Statistical analysis of facial landmark data for optimisation of Fetal Alcohol Syndrome diagnosis

By

Tinashe E.M. Mutsvangwa

A thesis submitted to the

Faculty of Health Sciences in partial fulfilment of the

requirements for the degree of

MSc (Med)

in

BIOMEDICAL ENGINEERING

University of Cape Town
February 2006

DECLARATION

Statistical analysis of facial landmark data for optimisation of Fetal Alcohol Syndrome diagnosis

I, Tinashe Mutsvangwa, hereby declare that:

1) This thesis with the above title is my own unaided work, and that apart from the normal guidance from my supervisor; I have received no assistance except as stated below.

2) Except where indicated to the contrary, neither the substance nor any part of this thesis with the above title has been submitted in the past, or is being, or is to be submitted for a degree in this university or any other university.

This thesis has been presented by myself and is being submitted for examination for the degree of Master of Science in Medicine in Biomedical Engineering at the University of Cape Town

---

Signature of Author

Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Tania Douglas, my supervisor, for all her guidance and advice in my Masters project.

Thanks, also, to the National Research Fund (NRF) for partially funding my Masters degree.

Thanks to my mother for inspiring me to do my best, and my family for believing in me. I would also like to thank my friends for their support.

ABSTRACT

This project involved the statistical analysis of facial landmark used in Fetal Alcohol Syndrome (FAS) diagnosis. FAS is a clinical condition caused by excessive maternal consumption of alcohol during pregnancy. Diagnosis of FAS depends on evidence of growth retardation, CNS neurodevelopment abnormalities, and a characteristic pattern of facial anomalies, specifically a short palpebral fissure length, smooth philtrum, flat upper lip and flat midface. The unique facial appearance associated with FAS is emphasized in diagnosis that relies, in part, on the comparison of linear measurements of facial features to population norms.

The suspected high prevalence of FAS in South Africa requires large scale screening programs, which necessitates cheap, cost-effective, and easy to use diagnostic tools. A stereo-photogrammetric screening tool for FAS has previously been developed in the MRC/UCT Medical Imaging Research Unit. This report describes the use of discriminant analysis and facial shape analysis on landmark data obtained using the stereo-photogrammetric tool. Discriminant analysis has previously been used for the diagnosis of the FAS face. Facial shape analysis has been used in the diagnosis of a number of genetic disorders, but its use in analysing the FAS face has been limited.

The study population consisted of images obtained during the screening of first-grade children from disadvantaged communities in the Gauteng and Northern Cape Provinces of South Africa for FAS. The subjects were appraised by three independent

dysmorphologists and images of subjects with a clinical diagnosis of FAS and normal controls were used in this project.

Discriminant analysis was performed to determine those facial features (concentrating on the eyes) that can best be used to diagnose FAS. Occipital frontal circumference (OFC), palpebral fissure length (PFL) and inter-pupillary distance (IPD) alone were used to classify subjects with a specificity of 86% and sensitivity of 100% in the first discriminant analysis study on a test sample of 25 subjects, after training on 42 subjects. A specificity of 95% and sensitivity of 100% in classifying subjects with FAS and normal controls was obtained across the total sample of 67 subjects using the discriminant equation. In the second discriminant analysis study using upper lip circularity as a proxy measure for upper lip thinness, OFC, PFL and IPD and upper lip circularity (ULC) were used to classify a test sample of 20 subjects with a specificity of 85.7% and sensitivity of 100% after training on 29 subjects. A specificity of 94% and sensitivity of 100% was obtained across the total sample of 49 subjects using the discriminant equation.

Facial shape analysis of stereo-photogrammetric image pairs was performed to see how the FAS facial shape differs from that of the normal face, and the extent of overlap of the two categories of faces was assessed. The use of Procrustes analysis followed by principal component analysis to reduce dimensionality and explore shape variability was also shown to be a successful method for analysing the modes of facial variation between FAS and normal controls. The first facial shape analysis using facial features

that are measured and used in FAS diagnosis did not reveal any statistically significant difference between the FAS and the normal facial shape. Inclusion of additional landmarks relevant to the FAS "gestalt" in a second facial shape study resulted in better discrimination between the FAS and normal subjects. Head circumference and midfacial features together with eye features were found to play a role in differentiating between FAS and normal subjects.

While normal population reference values for linear facial measurements may differ in South African children, facial shape analysis indicates that the shape variation associated with FAS in this country corresponds to that observed elsewhere. The landmark based, stereo-photogrammetric approach presented here was limited by the number of biologically homologous landmarks that are easily reproducible. The size of the study sample was also a limitation and a larger study sample, particularly for discriminant analysis, is recommended in future work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Fetal Alcohol Syndrome (FAS) is a clinical condition caused by excessive maternal consumption of alcohol during pregnancy. A syndrome is a recognizable pattern of malformations, disruptions and/or deformations occurring together to characterise a particular disease (Hunter, 2002). In 1968, a French physician first noted the pattern of behavioural and physical characteristics of children exposed to alcohol during prenatal development. The term Fetal Alcohol Syndrome was first used to describe growth retardation, facial anomalies and the neurological abnormalities resulting from prenatal alcohol exposure by Jones and Smith (1973).

The disease has been recognised world wide with a prevalence of 1 to 3 per 1000 live births in developed nations (May, et al., 2000). FAS has come to be accepted as the "leading identifiable preventable cause of mental retardation and neurological deficit in the Western world" (Burd, et al., 2003). Data from poorer areas of the world, which might have higher prevalence, are not readily available. Diagnosis of FAS depends on evidence of 1) growth retardation, 2) central nervous system neurodevelopment abnormalities and 3) a characteristic pattern of facial anomalies such as short palpebral fissure length, smooth philtrum, flat upper lip and flat midface (Astley and Clarren, 1995) (refer to Figure 1.1). Compounding the primary features of the disorder are secondary effects such as behavioural problems in later stages of life including low self esteem, depression and school failure because of the associated neurological problems

and social chastisement that result if the disorder is not diagnosed early enough (Astley and Clarren, 1996).

1.2 Definition of Problem

The traditional method of diagnosis has been the so called "gestalt" method where a patient had to present evidence of all three of the above main hallmarks of the disease. At present, a confident diagnosis of FAS can generally only be made by an expert dysphormologist. In the hands of trained professionals, this method can be sufficiently accurate and reproducible for diagnosing the condition (Clarren, et al., 1987). The problem arises, however, when the diagnosis is made by untrained professionals and misdiagnosis is possible. Misdiagnosis can adversely affect the patient through stigmatisation and can severely stifle any screening and prevalence efforts that could help track the disorder (Astley and Clarren, 1995).

There is much research aimed at providing a quantitative case definition for the disorder based on the facial anomalies characteristic of FAS. The advantage in concentrating on facial abnormalities as a case definition is that the growth retardation and brain dysfunction typical in FAS are also characteristic of other birth defects. Determining a subset of facial anomalies that are accurately and consistently capable of diagnosing FAS has proven to be a challenge because of lack of objective, quantitative scales to measure and report the magnitude of expression of key diagnostic facial features. For instance, of the key diagnostic facial features, quantitative guidelines have never been set for how thin the upper lip needs to be or how smooth the philtrum needs to be (see

Figure 1.1). Distinctions between lesser extremes are difficult to consider quantitatively (Astley, 2004).

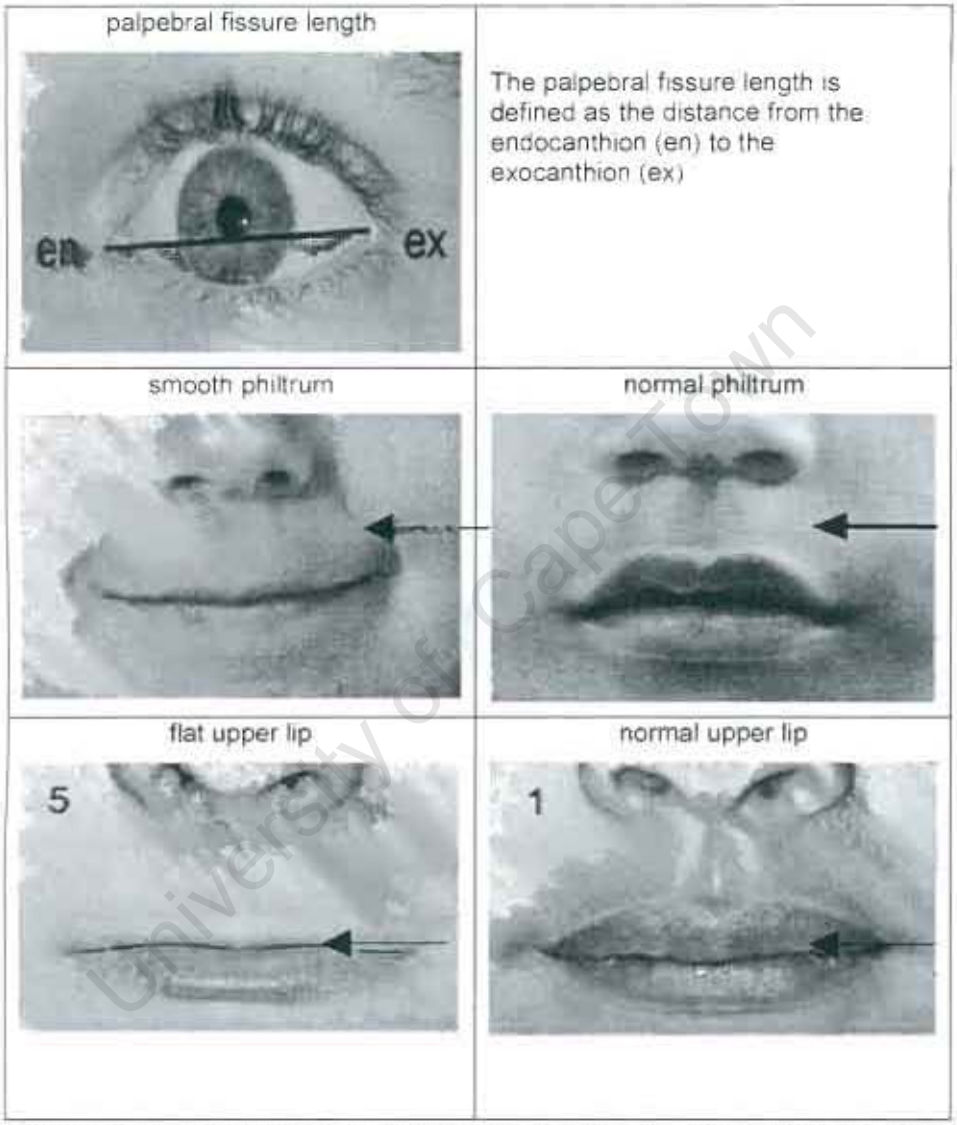


Figure 1.1: Key features characterizing the facial phenotype. There is a clear distinction between the very abnormal and the normal smooth philtrum and thin upper lips but the distinction is not easily quantifiable. The images are from Astley (2004).

