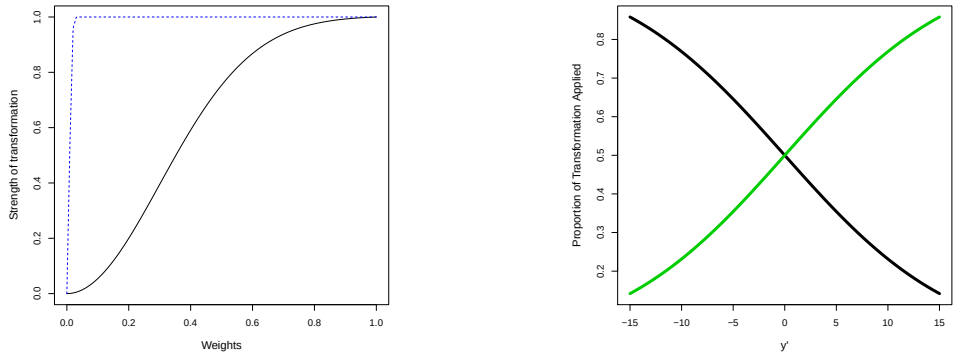

FIGURE 7.5: *Left:* The drop off rate of our exponential function when α=0.1 (black) and α=0.001 (blue). *Right:* The rate of transition between the upper (black) and lower(green) transformations in the overlap region, where the latter is simply 1 minus the former.

It was also required to determine how the surface behaved within $R_o = R_u \cap R_l$, the area where the upper and lower regions overlapped. Within this region, the points had a weight value for both the upper and lower transformations, and a final parameter ψ was introduced to determine how the two were best combined. To do so, the vertical position of the points, y', were also taken into account. Each y'_i was taken to be the vertical position of the point relative to the midline of the lip, and was used as an indicator of the distance that each point was from each of the region centres. The function used to smooth between the transitions was

$$\left(\frac{e^{\psi y'_i}}{1 + e^{\psi y'_i}} \right). \tag{7.9}$$

The shape of this function for a particular ψ can be seen in the right panel of Figure 7.5. Altering the parameter ψ determines how gradual or otherwise the transition between the two transformation is.

Combining these functions, the overall model for prediction for a point r_i can therefore be given by

$$P_i = \begin{cases} P_{ui}, & r_i \in R_u \backslash R_o \\ P_{li}, & r_i \in R_l \backslash R_o \\ \left(\frac{e^{\psi y_i'}}{1+e^{\psi y_i'}}\right) P_{ui} + \left(1 - \frac{e^{\psi y_i'}}{1+e^{\psi y_i'}}\right) P_{li}, & r_i \in R_o \end{cases} \tag{7.10}$$

7.3.1 Refining the Model

The final model then consisted of five parameters, namely α_u, β_u, α_l, β_l, and ψ, for which the values required for an optimal prediction must be deduced. As the post-operative data was available, it was simply required to find the values that minimised the distances between the prediction and the actual post-surgical surface.

However, upon a first optimisation it was found that the values of α_u and α_l, the parameters controlling the drop-off rates in the upper and lower regions respectively, were occasionally chosen to be very small. This meant that the application of the transformation would fade off very rapidly towards the edge of the prediction region.

This caused a slight ridge in the tissue, an example of which can be seen at the base of the neck in Figure 7.6, giving a predicted surface that was not smooth and therefore did not represent a plausible human face. It was therefore decided that a condition should be included in the model to ensure that these parameters had to take a value of at least 0.1. This figure was chosen as it was found to give a reasonably smooth surface in all cases and would therefore gives us a realistic representation of the post-surgical face. Prior to this constraint being applied, the smallest α that had been found for either region was 0.001. The left panel of Figure 7.5 shows the improvement in the smoothness of the exponential function when this value was increased to an α_u of 0.1.

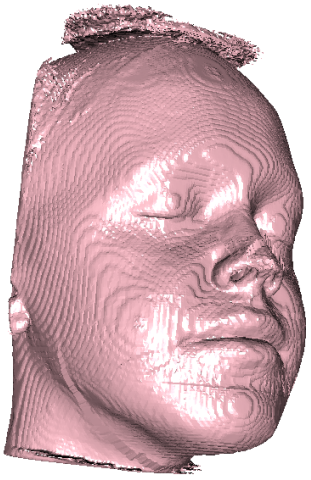

FIGURE 7.6: Slight ridges on predicted surface, in this case due to a small α_l and ψ.

It was also found necessary to ensure that ψ was sufficient so that the function controlling the transition between the two transformations would run from 0 to 1 within the range of y' values seen in the overlap. This was previously not always the case, as can be seen in the right panel of Figure 7.5. If the values did not span this entire range, the edges of the overlap would not merge back into the rest of the surface in a satisfactory manner. Looking at Figure 7.6, an example of this is apparent in the slight ridges occuring above and below the lips. By setting

$$\left(\frac{e^{\psi y'_{max}}}{1+e^{\psi y'_{max}}}\right) < 0.01, \tag{7.11}$$

where y'_{max} is the maximum value of all the y' values within the overlap, and solving for ψ, it was deduced that constraining ψ to be less than

$$-\frac{4.6}{y'_{max}} \tag{7.12}$$

would ensure that the function covered the whole range (0, 1) in all cases and gave a reasonable prediction in this area.

There was one further issue which arose with the ψ parameter. Where there had been large shift in one bone compared to the other, the value chosen for ψ occasionally did not give a gradual enough transition to ensure a smooth changeover within R_o. An

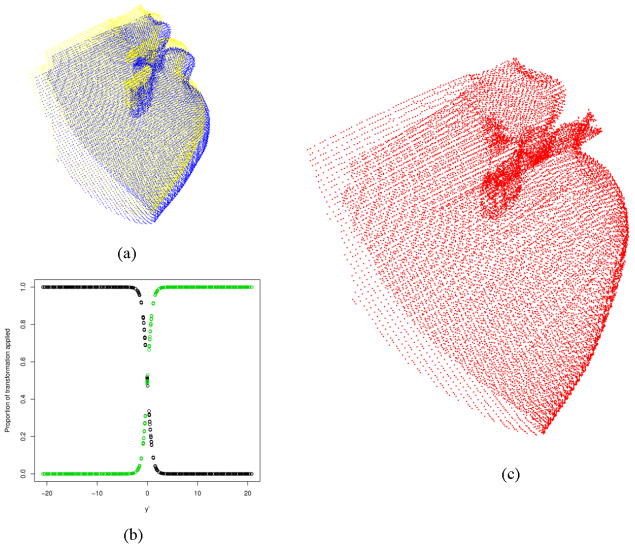

FIGURE 7.7: (a) The upper (blue) and lower (yellow) transformations applied separately to our prediction region (b) The rate of transition between the two that was found by our original unconstrained optimisation (c) The resulting surface, with disjoint region in the overlap

example can be seen in Figure 7.7, where the ψ value of -1.99 that was found for this particular case led to a sharp change between the two transformations. There was a reasonably large difference between the two transformations in this case which, combined with the fast rate of transition selected, led to a disjointed surface. It was found through graphical exploration over several values of y'_{max}, however, that constraining the parameter to be greater than

$$-\frac{20}{y'_{max}} \tag{7.13}$$

ensured a more gradual change between the two transformations in all cases.

7.4 Results

7.4.1 Accuracy of the Model

Individual Parameter Values

As a first step the model was fitted for each patient individually, with the parameters for these minimal distances being deduced using the *optim* function in R . The errors arising from the optimisations can be seen in Table 7.1, for both the unconstrained and constrained parameters.

Patient	Original Optimisation	Constrained optimisation	Global
1	0.944	0.957	0.974
2	0.956	0.981	0.996
3	1.618	1.664	1.762
4	1.368	1.367	1.423
5	1.458	1.453	1.639
6	1.281	1.354	1.479
7	2.196	2.261	2.262
8	0.790	0.858	0.939
9	1.375	1.391	1.469
10	1.284	1.281	1.397
11	0.971	0.988	1.088
12	1.221	1.222	1.447
13	1.310	1.352	1.423

TABLE 7.1: Average distances (mm^2) between the prediction and target surfaces, for the unconstrained, constrained and overall global optimisations.

As the original optimisations were based purely on the distances between the two surfaces, naturally the placing of these constraints on the parameters should increase the errors between the prediction and the target surface. Surprisingly though, a slight decrease was actually seen in the errors for three of our patients. However, looking at Table 7.1, this difference in errors was less than 0.006mm for each of these cases and the parameter values were very similar, suggesting that this was simply down to slight differences in the convergence of the optimisation function.

For the majority of cases (where the errors increased as expected) the differences were relatively small, with the average error over all cases only increasing by 0.03mm (from an original value of 1.29mm). As the constrained parameters resulted in an improved visualisation, a factor which was of ultimate importance given the purpose of the model, this was deemed an acceptable increase and the constrained model was adopted. The average errors for this model were all within 1.7mm with the exception of Patient 7, who had a larger mean error of 2.3mm.

The predicted surfaces, shown alongside their post-surgical target, can be seen in the third rows of Figures 7.8 - 7.12. The corresponding distribution of errors for the surfaces can be seen in the bottom row of these same figures, where negative values indicate a predicted surface that lay behind the post-surgical target surface and vice versa. The colour maps show a tendency for higher errors to be present around the lips in most cases, something that has also been reported as an issue in other studies, e.g. Mollemans et al. (2007) and Popat and Richmond (2010). The errors in this area were particularly high in Patient 3, shown in the right hand column of Figure 7.8. This was due to the patient's natural resting position having gone from an open to a closed mouth as a result of the surgery, a change which proved very difficult to model.

Examining the results for Patient 4 (the left hand column of Figure 7.9), the majority of the errors found in this case seem to be gathered around the neck area. As can be seen by comparison of the pre- and post-surgical surfaces in the upper panel, this was actually an area of high error prior to the model being implemented, and seems to be due to the differences in the patient's neck position during image capture. This neck position affects the shape of the surface but is unrelated to the surgery, and therefore can not be modelled based on the change in bone, an issue that was also found by Chabanas et al. (2004).

While this difference in neck position was fairly obvious and led to a greatly increased error in this particular patient, it was thought that it could also be having an (albeit smaller) effect in the other cases too. To investigate this, the optimal parameters and overall errors were again calculated, but over the area on the front of the face only.

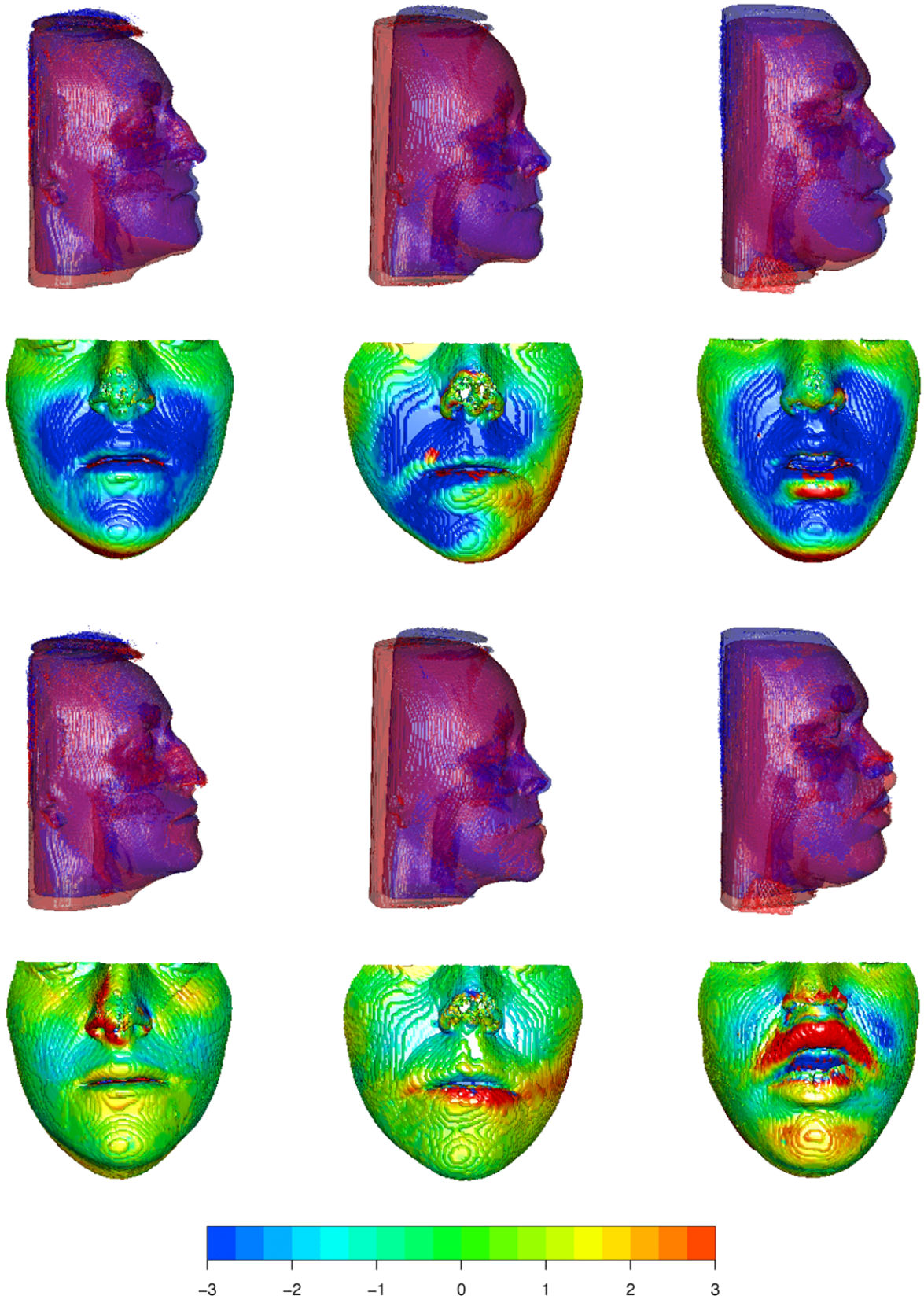

FIGURE 7.8: For patients 1-3, from top to bottom: The pre-surgical (red) and post-surgical (blue) surfaces; the differences between the pre- and post-surgical surfaces (mm); the predicted (red) and post-surgical (blue) surfaces; the differences between the predicted and post-surgical surfaces (mm).

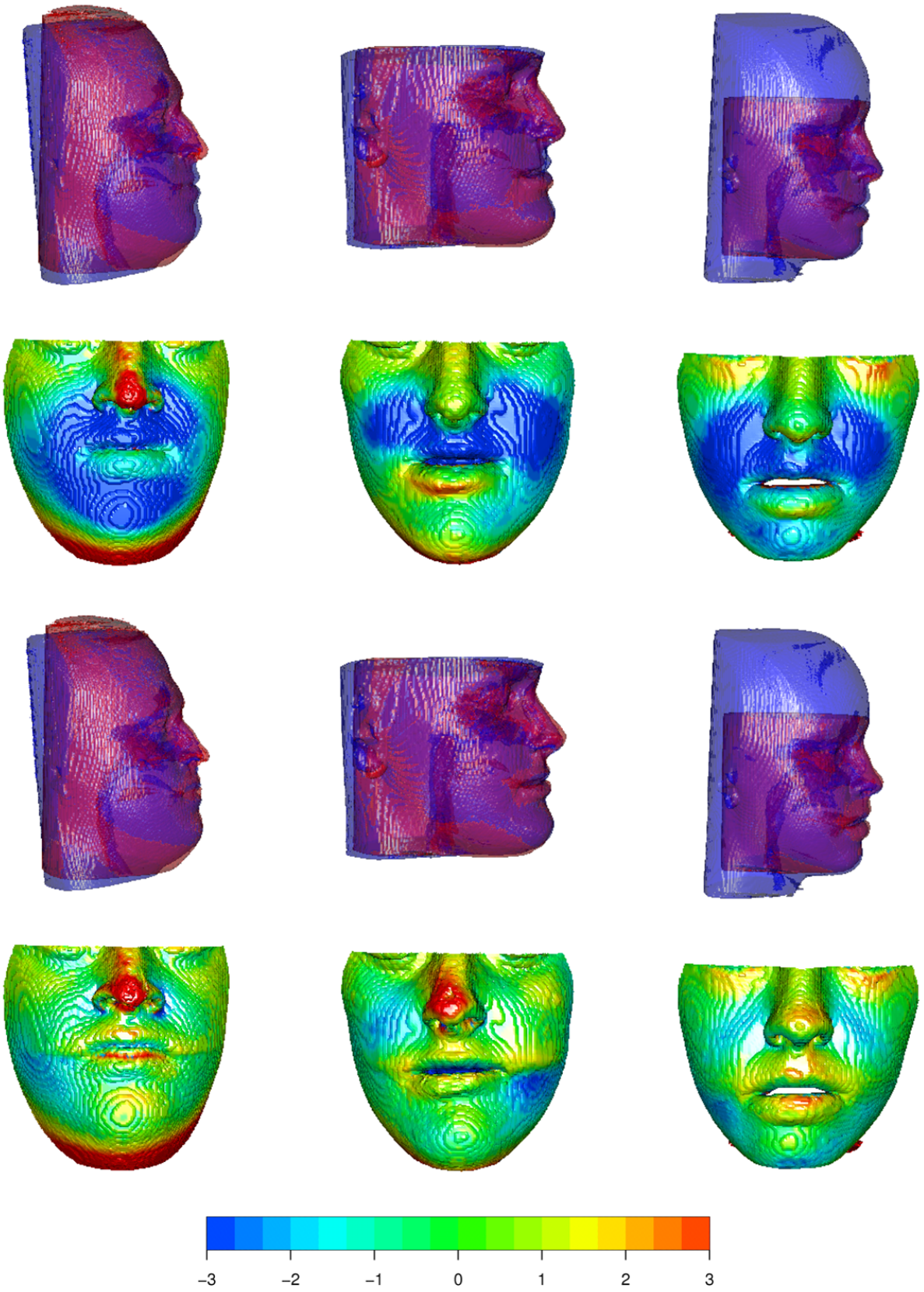

FIGURE 7.9: For patients 4-6, from top to bottom: The pre-surgical (red) and post-surgical (blue) surfaces; the differences between the pre- and post-surgical surfaces (mm); the predicted (red) and post-surgical (blue) surfaces; the differences between the predicted and post-surgical surfaces (mm).