Objective quantitative scales have the potential to improve accuracy and precision of diagnosis (Astley, 2004). Statistical analysis on data for FAS diagnosis has thus suffered

from a lack of quantitative scaling. Attributes like smooth philtrum can only be subjectively measured, although the use of Likert scoring brings some objectivity to measurements made in diagnosis.

The highest prevalence of FAS worldwide was reported among first-grade children in a wine-growing region in the Western Cape province of South Africa at approximately 40.5 to 46.4 per 1000 children (Warren, et al., 2001). The frequency of occurrence of the condition in other regions of South Africa is not known and must be determined. A stereo-photogrammetric method for FAS diagnosis using facial dysmorphology has been designed by Meintjes, et al., (2002) for assessment of FAS prevalence in South Africa. It is a fast, accurate screening tool for FAS designed to avoid the costly and labour intensive methods currently used to diagnose the condition which do not lend themselves well to the very large sample prevalence studies necessary in South African communities. These large-scale incidence and prevalence studies are necessary for the implementation of appropriate preventative measures and also for effective provision of clinical services. This thesis explores alternative methods for diagnosis and screening of FAS using data obtained using this stereo-photogrammetric tool.

1.3 Objectives

The aim of this project was to use statistics and statistical shape analysis to compare facial features of children with and without FAS, and hence to derive a facial shape range to which FAS faces belong. Both measurements obtained directly and those derived from the stereo-photogrammetric tool already developed were subjected to:

- Discriminant analysis to determine those facial features that best represent the FAS phenotype.
- Statistical shape analysis on landmarks representing the phenotypic facial features of FAS cited in the literature, to describe the modes of facial shape variation between FAS and non-FAS subjects.
- An assessment of the relevance of facial features indicated in the literature to be useful in FAS diagnosis on a local population of children.

1.4 Thesis outline

Chapter 2 gives a literature review of present methods of measuring and assessing the FAS facial anomalies including direct anthropometry, photogrammetry and three-dimensional facial scanning.

Chapter 3 discusses discriminant analysis theory in detail. Two-group discriminant analysis theory and the assumptions under which the method is performed are outlined.

Chapter 4 describes the application of two-group discriminant analysis. The materials and methodology of the analysis are presented, including a description of the study population and the facial landmarks used in the analysis. The results of the analysis are then presented and conclusions drawn from them.

Chapter 5 covers the theory of facial shape analysis. Shape is defined and Procrustes analysis is introduced as a shape alignment procedure and principal component analysis as a dimensionality reduction technique. Visual assessment of shape variability is also made.

Chapter 6 is concerned with the application of facial shape analysis. The methodology and materials are described, including the study sample, landmarks used and software used. The results of the analysis are presented and discussed.

Chapter 7 compares and contrasts the results of the discriminant analysis and the facial shape analysis.

Chapter 8 presents the overall conclusions drawn in the thesis and recommendations for future modifications that might be relevant to the goal of improving the diagnosis of the FAS facial phenotype.

1.5 Ethics Approval

Ethics approval for the study was granted by the Research Ethics Committee, Faculty of Health Sciences on 21 May 2003 (Rec. Ref 142/2003).

2 LITERATURE REVIEW

Syndrome diagnosis requires the definition of characteristic abnormal patterns associated with a given syndrome. Quantitatively determining the extent of deviation of an individual's facial pattern from the normal state requires the collection of data on normal individuals in order to establish numerical descriptions of normal measurement ranges. Data used in syndrome diagnosis based on facial appearance are typically acquired through:

1) Direct anthropometric measurements.
2) Photogrammetry - Either two-dimensional or three-dimensional.

2.1 Direct anthropometry

Anthropometric measurements for the diagnosis of FAS are done by expert dysmorphologists using hand-held rulers, calipers, or a cloth retractable tape measure. The advantages of direct anthropometric measurements are the relative ease, non-invasiveness and cheapness of the procedures (Moore, et al., 2002). Measurements made close to the eyes e.g. palpebral fissure lengths and inner canthal distances, (especially on young non-cooperative children) are generally made using a clear plastic ruler as the use of calipers is deemed dangerous. The individual is asked to open his/her eyes widely to allow accurate identification of the endocanthion and exocanthion landmarks. Measurements are also normalised to race where appropriate; for example, palpebral fissure length and upper lip thinness vary according to ethnicity (Astley and Clarren, 2001). The facial measurements used in FAS diagnosis in South Africa are palpebral

fissure length (PFL), inter-pupillary distance (IPD), inner canthal distance (ICD) and outer canthal distance (OCD) (Douglas and Viljoen, 2006).

2.2 Measuring facial dysmorphology using photogrammetry

Photogrammetry is the process of obtaining measurements by means of photographs, while stereo-photogrammetry refers to the special case where two or more cameras are used to obtain three-dimensional information of a scene (Douglas, 2004). Indirect methods such as photogrammetry have several advantages over direct surface anthropometry. The measurement inaccuracies introduced by indentation of some features in direct contact with instruments can be avoided. The outlines of a photograph do not move during a measurement in contrast with direct measurements on young subjects who may become restless. The duration of obtaining photographs is generally shorter than for measurements taken directly. The measurements can be repeated, if necessary, from photographs, whereas repeated direct measurements of the subject may not be possible. However, fewer facial measurements may be taken from a photograph than through direct means. Measurements from photographs may vary with changes in lightning conditions and photographs may present less diagnostically important information about the face. Of concern in photogrammetry are the introduction of errors in defining landmarks, especially bony landmarks, and calibration inaccuracies.

Previous studies using photographs have primarily employed two-dimensional photographs. The use of stereo-photogrammetry has been plagued by cost and size

(Meintjes, et al., 2002) although the costs of digital cameras have decreased and the ability of modern computers to deal with larger images has also improved dramatically. In their study on the facial features characteristic of FAS, Meintjes, et al. (2002) used images from a pair of stereo photographs of each child's face obtained using the imaging tool in Figure 2.1. Their results were then compared with measurements that were performed by dysmorphologists in the conventional manner.

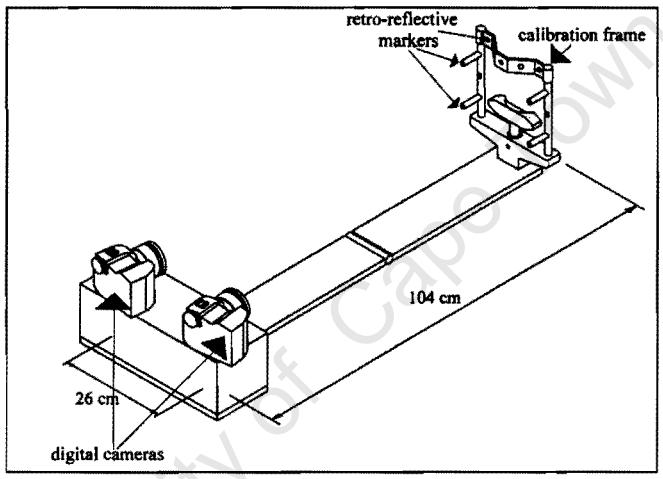


Figure 2.1: Image acquisition tool (Meintjes, et al., 2002). The two high-resolution digital cameras are mounted the same distance from the calibration frame where the child's head is placed, and triggered simultaneously by remote control.

The children were asked to have a relaxed facial expression with eyes fully open, lips gently closed, and not to smile. The cameras used to obtain the images were simultaneously triggered by remote control. The control frame comprises head and chin-rests with vertical supports on each side. The tool was designed with eleven well distributed retro-reflective control markers. The three-dimensional coordinates of these makers are known, and this was later exploited in the calibration of the images.

To produce better depth-of-field, some of the control markers are raised above the control surface of the frame and the face of the subject is semi-immersed within a three-dimension volume defined by the control markers as in Figure 2.2.

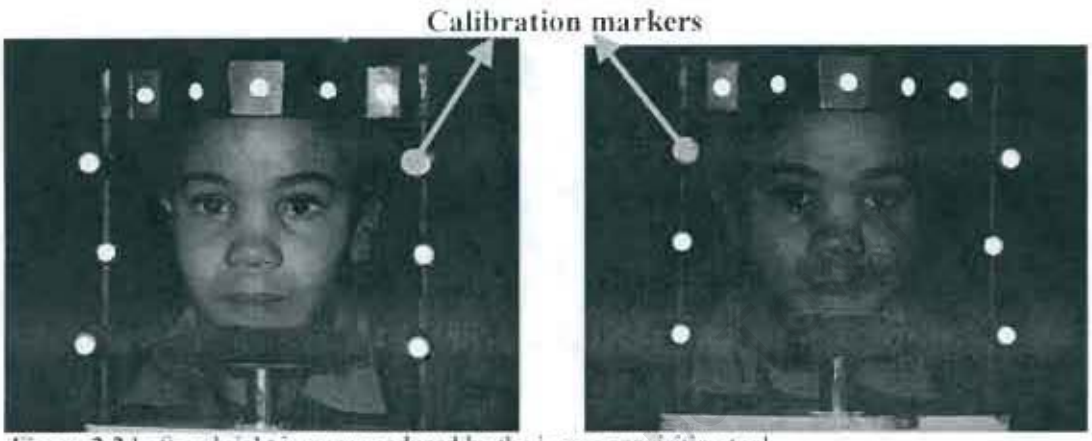


Figure 2.2 Left and right images produced by the image acquisition tool.

After loading the images from a database, software automatically detects the eleven control markers in both the left and right images. The relative positions of the markers in the frame are used to calibrate the image space. A Direct Linear Transformation, as first proposed by Abdel-Aziz and Karara (1971), is then used to transform the two-dimensional image coordinators (x,y) into three-dimensional object-space coordinates (X,Y,Z). Thus the three-dimensional coordinates of any point on the image pair is obtained. Calibration accuracies are reported to be better than 0.5 mm in each direction (Meintjes, et al., 2002).

2.3 Other three-dimensional image acquisition techniques.

Various methods of obtaining three-dimensional surface information about the face exist including laser scanning and computer tomography (CT) images. The limitations of CT scans are the costliness of obtaining the images, absence of surface detail, e.g. eye landmarks, and also the high levels of ionizing radiation subjects are exposed to. Laser based scanners were previously limited by the amount of time it took to capture images. This was especially problematic when images of infants were taken, as they are prone to being restless and fidgety. In their study on sexual dimorphism in craniofacial development, Hennessy, et al., (2002), applied a handheld three-dimension laser surface system to obtain three-dimensional information of the face. Their image acquisition device combined laser line and two cameras with tracking technology to track the position of 1) a sensor on the cameras and 2) the movement of a sensor attached to the subject's head. They produced, after post-processing their captured data, facial surfaces consisting of approximately 80000 points. From these, landmark data were extracted using a combination of manual and surface fitting methods. Four vertical sweeps were needed to get a full surface scan of the face and the subjects were told to keep their eyes closed throughout this time. Thus image acquisition was lengthy and prone to errors especially in the case of young uncooperative subjects. This system was also plagued by the loss of some surface detail, especially in the eye region.

Photogrammetric face scanners, which are rapid and non-invasive, are now able to obtain three-dimensional information about the face. These have the advantage of producing surfaces, which can be viewed from various angles and are not as dependent

on lighting conditions as two-dimensional images. Landmark data is easily locatable on a three-dimensional surface scan compared to two-dimensional images. Hammond, et al., (2004) used two non-contact scanners to simultaneously capture photographic images of the face from four viewpoints using separate charged couple device (CCD) cameras with a speckle pattern projected on the face. They used the speckle pattern in each of the four images to compute a three-dimensional surface that was then overlaid with the subject's appearance, obtained from photographs, to obtain the final result. The captured surface they produced contained between 4000 and 20000 three-dimensional points. Landmark points were then annotated using manual techniques. The landmark points were used in guiding common sets of points into a dense correspondence across all face surfaces in the study in order to compute dense surface models as described in Section 2.5.2 (Hammond, et al., 2004). Hammond, et al., (2005) later used commercial photogrammetric facial surface devices to capture three-dimensional images in their exploration of facial morphology.

2.4 Image processing in anthropometry

Automated image processing algorithms can greatly reduce the amount of time spent in evaluating facial images in syndrome diagnosis (Douglas, 2004). The digitised images can be manipulated on computer since various research areas exist which pertain to the automated extraction of facial features including the eyes, mouth and any features with a high contrast compared to background tissue like the skin.

A method of automatically extracting the eye was reported by Douglas, et al., (2003). After image acquisition, their image-processing algorithm automatically delineated the iris and eye contours to assist them in locating points relevant to FAS diagnosis. The three-dimensional coordinates of these points were calculated using the Direct Linear Transform (DLT) mapping and the relevant distances measured. Distances measured automatically were then compared to those obtained through the manual selection of the digital images. While methods of extracting eye features have become well established, other areas of the face with low contrast compared to the surrounding tissue, e.g. the mouth, are difficult to extract automatically, especially from two-dimensional images of the face.

2.5 Statistical analysis in facial dysmorphology

2.5.1 Direct anthropometry

Astley and Clarren (1995) proposed a FAS screening tool that could differentiate between children with and without FAS as diagnosed by a single dysmorphologist. Discriminant analysis identified thin upper lip, smooth philtrum and palpebral fissures as the three features that best discriminated between children with and without FAS. Using 21 craniofacial measurements, Moore, et al., (2001) took measurements of 100 individuals exposed prenatally to alcohol (41 FAS, 59 Partial FAS) and 31 control subjects, in order to obtain an objective, multivariate case definition of FAS and Partial Fetal Alcohol Syndrome (PFAS) through the use of discriminant analysis with stepwise variable selection. Their discriminant equations identified bigonial breadth and head

circumference as the key features that best differentiated subjects at risk for FAS from normal subjects with 100% accuracy and 100% sensitivity. Discriminant equations also identified all subjects as FAS, PFAS, or non-alcohol exposed with an accuracy of 88% and a sensitivity of 86%. Finally, discriminant equations were derived that could distinguish between alcohol exposed and non-alcohol exposed subjects with 98% sensitivity and 90% specificity. Moore, et al. (2002) later demonstrated that a consistent pattern profile exists for alcohol exposed individuals.

2.5.2 Three-dimensional laser scanning

Hammond, et al., (2004) used manually annotated landmarks on three-dimensional facial surfaces to construct dense surface models (DSM) using the anatomical landmarks to warp a template mesh onto all shapes in a training set. Using Procrustes analysis to align the revised three-dimensional facial surfaces (after the warping) an average shape was calculated. The differences between points on the original surfaces and corresponding points on the calculated average shape are called residuals. Reducing dimensionality using principal component analysis on the residuals allowed the major modes of shape variation to be explored. They used the dense surface models in facial shape analysis for discriminating between 280 control subjects, 90 with Noonan syndrome and 60 with Velo-cardio-facial syndrome (VCFS). Using different pattern recognition techniques they had the highest discrimination success rates, in terms of specificity and sensitivity, of 1) 92% and 93% for children and 83% and 94% for adults, and 2) 88% and 94% for children and adults combined, when discriminating between Noonan syndrome and control subjects. They also obtained 83% sensitivity and 92% specificity for

discriminating between VCFS and control subjects. An accuracy of 95% was obtained in discriminating between Noonan syndrome and VCFC. Hammond, et al., (2005) later extended the use of their dense surface models and pattern recognition approach to inter-syndrome discrimination of Williams (130 subjects), 22q11 deletion (115 subjects), Smith-Magnesis (54 subjects) and Noonan (80 subjects) syndromes. They achieved a worst discrimination rate of 88% for Williams vs. Smith-Magnesis syndromes using the whole facial surface. Their worst syndrome versus control discrimination rate was 88% for Smith-Magnesis syndrome. Localised face patches concentrating on the periorbital, perinasal and perioral regions produced similar results. The lowest inter-syndrome discrimination rate was 76% for the perioral region between the 22q deletion and Noonan syndromes. The worst syndrome versus control discrimination rate was 88% for the periorbital patch in 22q11 deletion syndrome.

3 DISCRIMINANT ANALYSIS THEORY

3.1 Overview

Discriminant function analysis or discriminant analysis (DA) is used to determine which variables (usually continuous) discriminate between two or more naturally occurring groups. The purposes of DA include (Garson, 2005) :

- To classify cases into groups using a discriminant prediction equation.
- To investigate predictor variable mean differences between groups formed by the dependent variable.
- To determine the percent of variance in the dependent variable explained by the predictors.
- To determine the percent of variance in the dependent variable explained by the predictors over and above the variance accounted for by control variables, using sequential DA.
- To assess the relative importance of the predictor variables in classifying the dependent variable.
- To discard variables which are not important in group distinctions.
- To test theory by observing whether cases are classified as predicted.

DA is a three-part process. The first step determines a discriminant function and assesses its significance through an F-test (Wilk's lambda). If the function is significant then the second step is an analysis on the nature of the difference in means of the predictor

variables to determine which are significant. Classification of new cases using this information is the last step generally performed.

The sections that follow are based on information found in Statsoft.Inc (2004) and Garson (2005).

3.2 Two-group discriminant analysis

Two-group DA with stepwise variable selection includes many independent variable measures in order to find the ones that are most discriminatory. Specifically, stepwise discriminant function analysis builds a model of discrimination step-by-step. In each step all variables are reviewed and evaluated to determine which one will contribute most to the discrimination between the two groups. That variable will then be included in the model, and the process starts again. In the two-group case, a linear equation is obtained of the form:

$$D.Score = \text{constant} + c_1 x_1 + c_2 x_2 + \ldots + c_m x_m \qquad \text{Eq (3.1)}$$

where c_1 through c_m are coefficients and here *D.Score* is a continuous variable. When there are more than two groups, more than one discriminant function is obtained, although each function is structured as the linear equation above. The coefficients in the two-group and multi-group cases are interpreted the same way. *D.Score* is the value resulting from the product of the unstandardised coefficients with the observations and is

used for classification purposes. Standardised coefficients are used in assessing the relative importance of each predictor variable in the discrimination between groups. The standardised coefficients are partial coefficients and thus only compare the unique explanation of each predictor not considering any shared explanation. The coefficients are chosen such that were a *D.Score* for each subject to be calculated and an analysis of variance or ANOVA performed on *D.Score*, the ratio of the between groups sum of squares to the within groups sum of squares is as large as possible. The value of this ratio is called an eigen-value. Thus formally the eigen-value λ (the quantity maximised by the coefficients obtained) on *D.Score*:

$$\lambda = \frac{SumofSquares_{between\_groups}}{SumofSquares_{within\_groups}} \qquad \text{Eq (3.2)}$$

Wilk's lambda can be used to test the null hypothesis that the two populations have identical means on *D.Score*. Wilk's lambda tests the significance of the function as a whole and is given by:

$$\text{Wilk's lambda} = \frac{\left|SumofSquares_{within\_groups}\right|}{\left|SumofSquares_{total}\right|} \qquad \text{Eq (3.3)}$$

The smaller the value of Wilk's lambda, the greater the discrimination between the two groups. Partial lambda is used in an ANOVA (F) test of mean differences such that the smaller the lambda for a predictor variable, the more that variable contributes to the discriminant function. Specifically, the partial lambda is the ratio of Wilk's lambda after

adding the respective variable, over the Wilk's lambda before adding the variable and is given by:

$$\text{Partial lambda} = \frac{lambda(after)}{lamda(before)} \qquad \text{Eq (3.4)}$$

3.3 Classification

Classification is the process by which it is determined to which group a case belongs. Classification functions are used for this purpose and they are of the form:

$$C.Score_i = \text{constant}_i + w_{i1} \times x_1 + w_{i2} \times x_2 + \ldots + w_{im} \times x_m \qquad \text{Eq (3.5)}$$

where subscript *i* denotes the respective group. The subscripts *1, 2, ..., m* denote the m predictor variables; constant_i is a constant for the *i'th* group, w_{ij} is the weight for the *j'th* variable in the computation of the classification score for the *i'th* group; x_j is the observed value for the respective case for the *j'th* variable. C.Score is the resultant classification score. A case can be classified into the group for which it scores the highest classification score. To get the posterior probability (the probability, based on knowledge of the values of other variables, that the respective case belongs to a particular group), the concept of Mahalanobis distance needs to be introduced.

For each group, we can determine the location of the point that represents the means for all variables in the multivariate space defined by the variables in the model. These means are generally called group centroids. The Mahalanobis distance is defined as a measure

of the distance between two points in the space defined by two or more correlated variables. The difference between Mahalanobis distance and normal Euclidean distance is that the former accounts for uncorrelated distances in more than three dimensions whereas the latter is only a true distance measure for uncorrelated dimensions up to three dimensions. After computing the Mahalanobis distances for each case, each can be classified according to how close it is to a group centroid. Thus, the smaller the Mahalanobis distance, the closer the case is to the group centroid and the more likely it is to be classed as belonging to that group. Posterior classification probabilities are proportional to the Mahalanobis distances for each case. An additional factor that needs to be considered when classifying cases is *a priori* probability. If the prevalence of cases in a particular group is already known then this can be factored into classification.

3.4 Cross validation

Cross validation is the process of assessing the predictive accuracy of a model based on new sample relative to the accuracy on the sample the model was produced from. Classification of cases on the sample the functions were modelled from can bring about erroneously good results, which might not be accurate representations of the predictive power of the model. The usual procedure is then to divide the sample into two. The first sample will be used to build the classification function and test it through classification. The second will be used in cross validation of the model. It is common to refer to the sample used in the function building as the "Analysis" sample and the cross validation sample as the "Hold out" sample.

3.5 Assumptions in discriminant analysis

- **Normality assumption**

  The predictor variables must be multivariate normal. DA is generally robust against violation of this assumption if the smallest group has more than 20 cases and the number of independents is fewer than six.

- **Homogeneity of variances /co variances**

  The variance/covariance matrices of variables should be homogeneous across groups. For the same predictor variable, the groups formed by the dependent variable should have similar variances and means on that independent variable. Small violations of this assumption do not greatly affect accuracy Lachenbruch (1977) but DA is very sensitive to outliers and violation of this assumption might imply that some outliers exist.