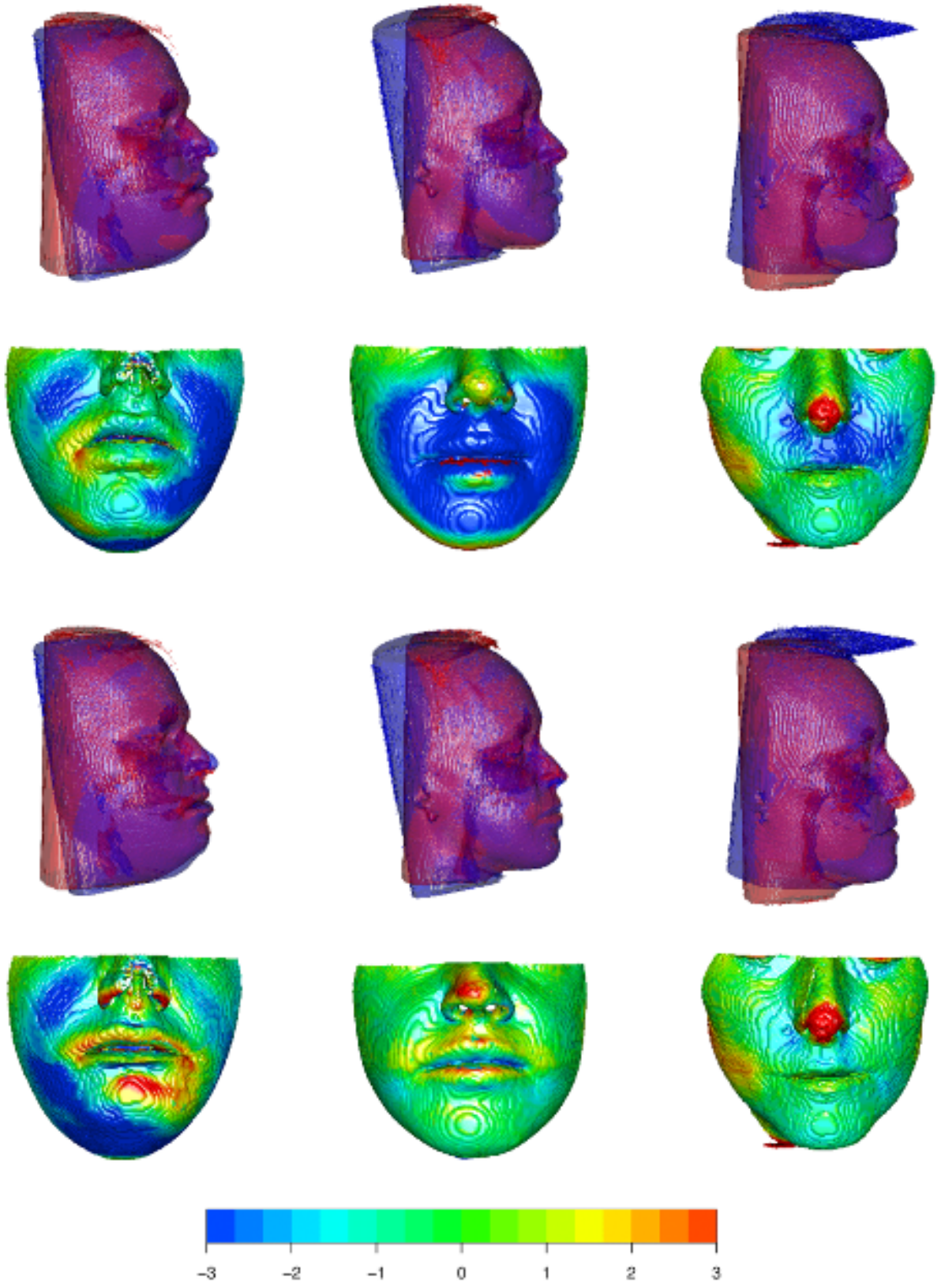

FIGURE 7.10: For patients 7-9, from top to bottom: The pre-surgical (red) and post-surgical (blue) surfaces; the differences between the pre- and post-surgical surfaces (mm); the predicted (red) and post-surgical (blue) surfaces; the differences between the predicted and post-surgical surfaces (mm).

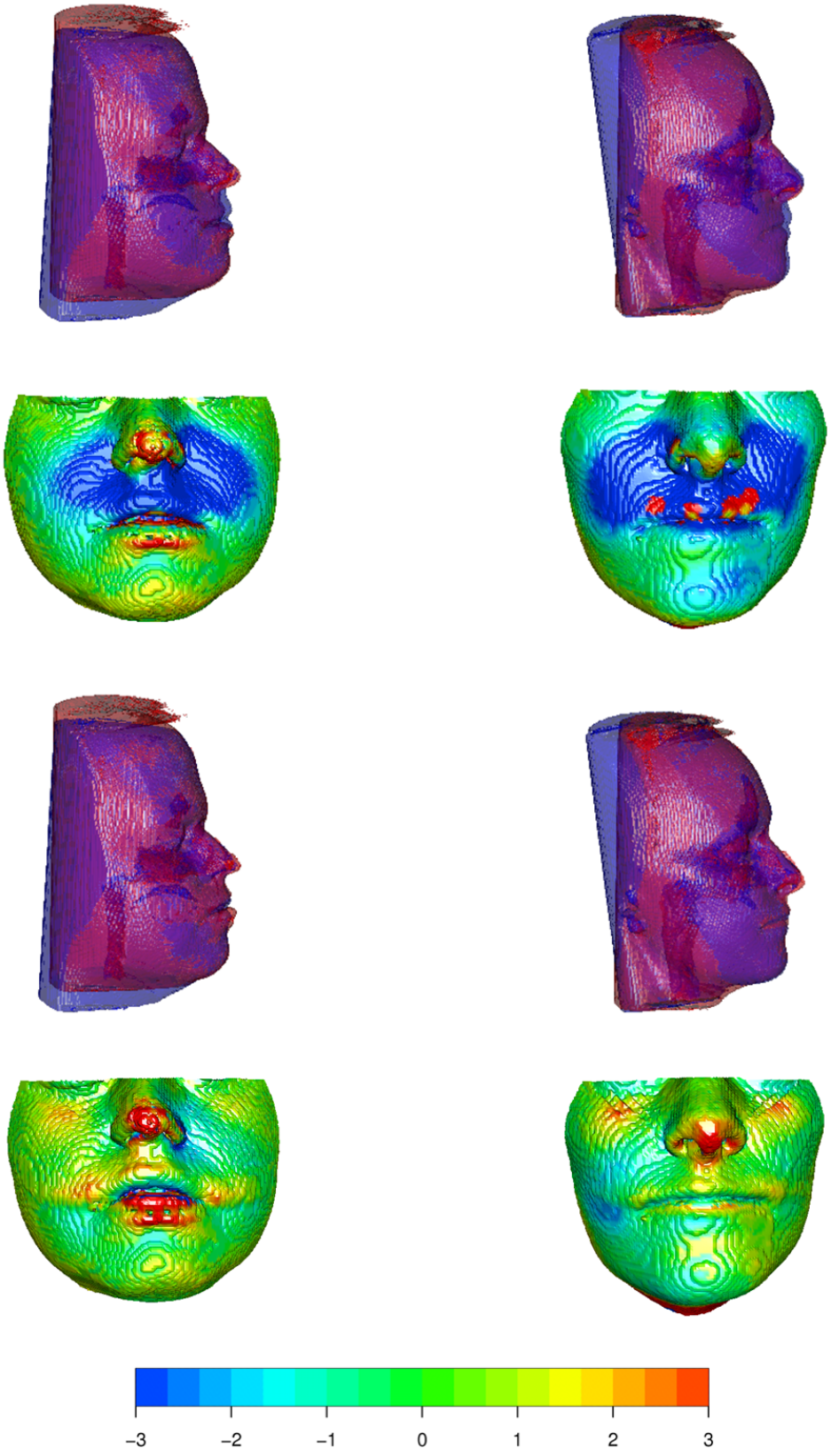

FIGURE 7.11: For patients 10-11, from top to bottom: The pre-surgical (red) and post-surgical (blue) surfaces; the differences between the pre- and post-surgical surfaces (mm); the predicted (red) and post-surgical (blue) surfaces; the differences between the predicted and post-surgical surfaces (mm).