- **Other minor assumptions**

  The variables that are used to discriminate between groups are not completely redundant; all cases must belong to only one group on the dependent variable; group sizes of the dependent variable should not differ too much.

3.6 Types of discriminant analysis

Forward stepwise DA builds, step-by-step, a discriminant model based on the predictor variables. Specifically, at each step all variables are reviewed and evaluated to determine which one will contribute most to the discrimination between groups. That variable will then be included in the model, and the process starts again. In reverse stepwise DA all variables are included in the model and then, at each step, the variable that contributes least to the prediction of group membership is eliminated. Thus only those that contribute the most to the model are retained. A model based on all the predictor variables can also be obtained, but it may contain many redundant predictor variables which might not make any significant difference in the discriminating power of the model.

4 DISCRIMINANT ANALYSIS APPLICATION

4.1 Method and materials

The aim of this study was to classify subjects into one of two groups using a discriminant prediction equation; to assess the relative importance of the predictor variables in classifying the dependent variable; and to discard predictor variables that are not significant to group distinctions. In the study the dependent variable was the diagnosis that subjects received after a clinical evaluation. It was thus dichotomous on "FAS positive" and "normal". Two analyses were performed. The first analysis included occipital frontal circumference (OFC), inner canthal distance (ICD), palpebral fissure length (PFL) and inter-pupillary distance (IPD) as the predictor variables under investigation. The second study included the all the above and also a mouth measurement of upper lip circularity (ULC). The five measurements were selected on the basis of their current use in diagnosis of the FAS phenotype, but were limited to those measurements, which could be made from the stereo-photogrammetry images by mouse click. Inclusion of ULC limited the number of subjects who could be included in the second study because of the poor picture quality of some images; it was not possible in all the images to identify the corners of the mouth in an effort to obtain a measurement for upper lip thinness. Images of four of the twenty-four FAS subjects could not be included in the second study because the subjects did not have their lips closed as recommended by dysmorpholists when taking measurements of the upper lip.

4.2 Study population

The study population consists of data obtained during the screening of first-grade children from disadvantaged communities in the Gauteng and Northern Cape Provinces of South Africa for FAS. The children had their height, weight and head circumferences measured. Only those children presenting signs of growth retardation were then evaluated for FAS. Each child was appraised by three independent dysmorphologists. Fifty-six normal subjects, 24 FAS subjects and 1104 growth-retarded but FAS negative subjects were photographed using the stereo-photogrammetric tool described in chapter 2.2 (Meintjes, et al., 2002). Of those evaluated, those with a clinical diagnosis of FAS and normal controls (i.e. no growth retardation and no FAS), were used in this study. The FAS diagnostic criteria in South Africa consists of the presence of facial dysmorphology, growth retardation, developmental delay and a positive maternal history of consumption of alcohol during pregnancy (Douglas and Viljoen, 2006). The use of young children in the study supports the belief that diagnosis is more accurate before puberty as the face becomes less specifically anomalous after puberty. "The human body is expected to grow proportionally and predictably and as such the relationship of measurements to each other is expected to be constant" at specific ages, for specific sexes, and for specific ethnicities (Hall, et al., 1989). In this study, an attempt was made as much as possible to balance the population on these three demographs for an accurate comparison of the subjects. The data was pair matched so that the ratio of FAS to normal subjects was approximately 1:2. All the subjects were of mixed ancestry. The descriptive statistics are presented below (Table 4.1).

First Study	FAS (n=20)	NORMAL(n=47)	TOTAL(n=67)
AGE (yr)			
Mean (SD)	7.11 (0.40)	6.56 (0.37)	
Minimum-Maximum	6.6-7.99	6.12-7.91	
GENDER			
Female: n(%)	12 (66)	29(62)	41(61)
Male: n(%)	8(34)	18(38)	26(39
Second Study	***FAS (n=17)***	***NORMAL(n=32)***	***TOTAL(n=49)***
AGE (yr)			
Mean (SD)	7.10 (0.39)	6.52 (0.37)	
Minimum-Maximum	6.7-7.99	6.12-7.91	
GENDER			
Female: n(%)	10 (58)	20(62)	30(61)
Male: n(%)	7(42)	12(38)	19(39)

Table 4.1: Descriptive statistics of study population.

4.3 Feature extraction

The coordinates of the facial features below were obtained by mouse click and these coordinates were used to calculate distance measurements of the ICD, PFL, IPD and ULC in three dimensions (Meintjes, et al., 2002).

Eye landmarks and measurements (Figure 4.1)

- **en**-The right and left endocanthi (medial corners of the eyes); from these the ICD was measured as the distance between the coordinates of the left and the right endocanthi.
- **ex**-The right and left exocanthi (lateral corners of the eyes); from these and the left and right endocanthi, PFL was measured as the distance between the coordinates of the left exocanthion and the left endocanthion or right exocanthion and the right endocanthion.

- The centres of the left and right pupils denoted by **pc**; from these the IPD was measured as the distance between the coordinates of the left and the right pupils.

Mouth landmarks and measurements (Figure 4.2)

- **st**-The stomion; lowest point on the upper lip of the mouth directly below the labiale superius.
- **ch**-The right and left cheilion (corners of the mouth).
- **ls** -The labiale superius.

st, **ch** and **ls** were used to calculate the upper lip circularity as described below

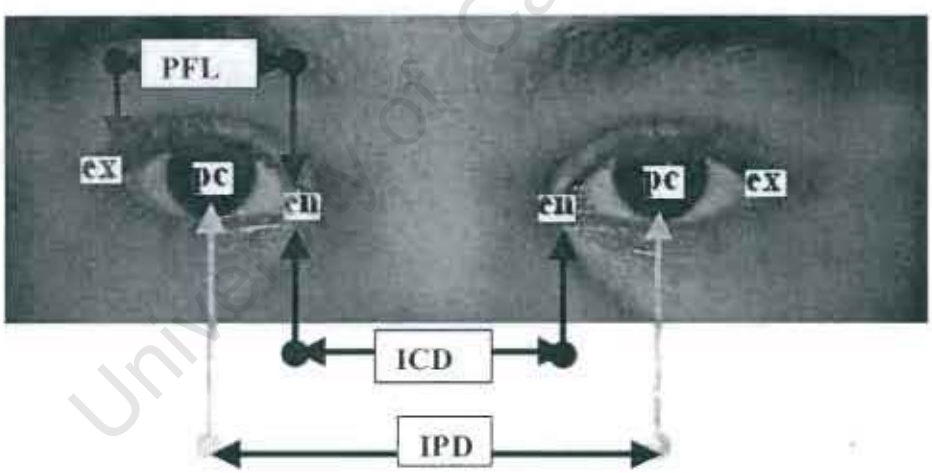
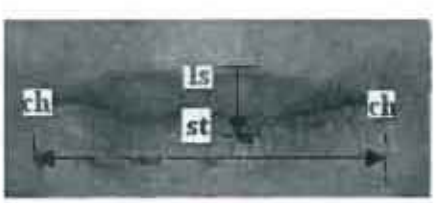


Figure 4.1: The distance measurements made around the eye region.



ls-st = minor axis

ch-ch =major axis

Figure 4.2: The distance measurements made around the mouth region.

Upper lip circularity is used here as an objective measure of shape. The circularity of an object is defined in terms of its area and its perimeter (Astley, et al., 2002).

$$Circularity = \frac{(perimeter)^2}{Area} \qquad \text{Eq (4.1)}$$

The circularity of a circle is 4π, which is the smallest value it can theoretically have. The circularity of an object tends towards infinity as the object's shape becomes more of a line (Astley, et al., 2002). Astley, et al., (2002) obtained the perimeter and area of the upper lip from digital images by outlining the upper lip with a computer mouse on a frontal facial image. In this project the lip region is approximated by a semi-ellipse. The approximation method is faster and easy to implement by mouse click.

Using a semi-ellipse as an approximation to the upper lip shape, we can substitute the area and perimeter of the semi-ellipse to get an estimate of the upper lip circularity. The area and perimeter of an ellipse are given by (Spiegel, 1968):

$$Area = \pi ab \qquad \text{Eq (4.2)}$$

And

$$Perimeter = 2\pi\sqrt{\left(\frac{a^2 + b^2}{2}\right)} \qquad \text{Eq (4.3)}$$

where a is the major axis and b is the minor axis. The stomion (**st**) and the right and left cheilon (**ch**) were used to obtain a (ch-ch) *and* b (ls-st).

The point **st** is obtained by bisecting line **ch-ch** in three dimensions. This bisection neglects the curvature of the mouth and places **st** at the depth of **ch**. Using the three-dimensional coordinates of **st** would therefore result in a distorted circularity because the minor axis (labialli superius to stomion) **ls-st** would be elongated. Depth was thus neglected when measuring **ls-st**. Lastly, the measurement of OFC made using tape during the appraisal of the subjects was included. OFC was included because microcephaly is of importance in the craniofacial appearance of FAS subjects (Moore, et al., 2002). All mouse click measurements for PFL, IPD, and ICD were made three times and then averaged to minimise intra-operator error. After all the measurements were taken they were entered into Statistica v7 (Statsoft.Inc, 2004) for analysis. A forward stepwise DA was done on the data. *A priori* probability for both studies was set to the sizes of the groups in the sample. The significance level used for all tests was p=0.05.

4.3.1 Precision and reliability

To give an indication of the quality of the data, precision and reliability must always be included in every anthropometric study (Jamison and Ward, 1993). Precision is defined as the closeness of measurements of the same quantity on repeating the measurement. The technical error of measurement (*TEM*) provides a measure of precision in the original units of measurements and is given by the formula (Jamison and Ward, 1993):

$$TEM = \sqrt{\frac{\sum d^2}{2n}} \qquad \text{Eq (4.4)}$$

where d is the difference between two measurements of the same quantity taken at different times and n is the number of subjects. A measure of relative variability is the coefficient of relative variation (CRV), which reports the magnitude of error relative to measurement size. CRV is presented as a percentage and given by the formula:

$$CRV = \frac{TEM}{Grandmean} \quad \text{Eq (4.5)}$$

where *Grandmean* is the average of all the measurements in the sample of the same quantity taken at different times.

Intra-observer measurement is an indication of the consistency with which the same observer measures the same subject at different times, and can be assessed by calculating the intra-class correlation coefficient of reliability R. R is given by the formula (Jamison and Ward, 1993):

$$R = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} \quad \text{Eq (4.6)}$$

where T represents individual "error-free" scores (the mean of time 1 and time 2 measurements) and e is the difference between time 1 and time 2 measurements for each individual. This measure of reliability accounts for the two components of variability among a series of measurements; variability among their steady-state values and variability of the random errors (Jamison and Ward, 1993). The higher the value of R the more reliable the measure is.

Douglas and Viljoen (2006) assessed the precision and reliability of PFL, IPD and ICD using data obtained through the stereo-photogrammetric method described in this project. Their study population included 56 six to eight year olds assessed to be healthy in the surveillance effort for FAS described in Section 4.2 and their measurement technique was identical to that used in the project described in this report. Table 4.2 shows the results they obtained. Precision and reliability measurements for upper lip circularity obtained for 49 subjects in the project described here, are also included in Table 4.2.