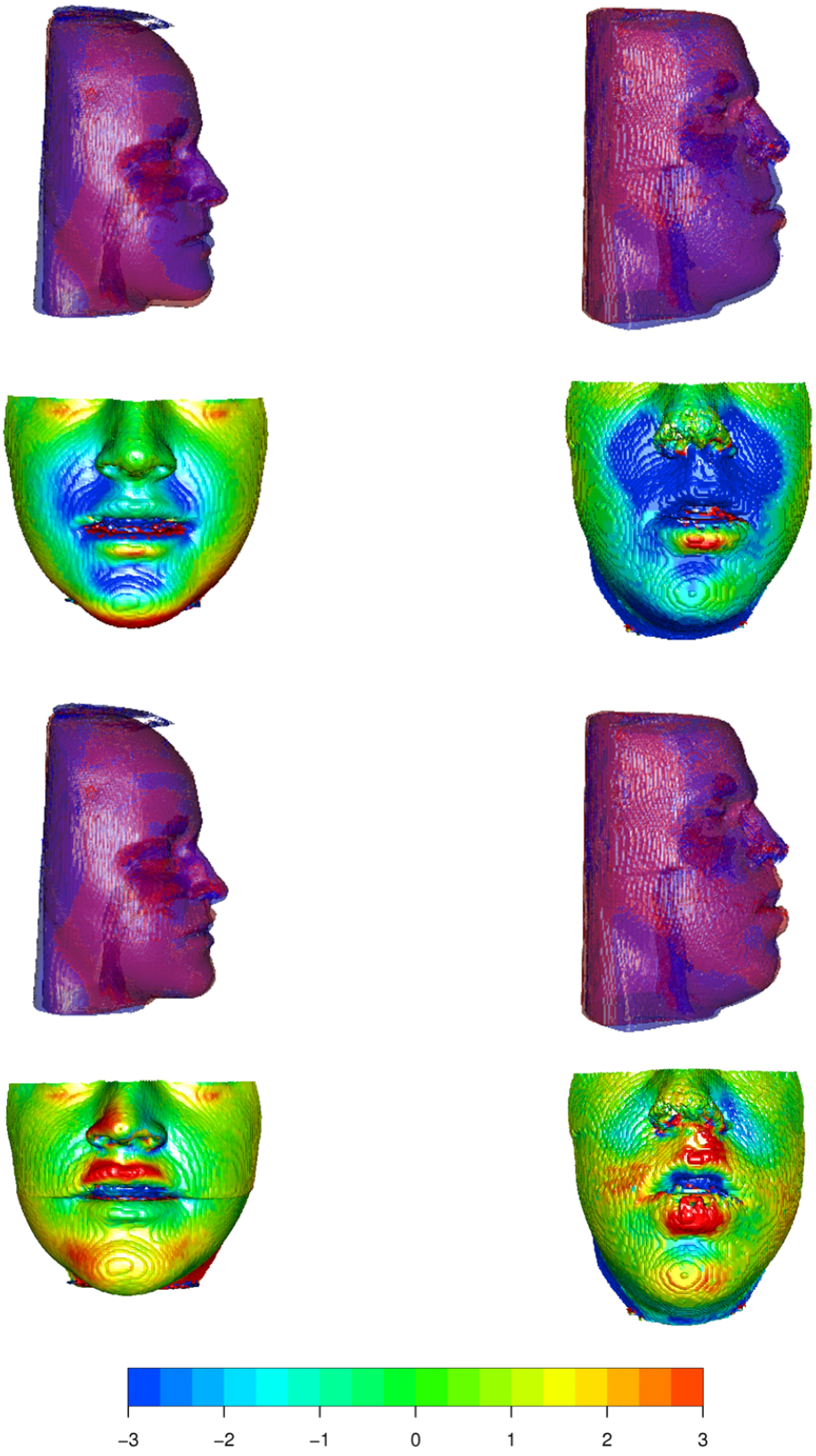

FIGURE 7.12: For patients 12-13, from top to bottom: The pre-surgical (red) and post-surgical (blue) surfaces; the differences between the pre- and post-surgical surfaces (mm); the predicted (red) and post-surgical (blue) surfaces; the differences between the predicted and post-surgical surfaces (mm).

This was done by determining where the jawline lay and including only points that lay above this boundary in the optimisation process. The methods used to do this were described in Chapter 4. However, calculating the predictions solely within this region was found to only slightly decrease the average errors between the predicted and post-surgical surfaces by 0.06mm. This was a relatively small difference and the errors were in fact higher than their original values in 6 of the 13 cases, suggesting that the neck position was not having an overly profound effect in the majority of the cases. Its removal was therefore deemed unnecessary.

Another issue raised upon examination of the error maps was the higher errors that were often seen within the nose area. Some of these errors, e.g. those of Patient 10, were in fact down to a section of the nose tip missing in either the pre- or post-surgical image. This can occasionally occur during the segmentation process, when the bone and soft tissue are being extracted from the original file, and could therefore also be seen in the original pre-to-post error maps (Fig. 7.9). This was not an artefact of the model and therefore not of concern. In Patient 1 however, for example, these errors were induced by the deformation of the soft tissue in the model. This is perhaps a somewhat unsurprising issue, as under the current model the nose is treated in the same manner as the rest of the prediction region; weights are assigned to the points and are then used to transform them according to the change in the maxilla (as the nose is within the upper prediction region). We would not necessarily expect this tissue to behave in the same way however, as its underlying bone had remained stationary.

In an attempt to deal with this issue, the region was modelled without including the nose, with the excluded nose area subsequently being warped to the resulting surface by means of a Thin Plate Spline. The nose boundary was defined using four of the landmark points, before the area was removed by means of a projection onto 2-d. Calculating the predictions by this more complex method, however, was found to give very similar errors to our original approach, with the overall average error actually increasing by 0.004mm and an improvement only seen in 6 of the 13 patients. Due to the extra time and calculation required to deal with the nose in this manner, it was therefore not deemed a worthwhile extension of the model.

Global Parameter Values

Of course, in order to use this type of model to predict in unseen cases a single set of optimal input parameters would need to be determined. Therefore, as well as assessing the levels of accuracy obtained for each case, we can also examine the variation seen in the parameter values in order to get an idea of how accurate a future prediction model might be. Similar parameters values across all cases would suggest that one model of this type could be sufficient to capture the variation across all cases reasonably well. The parameters previously found for each of the individual patients can be seen in Figure 7.13, and indicate possible ranges within which we would perhaps expect an optimal global set of parameters, calculated over a larger number of cases, to lie.

As can be seen in the upper left-hand panel of this figure, the values chosen for α_l often tended to be very small. However, as is illustrated in the left-hand panel of Figure 7.5, even with a small value of 0.1 the drop off is fairly gradual. The larger area of this lower region and less pronounced changes that were seen within it therefore meant that these small values of α_l were necessary in order to describe the transformation adequately.

As mentioned previously, the small number of cases available for this study is insufficient to calculate the overall optimal parameters required very accurately, given the huge number of surface points and the complexity of the procedure. However, to examine how the individual predictions were affected when one global set of parameters was applied to all cases, a single best set of parameters was calculated by optimising over all cases simultaneously. The errors arising from this overall set of parameters can be seen in the third column of Table 7.1. While these are generally slightly higher than the previous optimisations, where the parameters were tailored to each individual case, there did not seem to be a marked increase in the errors, with the overall average only rising by 0.09mm to 1.41mm. The visual impressions were also similar to the previous results. This suggests that a single model could possibly provide fairly good results across other unseen cases.

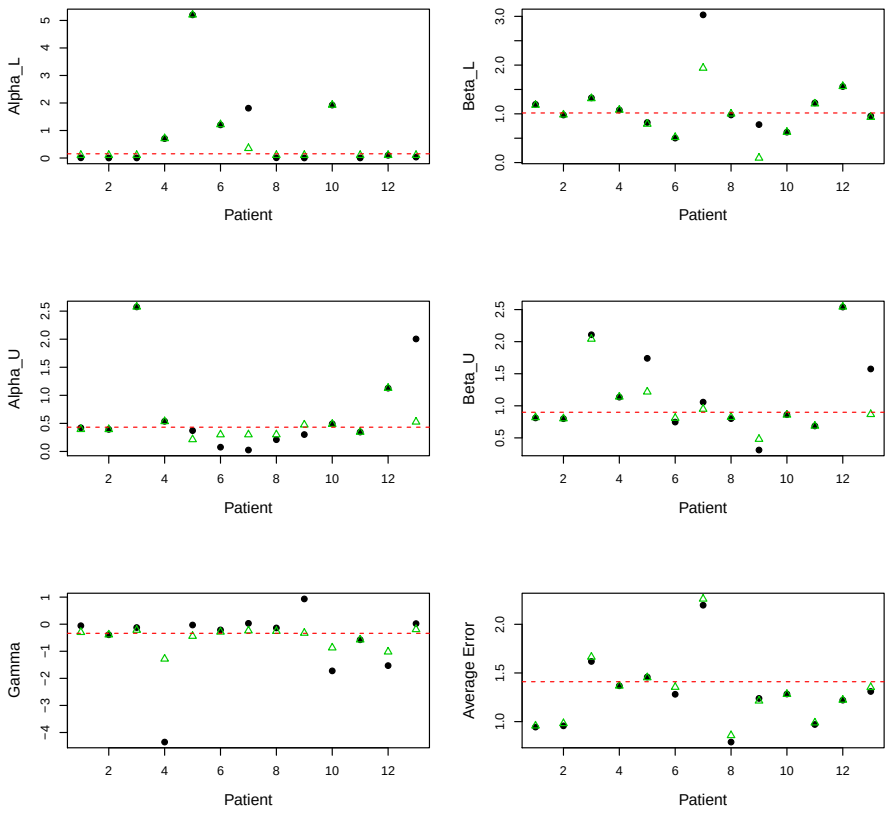

FIGURE 7.13: The optimum parameters and the associated average error for each case, with black representing the original parameters and green representing the constrained parameters. The red lines illustrate where the values lay for the overall global optimisation.

7.4.2 Comparison to Results from the Literature

As was mentioned in Section 7.1, there is a shortage of quantitative assessments available for the existing 3-d prediction methods, especially those based on a reasonable number of cases. An exception, however, is the study by Mollemans et al. (2007). This study compared four different computational methods for the prediction of soft tissue deformation, namely a Finite Element Model (FEM), a non-linear FEM, a Mass Spring Model (MSM) and a Mass Tensor Model (MTM). Each method requires all soft tissue points to be defined as either 'join points' if they are attached to the skull or 'free points' otherwise. The join points are shifted in the same manner as the corresponding section of bone, whereas the movement of the free points is determined by the elastic force that exists in these points. The accuracy and speed

of each method was assessed using a dataset consisting of ten patients, all of whom had undergone an orthognathic procedure.

Of the methods, the MTM and FEM were found to be the most accurate. Both methods gave very similar error distributions, with the average 50th percentile being equal to 0.6mm and the 90th percentile close to 1.5mm. Comparing this to results from our model, the corresponding values for were decidedly higher, with averages of 0.9mm and 2.5mm respectively. An advantage of our method, however, is the speed of computation. In the study by Mollemans et al. (2007), the MTM was found to be the fastest of the models with an average simulation time of 10 seconds, where Chabanas et al. (2004) reported a time of 3 seconds for small deformations and up to 3 minutes for larger deformations. This is in contrast to our method which, given a set of parameters, only requires around 0.7 seconds for the calculation, regardless of the extent of deformation.