	Grand mean (mm)	TEM (mm)	CRV%	R
Right PFL	25.6	1.5	5.70	0.78
Left PFL	25.4	1.4	5.53	0.75
Mean PFL	25.5	1.05	4.13	0.84
IPD	55.1	0.7	1.25	0.98
ICD	30.4	0.8	2.68	0.97
ULC	23.2	0.7	3.14	0.96

Table 4.2: The intra-operator technical error of measurement (*TEM*), coefficient of relative variation (*CRV*) and correlation coefficient *R* values in 56, six to eight year old, black South African subjects (Douglas and Viljoen, 2006). ULC was obtained from 49 subjects in this project (see second study, Section 4.4.2). The ULC measurement is a ratio and therefore has no units.

4.4 Results of discriminant analysis

4.4.1 FAS vs. normal using predictor variables of PFL, IPD, OFC and IPD

The data set was divided randomly into two groups with approximately two thirds in the analysis sample and one-third in the hold out sample. The analysis sample had 42 subjects; 12 with FAS and 30 normal subjects. The holdout sample had 25 subjects; 8 with FAS and 17 normal subjects. The results of the discriminant function analysis are summarised in Table 4.3.

Discriminant Function Analysis Summary (FAS vs. normal) No. of variables in model: 3 **Overall Wilk's lambda: 0.20890 , p<0.0000** Included cases: 1:42			
	Wilk's lambda after adding variable	**Partial lambda**	**p-level**
OFC (cm)	0.345519	0.604600	0.000014
PFL (mm)	0.349708	0.597358	0.000011
IPD (mm)	0.240302	0.869329	0.021915

Table 4.3: The results of the discriminant analysis. The values in bold are the overall Wilk's lambda after adding all the significant variables and the *p*-value.

The three predictor variables included in the model are PFL, OFC and IPD. Addition of ICD was not significant. (Overall Wilk's lambda= 0.20850 when ICD is included). The partial lambda indicates that the predictor variable PFL contributes the most, OFC the second most and IPD the least in the overall discrimination. The significance of the discriminant function is tested by the chi-squared test. From Table 4.4 it can be seen that the function is highly significant, p=0.000000.

Chi-Square Test Included cases: 1:42			
Eigen-value	**Wilk's lambda**	**Chi-Squared**	**p-level**
3.161439	0.240301	55.60858	0.000000

Table 4.4: The chi-squared test for significance of the discriminant model.

After variable selection by a stepwise process using Wilk's lambda, the discriminant function coefficients obtained are shown in Table 4.5

Standardised Coefficients Included cases: 1:42	
	Root 1
OFC (cm)	-0.741021
PFL (mm)	-0.976175
IPD (mm)	0.575663
Eigen-value	3.786956
Cumulative Proportion	1.000000

Table 4.5: The standardised coefficients showing how the variables contribute in the discrimination.

The standardised coefficients, which pertain to comparable scales, indicate that the discrimination between the two groups is weighted heavily on PFL and OFC although IPD also makes a contribution. Statistica v7 readily computes the classification functions to be used directly in classifying the cases.

Using the classification functions to compute the classification scores for each group there was 100% specificity and 100% sensitivity classifying the analysis sample. Specificity is the percentage of correctly classified FAS subjects and sensitivity is the percentage of correctly classified normal subjects. Cross validation involved application of the above discriminant model to the hold out sample using the classification coefficients obtained from the analysis sample. Application to the hold-out sample classified the subjects with 86% specificity and 100% sensitivity. There was one FAS misclassification and no misclassification of normal cases.

A DA was performed for all 67 subjects to determine the overall discriminant equation that could be used as a case definition for FAS in the sample. The overall discriminant function is shown below. It should be noted that the significance levels do not reflect the true alpha error rate, that is, the probability of erroneously rejecting H_0 (the null

hypothesis that there is no discrimination between groups). This is because stepwise DA capitalizes on chance because the F to enter and F to remove values, which are used to enter/remove a variable, are chosen arbitrarily. Inclusion of ICD still does not affect the overall results of the analysis. The discriminant equation derived across all 67 subjects was as follows:

$$\textit{D. Score} = 35.71743 - 0.55104\ (OFC) - 0.69307\ (PFL) + 0.16932\ (IPD) \quad \text{Eq (4.7)}$$

A *D.Score* of greater than 1.230 was the cut-off for classifying a subject as having FAS on the basis of the screening (sensitivity =95% and specificity =100%). There was one misclassification of FAS and no misclassification of normal subjects. Eq (4.7) reflects the most objective, sensitive and specific discriminant equation explaining 100% of the total variance. Table 4.6 shows the squared Mahalanobis distances between the two groups. This is a measure of the difference in *D.Score* means of the two groups. The closer the case is to a group centroid, the more confidence that it belongs to that group; the high difference in group means indicates high predictive power of the model.

Squared Mahalanobis Distances		
	normal	**FAS**
normal	0.00000	17.67246
FAS	17.67246	0.00000

Table 4.6: Squared Mahalanobis distances p=0.000000.

The distribution of *D.Scores* among the 67 subjects with and without FAS based on the discriminant equation above is shown in the histogram in Figure 4.3 below. The misclassified FAS subject is also highlighted in the histogram.

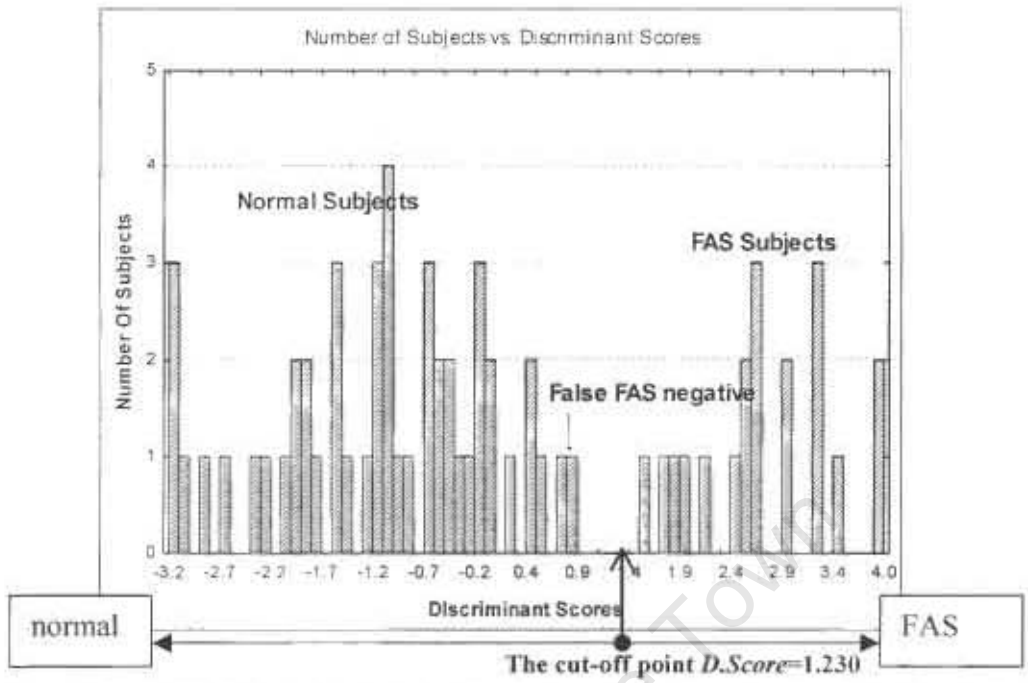


Figure 4.3: Distribution of *D.Scores* among all the 67 subjects based on the discriminant equation Eq. (4.7) above.

4.4.2 FAS vs. normal using predictor variables of PFL, IPD, OFC, IPD and ULC

In the second analysis there was an inclusion of ULC as a predictor variable. Some of the images could not be used as their quality made it impossible to accurately measure ULC. The data set was again divided randomly into two groups with approximately two thirds in the analysis sample and one-third in the hold out sample. The analysis sample had 29 subjects; 10 with FAS and 19 normal subjects. The holdout sample had 20 subjects; 7 with FAS and 13 normal subjects. The results of the DA are shown in Table 4.7.

Discriminant Function Analysis Summary (FAS vs. normal) Number of variables in model: 4; **Wilk's lambda: 0.13688, p<0.0000** Included cases: 1:29			
	Wilk's lambda after adding variable	**Partial lambda**	**p-level**
OFC (cm)	0.309252	0.442627	0.000012
PFL (mm)	**0.235353**	**0.581610**	**0.000356**
IPD (mm)	**0.154516**	**0.885887**	**0.091450**
ULC (mm)	**0.153330**	**0.892738**	**0.102417**

Table 4.7: The results of the discriminant analysis. The values in bold are the overall Wilk's lambda after adding all the significant variables and the *p*-value.

The four predictor variables included in the model are PFL, OFC, IPD and ULC. Addition of ICD did not change the overall Wilk's Lambda indicating that ICD is not significantly relevant in discriminating the two groups. (Overall Wilk's Lambda= 0.13688 when ICD is included). The partial lambda values show that OFC is the most important contributor to the model. PFL is the second most important followed by IPD then ULC. The discriminant model is highly significant as shown in Table 4.8.

Chi-Square Test Included cases: 1:29			
Eigen-value	**Wilk's lambda**	**Chi-Squared**	**p-level**
6.305486	0.136883	49.71564	0.000000

Table 4.8: The chi-squared test for significance of the discriminant model.

The discrimination between the two groups is weighted heavily on OFC, PFL and IPD although ULC also makes a contribution. This is summarized in Table 4.9.

Standardised Coefficients Include cases: 1:29	
	Root 1
OFC (cm)	-0.890533
PFL (mm)	-0.960120
IPD (mm)	0.519178
ULC (mm)	-0.399334
Eigen-value	6.305486
Cumulative Proportion	1.000000

Table 4.9: The standardised coefficients showing how the variables contribute in the discrimination.

Cross validation involved application of the above discriminant model to the hold out sample using the classification coefficients obtained from the analysis sample. Application to the hold-out sample classified the subjects with 85.7% specificity and 100% sensitivity. There was one FAS misclassification and no misclassification of normal cases. A *D.Score* of greater than 0.635 was the cut-off for classifying a subject as having FAS on the basis of the screening (sensitivity =94% and specificity =100%) of all 49 subjects using Eq 4.8.

$$D.Score = 41.42437 - 0.64788\,(OFC) - 0.64522\,(PFL) + 0.14720\,(IPD) - 0.04312(ULC) \quad \text{Eq (4.8)}$$

The overall accuracy = 98%. There was one misclassification of FAS and no misclassification of normal subjects out of the 49 subjects. The group centroids or distances between group means (Table 4.10) are again very high showing a large distinction between the two groups.

Squared Mahalanobis Distances		
	normal	Yes
normal	0.00000	17.47718
FAS	17.47718	0.00000

Table 4.10: Squared Mahalanobis distances with p=0.000000.