Two further quantitative assessments were carried out by Chabanas et al. (2004) and Zachow et al. (2004), again using finite element modelling approaches for prediction of the deformation. Chabanas et al. examined the predictions for 3 patients, finding them all to be within 1 - 1.5mm from the actual post-surgical surface on average, values which are highly comparable to our mean of 1.32mm. No information was presented on the variation of the errors in this case.

The study by Zachow et al. involved only one patient, with the average error being found to be 1.32mm. This mean error was exactly the same as that of our model and the differences in standard deviation were negligible, with values of 1.29mm and 1.30mm respectively. The medians of the errors were also similar, at 0.97mm for the FEM approach and 0.93mm for our model. Comparing their case to each of our patients individually, the average error was less than that found by Zachow et al. in 6 of 13 cases, while in the remaining 7 it was greater. For their unique patient, in 52% of the skin surface the prediction error was below 1 mm, for 78% it was below 2 mm, and for 10% of the surface the deviation was larger than 3mm. Of our 13 patients, 8 had a greater percentage of errors that lay within the 1mm limit, and 11 had a greater percentage of errors that lay within the 2mm limit. Combining all

sets of errors together, in 57% of points the prediction error was below 1 mm, for 84% below 2 mm, and for only 8% of the surface the deviation was larger than 3mm, suggesting a higher accuracy overall than was achieved for their individual patient. However, as mentioned previously, these conclusions should be interpreted with care due to the sample size of one.

7.5 Concluding Remarks on Soft Tissue Deformation

It was felt that our model was generally able to fit the original surfaces to their post-surgical targets reasonably well. It gave the flexibility of allowing different regions of the face to be affected by different bone transformations whilst still maintaining the smoothness of the surface, therefore giving us an outcome which was both close to the target surface and was a plausible representation of the human face. The lips however could be a problem area and tended to show higher errors than other areas of the face, an issue also found by other methods. This was particularly true in cases where the natural resting position of the mouth, which was the position in which the patients were captured, went from being open to closed as a result of the surgery. This proved to be very difficult to model.

While the predictions obtained by our model appear to fit well by visual examination and give a credible representation of the human face, they were not found to be as accurate as those produced by Mollemans et al. (2007) when assessed quantitatively. When comparing to the results of that study, the discrepancy is compounded by the fact that our errors were calculated on the basis of the smallest distance to the target surface at each point. This is in contrast to their method, where the predicted surface was warped to the target in order to find which was the 'true' corresponding point, and the distance to this point used as the error. As discussed in Chabanas et al. (2006), simply taking the distance to the closest point on the other surface can lead to a underestimation of the errors, meaning that the differences between these two sets of results may be of greater magnitude than first thought.

Looking at other studies presented by Zachow et al. (2004) and Chabanas et al. (2004), our results compared more favourably. These studies were based on only a small number of cases (1 and 3 patients respectively) meaning that the results should be interpreted cautiously, however our method appeared to be on a par with their methods in terms accuracy, and perhaps more precise than those reported by Zachow et al.

One advantage to our method is the increased speed of calculation. The calculation itself was sped up from an average of 10.2 seconds for the MTM (the fastest of the models assessed by Mollemans et al.) to an average of 0.7 seconds for our model. The time required for setting up the models must also be taken into consideration however, namely calculating the weights for the lower and upper prediction regions in our case, which took an average of 35 seconds for each region. It should be noted that these weights must only be defined once, to allow fast calculations of the predicted surface for any number of different treatment plans. The time taken to define the 'free' and 'join' points necessary for the MTM was not reported. The calculation time in Chabanas et al. (2004) greatly exceeded ours, and was reported as being anywhere from less than 3 seconds to almost 3 minutes, depending on the extent of the deformation.

As well as the perhaps less biologically informed model, the way in which the bone movement was dealt with could also be a reason for the lower accuracy of our method when compared to Mollemans et al. Both this study and Zachow et al. (2004) used differences between the pre- and post-surgical bone surfaces to find where the actual osteotomies were performed in each bone, whereas in the method described here the displacement seen in the frontal region of each bone was used to attempt to model the transformation in all of the relevant deformation region. Incorporating the osteotomy information into our model could possibly improve the performance.

One disadvantage of our method is its inability to deal with changes to the nasal area of the face. This is not necessarily well dealt with by a FEM approach either however, as suggested by larger errors seen around the nose by Mollemans et al., amongst others. It was found however that excluding the nose during the modelling

and warping it to the final predicted surface did not sufficiently improve the results. This is an area where further work may be necessary to improve the model.

A possible issue with our model is that it has so far only been tested for predicting one particular type of orthognathic surgery. Our data were a homogeneous patient group who had all been treated with a bone change in the anterior-posterior direction, therefore our method may need to be checked for robustness to other types of surgical procedure. For other procedures it would perhaps be required to calculate the direction of maximal displacement and fade the transformation effects from there.

While it was always unlikely that the accuracy of these biomechanically relevant FEM models could be matched by an approach such as this, the predicted surfaces still appear good in comparison to the post-surgical results, and were in fact as accurate as the method of Chabanas et al. (2004) and more accurate than the FEM approach of Zachow et al. (2004) in 6 of our 13 cases. The main advantage behind our system is the speed of calculation, something which is more likely to be of use when absolute precision is not as important as a real-time visualisation, for example in virtual surgery.

Chapter 8

Discussion

With ever improving and more readily available technology, it is likely that the fields of imaging and thus shape analysis will continue to grow. This trend is particularly apparent in the area of medicine, where patient images are being captured on a more and more regular basis - be it for the purposes of diagnosis, treatment planning or the evaluation of surgical outcome. The aim of this research was therefore the development of techniques to allow for the informative analysis of these images.

As with all types of data, the quality of the input must be high in order to achieve a valid set of results, and validation and reproducibility studies therefore play an essential part in the quality control of the data. Shape data can be validated in several ways, for example by a study into the reproducibility of the data or by comparison to a 'gold standard' method. With landmark data in particular, and therefore surface representations derived based on these landmarks, it is essential to ensure the accuracy of the landmarking process. Through previous validation studies as well as those conducted in Chapter 2, the landmarking and image capture systems were found to be satisfactorily reproducible for our data, allowing confidence to be placed in the resulting analyses.

The remainder of this thesis has explored a variety of ways in which three-dimensional anatomical images can be analysed, with focus lying mainly on the assessment of

shape through an appropriate surface representation of the object. These representations are composed of a fairly large number of surface points which correspond across the population and that describe the shape of the surface in a detailed manner. A major advantage of this approach in place of a more traditional, landmark-based analysis is that a far greater quantity of the information in the image is utilised. Landmark configurations tend to be a small set of representative points and define the key areas of a shape, however this leads to a loss of any more localised or small scale features of the image. While this may not be an issue with objects of a very simple form or where only global differences are of interest, in a medical setting it is very often the more subtle features or changes that it is desirable to assess.

Another incentive for the use of surface representations is that they lead to a more informative and detailed interpretation of the results. For example, visualising the set of average facial landmarks for the cleft palate data does not intuitively show us what the face itself would look like. The average surface representation, on the other hand, can provide a much clearer picture. The use of the surface representation also allows for a far more logical interpretation of the analyses, for example it was seen that the modes of variation found from a principal components analysis are easily interpretable, as are the averages faces obtained from the growth model of Chapter 7.

However, while there are some clear advantages to working with surface data in place of landmark points or surface curves alone, there are also some issues that can arise, and many of these have been illustrated by the applications presented here. For example, it was seen in Chapter 4 that it is often necessary to extract an area of an image in order that a subset of the surface can be analysed independently. This reduction can be necessary for many reasons, for example when areas that do not belong to the anatomical surface are included in the image, or when solely a particular area of the surface is to be analysed. In addition to allowing for a more focused analysis by concentrating purely on the area or feature of interest, this also allows for the size of the surface to be reduced. This can be a big advantage, as with continual improvements in technology come higher resolution images and therefore an increased number of data points. An analysis of the entire surface can therefore have very high computational costs.

Chapter 4 discussed several methods for the achievement of this surface reduction, with the required approach varying dependent on the complexity of the area to be extracted. A straightforward method of plane cuts was first detailed, however this would only be useful in loosely defining an area of the surface. For areas of any great complexity a more flexible approach would be required. A method was therefore presented which utilises the connectivity of a set of mesh points in order to track along the surface. This technique was seen to work well for area selection in the case where the outer extremities of that region could be easily defined, and proved a useful means of identifying the internal structures of images captured by CT scan. Finally, an approach to feature identification through curvature was also detailed, with an application shown for the determination of the breast boundary. This approach can usefully allow for the selection of features of a far more complex form.

8.1 Surface Correspondence

Following an extraction of the appropriate area, the number of data points can be further reduced by determining a set of point correspondences across the population. As well as reducing the number of points, the corresponding nature of the resulting points allows for a much more powerful analysis and the use of many existing shape analysis techniques.

As detailed in Chapter 3, a standard approach was utilised to determine the correspondences in the cleft palate data set, namely through the use of Thin Plate Spline warping. This was seen to produce a good representation of the surface in almost all cases, however there were a couple of issues that arose. Firstly, any image for which there was a lack of ear landmarks had to be removed from the dataset, due to the fact that these warps can only be reliably implemented when the landmarks that they are based on give a reasonable coverage of the entire object of interest. A second issue that arose was that of the presence of both open and closed-mouthed images. Clearly by using a standard base mesh to determine the correspondences across all cases, this will result in either images with closed mouths being warped to

an open-mouthed base mesh, or vice versa. This may cause the selection of points that are not in truly corresponding locations.