The distribution of *D.Scores* among the 49 subjects with and without FAS based on the discriminant equation above is shown in the histogram (Figure 4.4). The misclassified FAS subject is also highlighted in the histogram. The misclassified subject was the same subject in both analyses (see Figure 4.3 and 4.4)

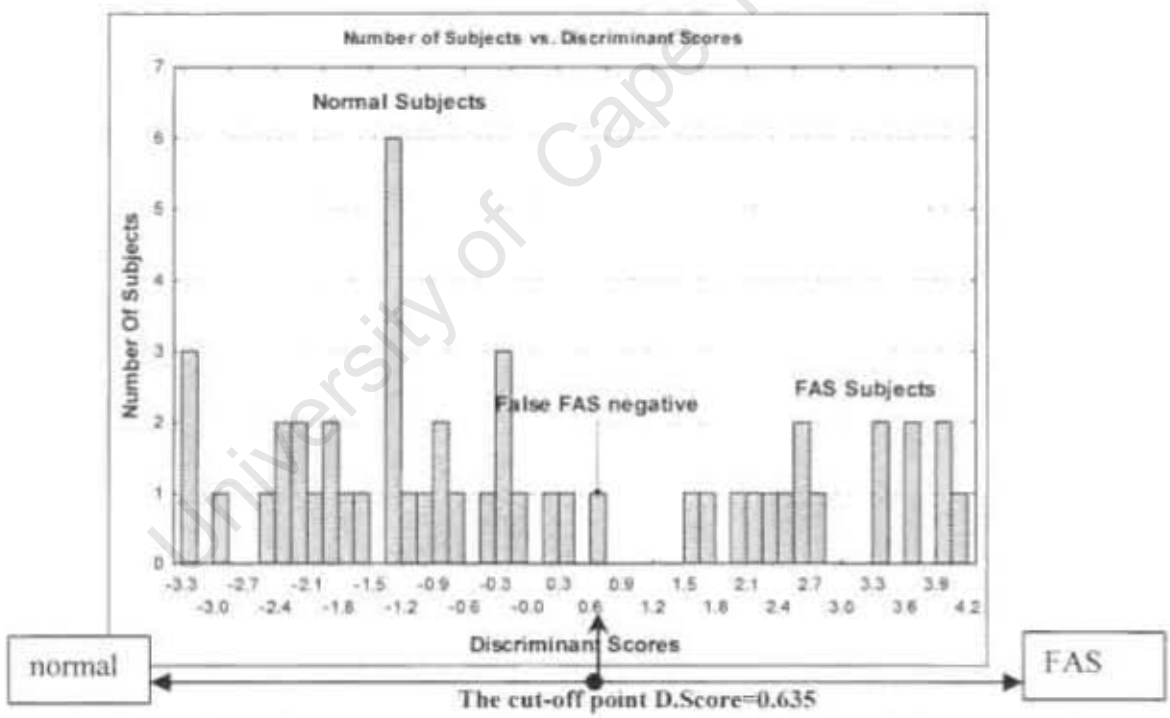


Figure 4.4: Distribution of *D.Scores* among all the 49 subjects based on the discriminant equation Eq. (4.8) above.

4.5 Discussion

Stereo-photogrammetric measurements of PFL, IPD and ICD are more reliable (have higher R) than direct anthropometric measurements of the same quantities (Douglas and Viljoen, 2006). There is no reliability measurement for ULC in the literature although the intra-operator correlation coefficient R obtained for ULC (0.96) suggests a high level of precision. Douglas and Viljoen (2006) found the precision (TEM) of PFL and ICD is less than that published in the literature and suggest that the lower precision of these measurements using stereo-photogrammetry is due to difficulty in the accurate location of endocanthion in the stereo image pairs resulting in depth errors not present in direct anthropometry. As the pupils are generally on the same plane, depth error is minimal and the precision of IPD is higher using stereo-photogrammetry than measured directly. Again there is no published precision measurement in the literature for ULC although the CRV is sufficiently low (<5%, see Table 4.2) to suggest a reasonable precision. Ward and Jamison (1991) suggest a CRV threshold of < 5% as being sufficiently precise.

The results of the first and the second analyses compare very well. Children with FAS can thus be segregated from normal subjects by direct measurement of OFC, and stereo-photogrammetric measurements of IPD and PFL. Microcephaly in FAS subjects is confirmed by the high discriminatory power of the OFC measurement. It should be noted, however, that the FAS sample was derived from a larger sample screened on the basis of growth retardation, specifically, body mass, physical stature and head circumference were below the 10^{th} centile as published by the National Centre for Health Statistics in the United States (Douglas and Viljoen, 2006). The high discriminatory

power of OFC was thus expected. OFC was emphasized by Moore, et al., (2001) who discriminated between subjects with FAS and a control sample with 100% accuracy based on head circumference and bigonial breadth. In Moore, et al., (2002) they go on to suggest that the commonly used clinical descriptors of long philtrum and wide spaced eyes often employed in the diagnosis of children with FAS must be seen in relative terms. That is, features such as small PFLs and IPDs are not measurably different from normal but appear so in relation to the overall reduction in cranial size.

The significance of PFL as suggested by its partial lambda values corroborates the work of Astley and Clarren (1995; 1996). They reported PFL as one of three facial features able to correctly classify FAS cases. The suggestion that the wide spaced eyes commonly associated with FAS are due to small PFLs is thus confirmed. Moore, et al., (2002) suggest that it is not that the eyes are widely spaced out (i.e. ICD is not enlarged) but that the palpebral fissures do not extend out as far laterally and thus biocular breadth is affected as well. In agreement with this suggestion, we have found that ICD is not important in FAS diagnosis and it is also independent of head circumference (Table 4.11).

The other facial features indicated in the literature to be useful in FAS diagnosis are smooth philtrums and thin upper lips. From the results presented here which could not include an assessment of philtrum smoothness (this could not be done using linear distances), upper lip thinness is not as significant in discriminating between FAS and normal subjects as can be shown by its partial lambda value and *p*-value in Table 6.5. IPD seems to be a more important linear measurement in FAS discrimination from the

results of both studies than upper lip thinness. This might seem to contradict the work of Astley and Clarren (1996) but it should be noted that there is a significant positive correlation of IPD and PFL with head circumference whereas upper lip thinness seems to be independent of head circumference as shown in Table 4.11.

Correlations (FAS vs. normal)					
	OFC (cm)	IPD (mm)	ICD (mm)	PFL (mm)	ULC (mm)
OFC (cm)	1.00	0.45	-0.07	0.65	-0.06
IPD (mm)	0.45	1.00	0.46	0.70	0.12
ICD (mm)	-0.07	0.46	1.00	-0.19	0.05
PFL (mm)	0.65	0.70	-0.19	1.00	0.13
ULC (mm)	-0.06	0.12	0.05	0.13	1.00

Table 4.11: Correlation matrix using data from the second study population of 49.

Growth retarded children who are not FAS positive are also expected to display microcephaly. The correlation of PFL and IPD with OFC was not explored in such a population as they were not part of the sample used in the analysis. Future work should be done to ascertain if PFL and IPD reduction correlates similarly with the reduction of cranial size in growth retarded individuals without FAS as in FAS individuals.

The same FAS subject was misclassified in both the first and second study. The range of variability displayed in the facial anomalies of individuals subjected to prenatal alcohol exposure might explain the misclassification of the subject i.e. the subject presented growth retardation, CNS deficiencies and was confirmed to have been subjected to prenatal alcohol exposure but the facial abnormalities they displayed were closer to normal than they were to a typical FAS facial phenotype.

5 FACIAL SHAPE ANALYSIS THEORY

5.1 Overview

Quantitative descriptions of the FAS phenotype to date have been based on linear measurements. This chapter outlines the theory of shape analysis which will be used in a later chapter to distinguish between subjects with FAS and normal subjects. Shape is defined as all the geometrical information that remains when location, scale and rotational effects are filtered out from an object (Dryden and Mardia, 1998). Two shapes can be compared by adjusting for size and superimposing one shape on the other using corresponding landmarks as references. The differences that remain are then due to shape dissimilarity. Dryden and Mardia (1998) discriminate 3 subgroups of landmarks:

- **Anatomical landmarks** are points assigned by an expert that correspond between organisms in some biologically meaningful way.
- **Mathematical landmarks** are points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extreme point.
- **Pseudo-landmarks** are constructed points on an object either on the outline or between landmarks.

A two-dimensional example is illustrated in Figure 5.1, which shows 4 landmarks for two shape images, a rectangle and a trapezoid. The square (black) and circular (grey) points can be represented by two-dimensional coordinates and they can be considered to be corresponding between the two shapes as the corners of four sided shapes. For higher-

dimensional data, the mathematical details are more complex, but the concepts involved remain the same. The alignment of the two shapes involves translating one of the shapes, rotating it relative to the other and then scaling both shapes by normalizing both of them to some size metric.

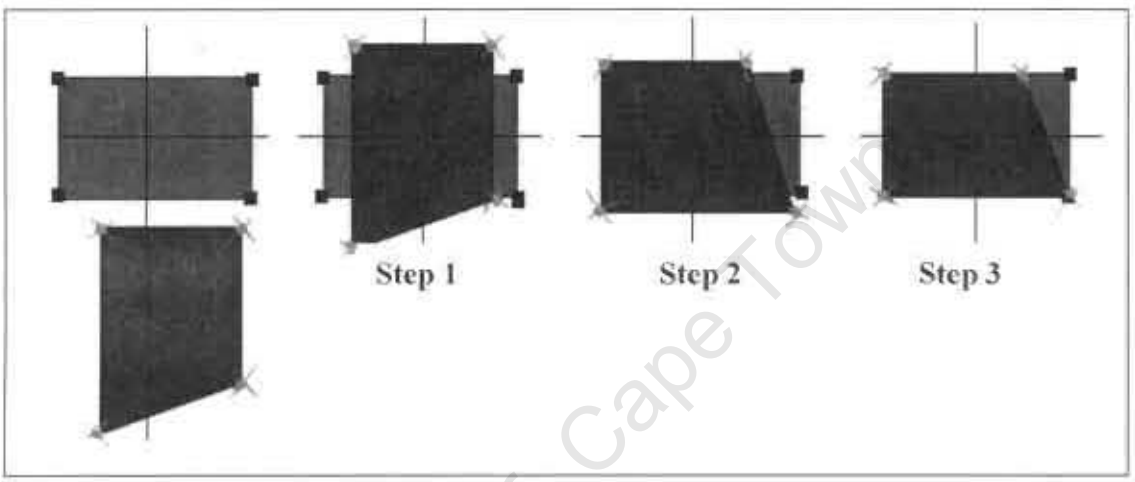


Figure 5.1: Two shapes, a rectangle and trapezium are superpositioned to assess the difference in their shapes.

5.2 Procrustes analysis

Two ways of describing shape are by locating a finite number of points on the shape outline or through landmarks. An *n*-point/landmark shape in *k*-dimensions can be mathematically represented by concatenating each dimension into a *(kxn)* vector. By establishing a coordinate reference with respect to position, scale and rotation, true shape representation can thus be obtained. The coordinate reference or pose aligns or superimposes all the shape objects in question. One way of obtaining such a coordinate reference is through Procrustes analysis (PA). The Procrustes method for comparing

shapes is a method used to fit all n landmark points for N objects with optimal superimposing of landmarks. Optimum superposition of shape objects is achieved when translation and rotation effects are adjusted so as to minimize the distances between landmarks (Halazonetis, 2004). Various minimization criteria exist but the most popular is that which minimizes the sum of the squared distances between corresponding points (Halazonetis, 2004). The process of aligning the shapes using Procrustes alignment involves four steps. The following description is based on work by Stegmann and Gomez (2002) and extended to three dimensions :

1) Calculate the centroid size of each shape

The centroid of a shape is the centre of mass of a physical system consisting of unit masses at each landmark. This can be calculated in three dimensions as

$$(\bar{x},\bar{y},\bar{z}) = \left(\frac{1}{n}\sum_{j=1}^{n} x_j, \frac{1}{n}\sum_{j=1}^{n} y_j, \frac{1}{n}\sum_{j=1}^{n} z_j\right) \qquad \text{Eq (5.1)}$$

If the squared distances from the centroid to all landmarks are summed, and the square root taken the result is the centroid size:

$$S(x) = \sum_{j=1}^{n} \sqrt{(x_j-\bar{x})^2 + (y_j-\bar{y})^2 + (z_j-\bar{z})^2} \qquad \text{Eq (5.2)}$$

2) Each shape is normalised by dividing by the centroid size.

3) Each of the shapes is then aligned with respect to position at their centroids.

4) Each shape is then realigned with respect to rotational orientation about their centroids.

An algorithm developed by Bookstein (1997) to obtain the new coordinates of the shapes, after alignment is complete, is presented below (these coordinates are called the full Procrustes fits and defined below):

1. Choose an initial estimate of the mean shape or configuration. The initial shape in the set will do.
2. Align all the remaining shapes to the mean shape using the Procrustes alignment as described above.
3. Re-calculate the estimate to the mean from the aligned shapes.
4. If the estimate mean has changed, return to step 2.

Convergence is declared when the mean shape does not change significantly within an iteration (Stegmann and Gomez, 2002).

For a sample of landmark configurations S_i where $i=1,...N$ (number of shapes); the Procrustes mean shape $\overline{\alpha}$ after convergence can be estimated by:

$$\overline{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i \qquad \text{Eq (5.3)}$$

to give the Procrustes mean coordinates:

$$(\overline{\alpha}_{jx}, \overline{\alpha}_{jy}, \overline{\alpha}_{jz})$$

where $j=1,...n$ (number of landmarks)

The full Procrustes fit coordinates S_i^P are found by fitting each shape S_i to the Procrustes mean $\overline{\alpha}$ using, for example, a least squares method (Halazonetis, 2004), and thus each S_i^P has coordinates:

$$(s_{i_{jx}}^P, s_{i_{jy}}^P, s_{i_{jz}}^P)$$

Procrustes residuals S_i^{PR} (Robinson, et al., 2001) are the difference between the full Procrustes fit coordinates S_i and the Procrustes mean coordinates $\overline{\alpha}$, and are here represented as:

$$(s_{i_{jx}}^{PR}, s_{i_{jy}}^{PR}, s_{i_{jz}}^{PR})$$

When variation in shape is reasonably small, we can project the Procrustes fit shapes into a Euclidean space called the tangent space. The Procrustes residuals can be used as a good approximation to the Procrustes tangent coordinates. Tangent space projection makes linear assessment of the shape space possible. Procrustes residuals can be used in principal component analysis to explore shape variability (Hennessy, et al., 2002).

5.3 Variation of shape

The variability of shape around the mean can be viewed as scatter plots or connected scatter plots for easier visual inspection. The position of each point can vary along all orthogonal axes. As such, *(n×k)* variables can describe the variability of the object. Various statistical procedures can be carried out to investigate shape variation. One of them is principal component analysis (PCA). PCA is a procedure that can decrease or summarise the *(n×k)* variables into uncorrelated linear combinations of these variables. Principle component (r) can be given by

$$PC(r) = \sum_{j}^{n} (\gamma_{rjx} s_{jx}^{PR} + \gamma_{rjy} s_{jy}^{PR} + \gamma_{rjz} s_{jz}^{PR}) \qquad \text{Eq (5.4)}$$

where the weights $\gamma_{rjx}, \gamma_{rjy}$ and γ_{rjz} describe the directions of variation in each Procrustes coordinate about the mean shape (Robinson, et al., 2001) .

A set of the same class shape of biological objects will always have some degree of inter-point correlation (Stegmann and Gomez, 2002), especially since they belong to the same biological entity (Halazonetis, 2004). There therefore exists a shape representation between points that accounts for the correlation between points. This can be used to reduce dimensionality. There are as many principal components as there are landmark coordinates. Thus in this study where 20 landmark coordinates are measured in three dimensions, the number of principal components (PCs) is equal to 60. Using various cut-off criteria, a dimensionality of less than 60 can be obtained which can still account for

as much variance in shape as desired. The different sets of variables called principal components obtained have the following properties (Halazonetis, 2004):

- All components are orthogonal to each other and so are statistically unrelated.
- Part of the variability of the sample is represented by each component in decreasing order, starting from the largest variability being represented by the first component and the second largest variability by the second and so on.
- Every component is linear combination of the original variables.

The decreasing order of significance of the components in accounting for variability means that only some need be retained to account for a significant part of shape variability. The average shape of a sample has, by definition, all principal components equal to 0. Thus, to visualise the pattern shape variability represented by each principal component, the average shape can be warped by moving points according to the weights on the principal component. For *PC(r)* shapes with the following coordinates can be plotted:

$$(\bar{\alpha}_{jx} + c\sqrt{\lambda_r \gamma_{rjx}},\ \bar{\alpha}_{jy} + c\sqrt{\lambda_r \gamma_{rjy}},\ \bar{\alpha}_{jz} + c\sqrt{\lambda_r \gamma_{rjz}})$$

where λ_r is the variance captured by *PC(r)*. The value of c can range about the mean shape by some standard deviation in the negative or positive direction of *PC(r)* (typically $-3 \le c \le 3$ although this varies according to how interpretable, visually, the shape configuration is). Thus the shape configurations between the range spanned by the standard deviation may be obtained (Robinson, et al., 2001).

By setting the feature components (list of most significant components chosen on how much of the total variability they account for) to values of plus and minus some standard deviation the pattern of shape variability can be observed.

Shape is considered to be a continuous variable of smoothly varying patterns and not a discrete variable. The collection of all possible patterns of the same "shape" is called shape space (Halazonetis, 2004; Stegmann and Gomez, 2002). Each shape object or pattern will thus be a particular point in this space. It is useful to consider the principal components as representative of coordinates that uniquely address any shape pattern from the collection in shape space. Scatter plots of principal components against each other can give an indication of where a shape object lies.

6 FACIAL SHAPE ANALYSIS APPLICATION

6.1 Method and materials

Two studies were performed. Both compared facial landmark data of FAS vs. normal subjects. The difference in the two studies was the number of landmarks chosen. The first study included only the landmarks that are used in published work in linear measurements to diagnose the FAS facial phenotype. The second included other landmarks in the mid facial region that are affected by midface hypoplasia. Midface hypoplasia is underdevelopment of the middle of the face characterized by a flat nasal bridge, seemingly widely spaced eyes, an upturned nose, long philtrum and a generally concave-looking face. Midface hypoplasia has been documented as part of the facial anomalies associated with the FAS facial phenotype (Huang, et al., 2005; Clarren, et al., 1987; Astley and Clarren, 1995). The study population for both studies was identical.

6.1.1 Study population

Images of 56 normal and 24 FAS subjects were available from the database. Of these only 20 normal and 14 FAS image pairs were clear enough for measurement. The descriptive statistics of the subjects are shown in Table 6.1 below. Images in which all the features could not be clearly seen because of bad lighting conditions were discarded.

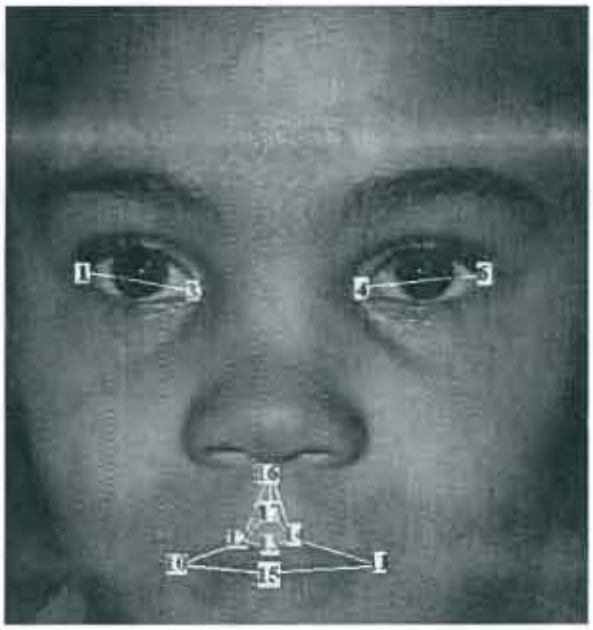
Characteristic	*FAS (n=14)*	*NORMAL(n=20)*
AGE (yr)		
Mean (SD)	7 11 (0.40)	6.56 (0.37)
Minimum-Maximum	6 6-7 99	6 12-7 91

Table 6.1: The descriptive statistics of demographics used in Procrustes analysis of FAS vs. normal subjects.