One way around this could perhaps be to allow the shape of the surface itself, as opposed to the mapping to a standardised mesh, to drive the determination of correspondences. For example, beginning with a set of landmark points it could be possible to interpolate between various points in order to create a set of pseudo-landmarks that represent curves along the surface. An interpolation could then be conducted between the psuedo-landmarks on different curves in order to connect pairs of curves, and this process could be completed until an adequately dense representation of the surface is defined. For the landmarks of the cleft data, for example, this interpolation would serve to trace along the upper edge of the lip regardless of whether the mouth itself was open or closed.

However, clearly this method is still heavily reliant on the original set of user-placed landmark points. A more automated method would of course be preferable, where solely information inherent in the surface is utilised and no manual intervention is required. It has been seen that curvature is a key descriptor of shape, and as such it would perhaps be possible to define the landmarks based on this information alone. While by no means fully-automated, the method presented for correspondence in the breast data was a step towards achieving this, with a combination of landmark points and inherent geometrical information being utilised. A very small number of landmarks were used to provide a means of indexing the surface and analysing the curvature, from which the boundary points of the breast could be determined. The interpolation of each of these boundary points along with one crucial landmark (namely the most prominent point of the breast) then provided the required set of correspondences. This type of approach was particularly useful in the setting of breast analysis, where only a handful of easily distinguishable landmarks are available for the surface and the previously discussed approach of warping would therefore not be reliable.

An improvement to this approach of determining surface correspondence could possibly be achieved through the use of local principal curves, as introduced by Einbeck

et al. (2005). By starting with an individual point in the dataset, these curves are calculated based on a principal components analysis of solely a localised set of points within some small bandwidth distance, and as such they could therefore prove useful in the case of problematic and complex surface shapes such as seen with the breast data. For example, when attempting to determine the breast boundary, provided that the original starting point truly lay on the boundary, this type of curve could potentially be used to trace around the edge of the breast without being affected by the extraneous points that were found further away on the surface, thereby negating the need for exclusion/acceptance criteria of candidate points. This more flexible type of curve could therefore make the approach highly applicable in a variety of settings.

However, whatever the approach taken, it is clearly of key importance to ensure the use of a valid method of determining point correspondence. Flawed surfaces will of course lead to invalid results, no matter how well thought out an analysis is applied to them.

Some brief discussion will now be given of each of the three independent studies, beginning first with the breast reconstruction data.

8.2 Breast Asymmetry

For this dataset, interest lay in the extent of breast symmetry achieved through the reconstructive surgery. Previous methods of breast assessment have been largely subjective, and it was therefore of interest to develop a quantitative means of assessing the asymmetry. As detailed in Chapter 5, an objective method for the assessment of asymmetry was succesfully derived. The approach was based strongly on the methodology of both Mardia et al. (2000) and Bock and Bowman (2006), although some adjustments to the methods were required in this particular setting. A technique was devised to determine the relative position of the reconstructed breast on the opposite side of the chest, specifically by an appropriate reflection of the points and their subsequent rotation around the midline of the chest. From this position a

more informed description of the shape differences could be obtained, with it being possible to evaluate the contribution to the asymmetry of the position, orientation, size and shape of the reconstructed breast.

The technique was applied to both the landmarks and the derived surface representations, with similar results being obtained by both methods. Both sets of asymmetry scores were found to be strongly related to a set of subjective scores from expert reviewers, which suggested that a valid method for the assessment of asymmetry had been derived.

There was one case, however, for which the landmark and surface methods led to very different scores. As was seen in Chapter 5, this was due to large differences that were seen in the upper boundaries of the two breasts. As there was no landmark defining this upper limit, this issue had not been reflected in the asymmetry score. This discrepancy highlights two important points - firstly that a comprehensive coverage of the shape is essential with landmark data, and secondly that the boundaries for surface data, and hence the areas that are included in the analysis, must be accurately defined.

Another issue highlighted through this study was the importance of control data. Although the methodology developed here could be validated by the subjective scores, it was difficult to interpret what the individual scores meant without the availability of a baseline for comparison. These reproducible methods could however be used for a comparison with control data in the future, or alternatively with data arising from different techniques for the reconstructive surgery.

8.3 Longitudinal Facial Growth

The second study investigated was concerned with describing the 'normal' facial growth patterns displayed by young children, specifically between the ages of 3 months and 5 years. Through the creation of representative facial meshes, the patterns of variation existing in the dataset were examined by means of a principal components analysis, from which it was found that more than 90% of the variability

seen across the 874 data points could be explained by only 8 components. The dimensionality of the problem could therefore be reduced significantly through a direct analysis of these principal component scores.

As the data was longitudinal in nature, with several images available for a given child, a linear mixed effects model was utilised. This type of model allowed for a particular trend over time for any given component, however the differences that would naturally be expected across the individual children were also accounted for by the use of subject specific deviations from this mean trend, or 'random effects'. It was also found necessary to include time as a non-linear effect in the model, suggesting that growth is not necessarily occurring at a constant rate and in a constant direction over this time period.

In addition to the method discussed in Chapter 6, an alternative approach to the calculation of principal components was also attempted, in which the residuals for a given child were calculated based on the deviances over time from that child's own particular mean face (as opposed to from the overall mean face). It was thought that using these residuals may serve to remove the effect of differences naturally occurring across the various facial shapes and perhaps place focus more clearly on the changes or growth patterns that were occurring over time. However, this approach was in fact found to give very similar results to the PCA based on residuals from the overall mean, and was therefore not thought to be a worthwhile extension of the model.

Apart from an issue relating to the natural resting position of the mouth, which is discussed in Chapter 7, the results arising from the model appear to be in keeping with what one would expect from facial growth at this age. The general trend over time showed an increase in facial size along with an overall lengthening of the face, describing well the changes seen from baby to young child. The model was also utilised to investigate the presence of a possible sex effect, from which it was concluded that while there is a significant size difference between boys and girls at this age, the actual shape changes occurring are in fact similar.

Generally, the methods developed proved to be very useful in the analysis of facial growth, allowing us to visualise the expected face at each observed timepoint and

also determine confidence regions for these expected faces. Through the application of a spline interpolation of the fitted values, the model also allows for the estimation of the average face at any time point within the studied age range. An alternative approach to this interpolation, however, is to model the principal component scores themselves in terms of a smooth non-linear trend over time, through the use of a P-spline in the mixed model for example. However, due to the small number of available timepoints and highly non-linear patterns in the PC scores, it was found that the splines required a large number of degrees of freedom in this setting and thus were not be a great improvement on the current approach with time as a factor. When modelling over a longer time period, however, this could possibly be a valuable extension to the model.

The scope and longitudinal nature of this dataset make it fairly unique, and it is currently hoped to extend this study further by capturing follow-up images from the same cohort of children, thereby extending the age range modelled to the 10 or 11 old year age group.

8.4 Orthognathic Prediction

In the final study investigated, a novel approach was presented to modelling the soft tissue changes that are seen in orthognathic surgery. Unlike the other studies discussed in this thesis, where a set of corresponding locations was found on each image, the approach taken here was slightly different. In place of determining a representation of the surface, the set of all surface points were instead utilised, with each point being allocated a corresponding 'weight'. These weights were based on the location of the point and therefore created a correspondences of sorts, by allowing points in similar positions to be treated in the same manner across all images.

In the literature, most approaches to this problem are based on biomechanical approaches such as finite element modelling. The method developed for this thesis, however, was based on simpler mathematical models, the parameters of which could be statistically determined. The main advantage to this approach lies in the speed

of calculation, which is far faster than that of the existing methods. The model presented has the flexibility of allowing different areas of the face to be modelled independently, based on the transformations that have taken place in the underlying bones. While the preliminary results presented here appear promising and compare well to those that have been published on existing methods, it is clear that more data would be required in order to determine an accurate set of optimal parameters. With this small exploratory dataset however, it was possible to model the post-surgical outcome with an average error of only 1.41mm across the surface.

Several improvements could yet be made to the model however. For example, there is still a slight issue with the smoothness of the predicted surfaces. While this had been greatly improved upon through the constraints placed on the modelling parameters, the surfaces occasionally remained slightly disjoint in the areas where the various regions of the surface met. This should be further explored, perhaps by determining an improved set of constraints or perhaps by forcing points that are connected to one another on the mesh to be within a certain distance.

Another possible improvement could be to include information on the osteotomies, or bone cuts, that have been made. At present the entire jaw area is treated in the same manner regardless of the position of these osteotomies, whereas knowledge of this could perhaps be included in order to give a more informed model. Further extensions to the method could also include a new approach to the way in which the nose and lips are treated. Under the current model, these regions are treated in the same way as the rest of the surface, which does not appear to be ideally capturing the changes that take place in these regions.

8.5 Final Remarks

In summary, this thesis has discussed a variety of the ways in which three-dimensional shape data can be analysed. The focus has specifically centred on the extension of standard landmark based methods to a more comprehensive surface based analysis, as well as discussion of the issues that often arise from this type of approach. There

are many natural extensions that could be taken to the work presented here, and these have been discussed at both at a general and application specific level. A key extension could perhaps be an improvement to the methods for determining surface correspondence, for example by using curvature to provide a more automated and less landmark driven approach to correspondence. As well as removing the scope for human error, this will help to overcome issues such as that seen with the mouth position in the cleft data. While the techniques discussed here have generally been illustrated in terms of a specific medical application, it is hoped that they would be more widely applicable and prove useful in a variety of situations.

Bibliography

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

Ayoub, A., Garrahy, A., Hood, C., White, J., Bock, M., Siebert, J. P., Spencer, R. and Ray, A. (2003), 'Validation of a vision-based, three-dimensional facial imaging system', *The Cleft Palate-Craniofacial Journal* **40**(5), 523–529.

Barry, S. J. E. (2008), Longitudinal analysis of three-dimensional facial shape data, PhD thesis, University of Glasgow.

Barry, S. J. E. and Bowman, A. W. (2008), 'Linear mixed models for longitudinal shape data with applications to facial modeling', *Biostatistics* **9**(3), 555–565.

Bishara, S. E., Ortho, D. and Jakobsen, J. R. (1985), 'Longitudinal changes in three normal facial types', *American Journal of Orthodontics* **88**(6), 466 – 502.

Bland, M. and Altman, D. (1986), 'Statistical methods for assessing agreement between two methods of clinical measurement', *The Lancet* **327**, 307.

Bock, M. T. and Bowman, A. W. (2006), 'On the measurement and analysis of asymmetry with applications to facial modelling', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**(1), 77–91.

Bookstein, F. L. (1986), 'Size and shape spaces for landmark data in two dimensions', *Statistical Science* **1**(2), pp. 181–222.

Bookstein, F. L. (1989), 'Principal warps: Thin-plate splines and the decomposition of deformations', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(6), 567–585.

Bookstein, F. L. (1996*a*), Combining the tools of geometric morphometrics, *in* 'Advances in Morphometrics', Plenum Press, pp. 131–151.

Bookstein, F. L. (1996*b*), Shape and the information in medical images: A decade of the morphometric synthesis, *in* 'Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis', pp. 2–12.

Bowman, A. W. and Bock, M. T. (2006), 'Exploring variation in three-dimensional shape data', *Journal of Computational and Graphical Statistics* **15**(3), 524–541.

Brignell, C. J., Dryden, I. L., Gattone, S. A., Park, B., Leask, S., Browne, W. J. and Flynn, S. (2010), 'Surface shape analysis with an application to brain surface asymmetry in schizophrenia', *Biostatistics* **11**(4), 609–630.

Brown, B. J. and Rusinkiewicz, S. (2007), 'Global non-rigid alignment of 3-d scans', *ACM Transactions on Graphics* **26**(3).

Brown, T., Ringrose, C., Hyland, R., Cole, A. and Brotherston, T. (1999), 'A method of assessing female breast morphometry and its clinical application', *British Journal of Plastic Surgery* **52**(5), 355–359.

Bryan, D. and Hunt, N. (1993), 'Surgical accuracy in orthognathic surgery', *British Journal of Oral and Maxillofacial Surgery* **31**(6), 343 – 349.

Burke, P. and Beard, L. (1967), 'Stereophotogrammetry of the face: A preliminary investigation into the accuracy of a simplified system evolved for contour mapping by photography', *American Journal of Orthodontics* **53**(10), 769–782.

Caloss, R., Atkins, K. and Stella, J. P. (2007), 'Three-dimensional imaging for virtual assessment and treatment simulation in orthognathic surgery', *Oral and Maxillofacial Surgery Clinics of North America* **19**(3), 287 – 309. Orthognathics.

Cardoso, M. J., Cardoso, J., Amaral, N., Azevedo, I., Barreau, L., Bernardo, M., Christie, D., Costa, S., Fitzal, F., Fougo, J. L., Johansen, J., Macmillan, D.,

Mano, M. P., Regolo, L., Rosa, J., Teixeira, L. and Vrieling, C. (2007), 'Turning subjective into objective: The bcct.core software for evaluation of cosmetic results in breast cancer conservative treatment', *The Breast* **16**(5), 456 – 461.

Cardoso, M. J., Cardoso, J., Santos, A. C., Barros, H. and de Oliveira, M. C. (2006), 'Interobserver agreement and consensus over the esthetic evaluation of conservative treatment for breast cancer', *The Breast* **15**(1), 52 – 57.

Chabanas, M., Marcaux, C., Chouly, F., Boutault, F. and Payan, Y. (2004), Evaluating soft tissue simulation in maxillofacial surgery using pre and post operative CT scans, *in* 'Proceedings of CARS', pp. 419–424.

Chabanas, M., Payan, Y., Marecaux, C., Swider, P. and Boutault, F. (2006), 'Comparison of linear and non-linear soft tissue models with post-operative CT scan in maxillofacial surgery', *Medical Simulation* (2003), 19–27.

Chvatal, B. A., Behrents, R. G., Ceen, R. F. and Buschang, P. H. (2005), 'Development and testing of multilevel models for longitudinal craniofacial growth prediction', *American Journal of Orthodontics and Dentofacial Orthopedics* **128**(1), 45 – 56.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F. and Ranzuglia, G. (2008), MeshLab: an Open-Source Mesh Processing Tool, *in* 'Italian Chapter Conference', pp. 129–136.

Cootes, T., Taylor, C., Cooper, D. and Graham, J. (1992), Training models of shape from sets of examples, *in* 'Proceedings of the British Machine Vision Conference', pp. 9–18.

Dellavia, C., Catti, F., Sforza, C., Tommasi, D. G. and Ferrario, V. F. (2010), 'Craniofacial growth in ectodermal dysplasia', *The Angle Orthodontist* **4**(80), 733–739.

Donatsky, O., Bjørn-Jørgensen, J., Hermund, N. U., Nielsen, H., Holmqvist-Larsen, M. and Nerder, P. H. (2009), 'Accuracy of combined maxillary and mandibular repositioning and of soft tissue prediction in relation to maxillary antero-superior

repositioning combined with mandibular set back: A computerized cephalometric evaluation of the immediate postsurgical outcome using the tiops planning system', *Journal of Cranio-Maxillofacial Surgery* **37**(5), 279 – 284.

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

Duchon, J. (1976), 'Interpolation of functions with two variables according to the principle of bending of thin plates. [Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces.]', *RAIRO Analyse Numerique* **10**, 5–12.

Einbeck, J., Tutz, G. and Evers, L. (2005), 'Local principal curves', *Statistics and Computing* **15**(4), 301–313.

Ercan, I., Ozdemir, S. T., Etoz, A., Sigirli, D., Tubbs, R. S., Loukas, M. and Guney, I. (2008), 'Facial asymmetry in young healthy subjects evaluated by statistical shape analysis', *Journal of Anatomy* **213**(6), 663–669.

Ferrario, V. F., Sforza, C., Poggio, C. E. and Schmitz, J. H. (1999), 'Soft-tissue facial morphometry from 6 years to adulthood: A three-dimensional growth study using a new modeling', *Plastic and Reconstructive Surgery* **103**(3), 768 – 778.

Fieuws, S. and Verbeke, G. (2006), 'Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles', *Biometrics* **62**(2), 424–431.

Fisher, R. A. (1947), *The Design of Experimments*, Vol. 4, Oliver and Boyd.

Fleiss, J. L. (1981), *Statistical Methods for Rates and Proportions*, Wiley.

Goldstein, H. (1986), 'Efficient statistical modelling of longitudinal data', *Annals of Human Biology* **13**(2), 129–141.

Green, P. J. and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall.

Gwilliam, J. R., Cunningham, S. J. and Hutton, T. (2006), 'Reproducibility of soft tissue landmarks on three-dimensional facial scans', *European Journal of Orthodontics* **28**, 408–415.

Hajeer, M. Y., Mao, Z., Millett, D. T. and Ayoub, A. F. (2005), 'A new three-dimensional method of assessing facial volumetric changes after orthognathic treatment', *The Cleft Palate-Craniofacial Journal* **42**.

Hastie, T. and Stuetzle, W. (1989), 'Principal curves', *Journal of the American Statistical Association* **84**(406), 502–516.

Hastie, T. and Weingessel, A. (2009), *princurve: Fits a Principal Curve in Arbitrary Dimension*. R package version 1.1-10.

Hennessy, R. J., Baldwin, P. A., Browne, D. J., Kinsella, A. and Waddington, J. L. (2007), 'Three-dimensional laser surface imaging and geometric morphometrics resolve frontonasal dysmorphology in schizophrenia', *Biological Psychiatry* **61**(10), 1187 – 1194.

Hennessy, R. and Moss, J. (2001), 'Facial growth: separating shape from size', *European Journal of Orthodontics* **23**, 275–285.

Henseler, H., Khambay, B. S., Bowman, A., Smith, J., Siebert, J. P., Oehler, S., Ju, X., Ayoub, A. and Ray, A. K. (2011), 'Investigation into accuracy and reproducibility of a 3-d breast imaging system using multiple stereo cameras', *Journal of Plastic, Reconstructive and Aesthetic Surgery* **64**(5), 577–582.

Henseler, H., Smith, J., Bowman, A., Khambay, B., Ji, X., Ayoub, A. and Ray, A. (2012), 'Comparison of objective breast symmetry assessment by three-dimensional imaging with subjective assessment after breast reconstruction using the latissimus dorsi flap', *Journal of Plastic, Reconstructive and Aesthetic Surgery* . To Appear.

Hood, C., Hosey, M., Bock, M., White, J., Ray, A. and Ayoub, A. (2004), 'Facial characterization of infants with cleft lip and palate using a 3d capture technique', *Craniofacial and Cleft Palate Journal* **41**, 27–35.

Hutton, T., Buxton, B., Hammond, P. and Potts, H. (2003), 'Estimating average growth trajectories in shape-space using kernel smoothing', *IEEE Transactions on Medical Imaging* **22**(6), 747 –753.

Hutton, T. J. (2004), Dense Surface Models of the Human Face, PhD thesis, University College London.

Hutton, T. J., Buxton, B. F. and Hammond, P. (2001), 'Dense surface point distribution models of the human face', *Mathematical Methods in Biomedical Image Analysis* pp. 153–160.

Jacob, H. B. and Buschang, P. H. (2011), 'Vertical craniofacial growth changes in French-Canadians between 10 and 15 years ofage', *American Journal of Orthodontics and Dentofacial Orthopedics* **139**(6), 797 – 805.

Kau, C. H. and Richmond, S. (2008), 'Three-dimensional analysis of facial morphology surface changes in untreated children from 12 to 14 years of age', *American Journal of Orthodontics and Dentofacial Orthopedics* **134**(6), 751 – 760.

Keeve, E., Girod, S., Pfeifle, P. and Girod, B. (1996), Anatomy-based facial tissue modeling using the finite element method, *in* 'Proceedings of the 7th conference on Visualization '96', IEEE Computer Society Press, Los Alamitos, CA, USA.

Kendall, D. G. (1984), 'Shape manifolds, procrustean metrics, and complex projective spaces', *Bulletin of the London Mathematical Society* **16**(2), 81–121.

Kent, J. T. and Mardia, K. V. (2001), 'Shape, procrustes tangent projections and bilateral symmetry', *Biometrika* **88**(2), pp. 469–485.

Kent, J. T., Mardia, K. V., Morris, R. J. and Aykroyd, R. G. (2000), Procrustes growth models for shape, *in* 'Proceedings of the First Joint Statistical Meeting', pp. 236–238.

Kent, J. T., Mardia, K. V., Morris, R. J. and Aykroyd, R. G. (2001), Functional models of growth for landmark data, *in* 'Proceedings in Functional and Spatial Data Analysis', pp. 109–115.

Klingenberg, C. P., Barluenga, M. and Meyer, A. (2002), 'Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry', *Evolution* **56**(10), 1909–1920.

Kusnoto, B., Figueroa, A. and Polley, J. (1999), 'A longitudinal three-dimensional evaluation of the growth pattern in hemifacial microsomia treated by mandibular distraction osteogenesis: a preliminary report.', *Journal of Craniofacial Surgery* (6), 480–486.

Laird, N. M. and Ware, J. H. (1982), 'Random-effects models for longitudinal data', *Biometrics* **38**(4), 963–974.

Landis, J. R. and Koch, G. G. (1977), 'The measurement of observer agreement for categorical data', *Biometrics* **33**(1), 159–174.

Lipman, Y. and Funkhouser, T. (2009), 'Mobius voting for surface correspondence', *ACM Transactions on Graphics* **28**(3).

Loncaric, S. (1998), 'A survey of shape analysis techniques', *Pattern Recognition* **31**, 983–1001.

Losken, A., Seify, H., Denson, D. D., Paredes, Jr., A. A. and Carlson, G. W. (2005), 'Validating three-dimensional imaging of the breast', *Annals of Plastic Surgery* **54**(5), 471–476.

Loughry, C. W., Sheffer, D., Price, T. E. J., Lackney, M., Bartfai, R. G. and Morek, W. M. (1987), 'Breast volume measurement of 248 women using biostereometric analysis', *Plastic and Reconstructive Surgery* **80**(4), 553–558.

Mao, Z., Siebert, J., Cockshott, W. and Ayoub, A. (2004), Constructing dense correspondences to analyze 3d facial change, *in* 'Proceedings of the 17th International Conference on Pattern Recognition', Vol. 3, pp. 144 – 148.

Mardia, K. V., Bookstein, F. L. and Moreton, I. J. (2000), 'Statistical assessment of bilateral symmetry of shapes', *Biometrika* **87**(2), 285–300.

Miller, J. (2009), Shape Curve Analysis Using Curvature, PhD thesis, University of Glasgow.

Mollemans, W., Schutyser, F., Nadjmi, N., Maes, F. and Suetens, P. (2007), 'Predicting soft tissue deformations for a maxillofacial surgery planning system: From

computational strategies to a complete clinical validation', *Medical Image Analysis* **11**(3), 282 – 301.

Morris, R. J., Kent, J. T., Mardia, K. V., Fidrich, M., Aykroyd, R. G. and Linney, A. (1999), Analysing growth in faces, *in* 'Proceedings of the Conference on Imaging Science', pp. 404–410.

Mueller, W. and Martorell, R. (1988), Reliability and accuracy of measurement, *in* 'Anthropometric Standardisation Reference Manual', Human Kinetics Books.

Mutsvangwa, T. and Douglas, T. S. (2007), 'Morphometric analysis of facial landmark data to characterize the facial phenotype associated with fetal alcohol syndrome', *Journal of Anatomy* **210**(2), 209–220.

Oehler, S. (2006), Breast analysis tool for volume and surface measurement of breast segment for clinical purposes, Master's thesis, Uniersity of Glasgow.

Papadopoulos, M. A., Christou, P. K., Christou, P. K., Athanasiou, A. E., Boettcher, P., Zeilhofer, H. F., Sader, R. and Papadopulos, N. A. (2002), 'Three-dimensional craniofacial reconstruction imaging', *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics* **93**(4), 382 – 393.

Popat, H. and Richmond, S. (2010), 'New developments in: three-dimensional planning for orthognathic surgery', *J. Orthod.* **37**(1), 62–71.

Power, G., Breckon, J., Sherriff, M. and McDonald, F. (2005), 'Dolphin imaging software: An analysis of the accuracy of cephalometric digitization and orthognathic prediction', *International Journal of Oral and Maxillofacial Surgery* **34**(6), 619–626.

Pulli, K. (1999), Multiview registration for large data sets, *in* 'International Conference on 3-D Imaging and Modeling', pp. 160–168.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ras, F., Habets, L. L. M. H., Ginkel, F. C. V. and Prahl-Andersen, B. (1995), 'Longitudinal study on three-dimensional changes of facial asymmetry in children between 4 to 12 years of age with unilateral cleft lip and palate', *The Cleft Palate-Craniofacial Journal* **32**(6).

Samir, C., Srivastava, A. and Daoudi, M. (2006), 'Three-dimensional face recognition using shapes of facial curves', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 1858–1863.

Savriama, Y. and Klingenberg, C. P. (2011), 'Beyond bilateral symmetry: geometric morphometric methods for any type of symmetry', *BMC Evolutionary Biology* **11**.

Smith, J. and Bowman, A. (2010), Asymmetry in breast reconstruction patients, *in* 'Proceedings of the 25th International Workshop on Statistical Modelling', pp. 505–508.

Smith, J. and Bowman, A. (2011), Boundary identification in 3-d images, *in* 'Proceedings of the 26th International Workshop on Statistical Modelling', pp. 565–570.

Sonat, F., Ercan, I., Ozdemir, S., Ozkaya, G. and Noyan, B. (2009), 'Statistical shape analysis of the rat hippocampus in epilepsy', *Anatomical Science International* **84**, 298–304.

Spivak, M. (1979), *Differential Geometry*, Vol. 2, Publish or Perish, Inc.

Takemura, C. M., Cesar, Jr., R. M., Arantes, R. A. T., Costa, L. d. F., Hingst-Zaher, E., Bonato, V. and dos Reis, S. F. (2004), 'Morphometrical data analysis using wavelets', *Real-Time Imaging* **10**(4), 239–250.

Toma, A., Zhurov, A., Playle, R., Ong, E. and Richmond, S. (2009), 'Reproducibility of facial soft tissue landmarks on 3d laser-scanned facial images', *Orthodontics and Craniofacial Research* **12**, 33–42.

Ulijaszek, S. J. and Kerr, D. A. (1999), 'Anthropometric measurement error and the assessment of nutritional status', *British Journal of Nutrition* **82**, 165.

Vandewalle, P., Schutyser, F., Van Cleynenbreugel, J. and Suetens, P. (2003), Modelling of facial soft tissue growth for maxillofacial surgery planning environments, *in* 'Proceedings of the 2003 international conference on surgery simulation and soft tissue modeling', IS4TM'03, Springer-Verlag, Berlin, Heidelberg, pp. 27–37.

Wang, Y., Peterson, B. S. and Staib, L. H. (2000), Shape-based 3d surface correspondence using geodesics and local geometry, *in* 'Proceeedings of IEEE Conference of Computer Vision and Pattern Recognition', pp. 644–651.

Westermark, A., Zachow, S. and Eppey, B. L. (2005), 'Three-dimensional osteotomy planning in maxillofacial surgery including soft tissue prediction', *Journal of Craniofacial Surgery* **16**(1), 228–238.

Xia, J., Ip, H., Samman, N., Wong, H., Gateno, J., Wang, D., Yeung, R., Kot, C. and Tideman, H. (2001), 'Three-dimensional virtual-reality surgical planning and soft-tissue prediction for orthognathic surgery', *IEEE Transactions on Information Technology in Biomedicine* **5**(2), 97 –107.

Yamada, T., Mori, Y., Minami, K., Mishima, K. and Tsukamoto, Y. (2002), 'Three-dimensional analysis of facial morphology in normal Japanese children as control data for cleft surgery', *The Cleft Palate-Craniofacial Journal* **39**(5), 517–526.

Zachow, S., Hierl, T. and Erdmann, B. (2004), On the predictability of tissue changes for osteotomy planning in maxillofacial surgery: a comparison with postoperative results, *in* 'Proceedings of CARS', pp. 648–653.

Zitov, B. and Flusser, J. (2003), 'Image registration methods: a survey', *Image and Vision Computing* **21**(11), 977 – 1000.