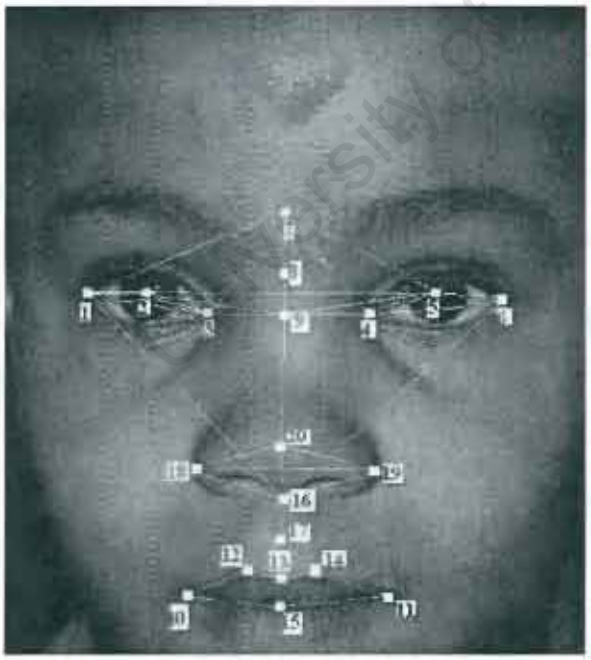
6.1.2 FAS related features

Figure 6.1 below shows the landmarks used in the facial analysis for both studies. From these landmarks the distances and angles in Figure 6.2 were observed. Midface hypoplasia was assessed by looking at landmarks around the middle of the face. Abnormalities in eye orbit region (especially length of the palpebral fissures) were assessed by looking at landmarks in the eye region. Philtrum smoothness and upper lip thinness were assessed by looking at landmarks defining the upper lip region. Connecting the distances, lengths and angles (Figure 6.2) by wire frame produces the shape in Figure 6.1

First Study

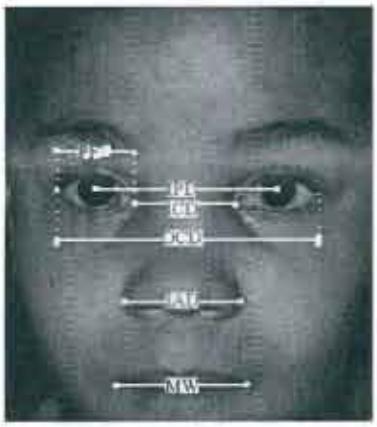
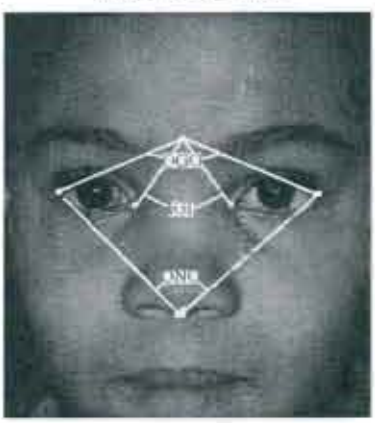


Second study



No	Landmark
1	left outer canthion
2	left pupil center
3	left inner canthion
4	right inner canthion
5	right pupil center
6	right outer canthion
7	glabella
8	nasion
9	sellion
10	left cheilion
11	right cheilion
12	left crista philtri
13	labiale superius
14	right crista philtri
15	stomion
16	subnasale
17	midpoint of philtrum furrow
18	left alare
19	right alare
20	pronasale

Figure 6.1: The facial landmarks used in the Procrustes analysis. The images are the right oriented of a pair of stereo images.

Horizontal Distances Observed 	• **IPD**-Projected horizontal distance between two pupil centers • **PFL**- Projected horizontal distance between the inner and outer canthion • **ICD**-Projected horizontal distance between the two inner canthi • **OCD***-Projected horizontal distance between the two outer canthi • **IAD***-Interaral Distance-Projected horizontal distance between the two aralae • **MW***-Mouth Width-Projected horizontal distance between two cheilion
Vertical Distances Observed 	• **UFH***-Upper Facial Height-Projected vertical distance between nasion and subnasale • **NBL***-Nasal Bridge Length-Projected vertical distance between nasion and pronasale • **PL***-Philtrum Length- Projected vertical distance between subnasale and labiale superius
Angles Observed 	• **OGO***-Outer canthal glabella outer canthal • **IGI***-Inner canthal glabella inner canthal • **ONO***-Outer canthal nasal outer canthal

*Included only in the second study

Figure 6.2: The distances and angles observed. The images are the right oriented of a stereo pair of images.

6.1.3 Feature extraction

The anatomical landmarks whose coordinates were extracted for Procrustes analysis were those that could be easily seen in the image pairs and they were consistent with some of those cited in literature as clinically relevant for FAS diagnosis (Clarren, et al., 1987; Astley and Clarren, 1995; Astley and Clarren, 1996; Moore, et al., 2002). The extraction of coordinates involved clicking on the landmark on both the right and left images displayed on a computer monitor to obtain the x and y coordinates. The X,Y,Z coordinates were calculated automatically as described in chapter 4.2. This procedure was done twice for each subject for assessment of precision and reliability. Software was developed in Matlab V6.50 (Mathworks Inc, 2004) to extract the three-dimensional coordinates of relevant points on the image pair.

Guidelines developed by Farkas (1994) were used to estimate the position of those landmarks that do not lie on extremes of curvature and contours. Those that were difficult to mark due to bad lightning or orientation of the photographs are indicated below.

- **Glabella**-if the subject has thick eyebrows then the glabella is the intersection of the vertical facial midline with the middle of the eyebrows; if the subject has thin, the glabella is the intersection of the vertical facial midline with the top of the eyebrows.
- **Nasion**- intersection of the vertical facial midline with the horizontal midline of the eyelid folds.

- **Sellion** - intersection of the vertical facial midline with the horizontal midline at the level of the endocanthion.
- **Arale** - most protruding point at sides of nose (difficult to measure in stereo images).
- **Subnasale** - most inferior point of nose on the facial midline (difficult to measure unless face is upturned).
- **Pronasale** - estimated on images as the intersection of the perceived midline of the face and the tip of the nose.
- **Midpoint of the philtrum** - halfway between subnasale and labiale superius along the perceived midline of the face.

It is important to note, however, that Procrustes analysis treats all points equally and thus an error in the location of one point does not necessarily significantly affect the whole configuration. In contrast, conventional analyses might result in significant errors in measurement, and therefore diagnosis, as a result of small errors in location of points, especially those that are used as reference points (Halazonetis, 2004).

6.1.4 Analysis of difference in shape

A statistical measure of difference in shape is the Goodall's F test. This test compares the Procrustes distance between the means of the two samples to the amount of variation found within the samples. The test accounts for all the sample variance and assumes that landmarks are symmetrically distributed around the Procrustes mean. This test returns a *p*-value under the null hypothesis of no difference in shape between the two samples.

Rohlf (2000) reported that Goodall's F test provides the highest statistical power compared to other conventional T^2 tests used in morphometrics, especially for small samples.

6.1.5 Software and file formats

The Procrustes superposition was implemented using the program IMP: Simple3 (Sheets, 2004) which is part of the Integrated Morphometrics Package (IMP) series of geometric morphometrics software and PAST- PAlaeontologicalSTatistics, v1.35 (Hammer, et al., 2005). The principal component analysis was performed using the software IMP: ThreeDPCA6 (Sheets, 2003) which is also part of the IMP software series and PAST. The three programs use different file formats. Conversion from one format to the other was performed in Simple3, which has the facility to make the conversions. The extracted landmark data has the format:

$$\begin{bmatrix} Landmark_1 \\ Landmark_2 \\ \ldots\ldots\ldots\ldots \\ Landmark_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 z_1 \\ x_2 y_2 z_1 \\ \ldots\ldots\ldots \\ x_n y_n z_n \end{bmatrix}$$

Where n is the *nth* landmark for each subject. IMP can use this format but PAST and ThreeDPCA6 use the format:

$$[Subject] x_1 y_1 z_1 (Landmark_1), x_2 y_2 z_2 (Landmark_2), \ldots, x_n y_n z_n (Landmark_n), (CentroidSize)$$

Where n is the *nth* landmark for each subject.

The three software packages perform landmark based geometric morphometrics software and are capable of visually displaying the results as coordinates of landmarks connected by wire frame.

Reliability tests and other statistical procedures were performed in Statistica v7.0 (Statsoft.Inc, 2004) and Excel.

6.2 Results of first study

6.2.1 Precision and reliability of landmark extraction

As mentioned in Section 4.3.1 an indication of the quality of the data, should be included in every anthropometric (Jamison and Ward, 1993). Selection of landmarks on the images was done by mouse click on a computer screen. This process is prone to errors since some landmarks were hard to identify on both images. Thus, although this study deals primarily with shape, an analysis of the precision and reliability of landmark extraction was performed. Intra-observer precision and reliability were assessed using the landmarks described for the first analysis. Procrustes analysis does not treat any landmark points as privileged (Halazonetis, 2004), and so a good measurement to use in precision and reliability studies is centroid size. The formulae for technical error of measurement (*TEM*), coefficient of relative variation (*CRV*) and intra-class correlation coefficient *R* were presented in Section 4.3.1. The *TEM* obtained for centroid size in the first analysis using two sets of measurements was 1.2mm (total number of subjects=34). The coefficient of relative variability (*CRV*) was 1.03%. The intra-class correlation coefficient obtained for centroid size was $R=0.98$.

6.2.2 Centroid size

The centroid size of the FAS subjects was smaller than that of normal subjects. This result is statistically significant and is shown below in Table 6.2.

	FAS Mean	normal Mean	t-value	p	Std.Dev. FAS	Std.Dev. normal
Centroid Size (mm)	100.7271	111.3365	-7.72704	0.000000	4.394643	3.596357

Table 6.2: Centroid size mean and standard deviation values. The centroid sizes are smaller for FAS subjects compared to normal subjects. The t-value shows a statistical difference between the FAS and normal centroid sizes.

6.2.3 Comparing difference in shape

The two groups (FAS and normal) were treated separately. The mean configuration for each group was obtained and the two were compared for significance in means. Table 6.3 below shows the results of Goodall's F-test. There is no statistically significant difference between the two shape means.

Between Group Distance (Partial Procrustes distance)	0.00264
F-Score	0.3339
df1	29
df2	928
p	0.99968161

Table 6.3: The results of Goodall's F- test. The *p* value reported supports the null hypothesis that there is no shape difference between the two samples.

No further analysis was done on this set of landmarks as no difference was found between the FAS and normal group means.

6.3 Results of second study

6.3.1 Precision and reliability of landmark extraction

Analysis of reliability and precision of extracted landmark data was performed to gauge the quality of the data before the facial shape analysis was done. The *TEM* obtained for centroid size in the second analysis using two sets of measurements was 1.0mm (total number of subjects=34). *CRV* was 0.69% and the intra-class correlation coefficient obtained for centroid size was R=0.98. These values were satisfactory to justify using either set of measurements as the data for further analysis. The first set of measurements was thus used in the analysis and no averaging was performed.

6.3.2 Centroid size

The centroid size of the FAS subjects was smaller than that of normal subjects (using landmark data from the second study). This result is statistically significant and is shown below in Table 6.4.

	FAS Mean	normal Mean	t-value	*p*	Std.Dev. FAS	Std.Dev. normal
Centroid Size (mm)	131.0139	146.0546	-8.10234	0.000000	5.774957	4.997750

Table 6.4: Centroid size mean and standard deviation values. The centroid sizes are smaller for FAS subjects compared to normal subjects. The t-value shows a statistical difference between the FAS and normal centroid sizes.

To discount age as a factor in any shape difference that would be obtained between the FAS subjects and the normal subjects the correlation coefficient between age and centroid size was obtained. The correlation coefficient matrix between age and centroid size is shown below in Table 6.5.

Correlation matrix (Age and Centroid Size) Significant at p < .05000		
	Age	**Centroid Size**
Age	1.00	-0.27
Centroid Size	-0.27	1.00

Table 6.5: Correlation matrix between age and centroid size. The correlation coefficient is small. It is important to note that the normal subjects are on average younger than the FAS subjects even though the FAS subjects have smaller centroid sizes.

6.3.3 Comparing shape differences

Again in this analysis the two groups (FAS and normal) were treated separately. The mean configuration for each group was obtained and the two were compared for significance in means. Figure 6.3 below shows the mean configurations or shapes of the two groups.

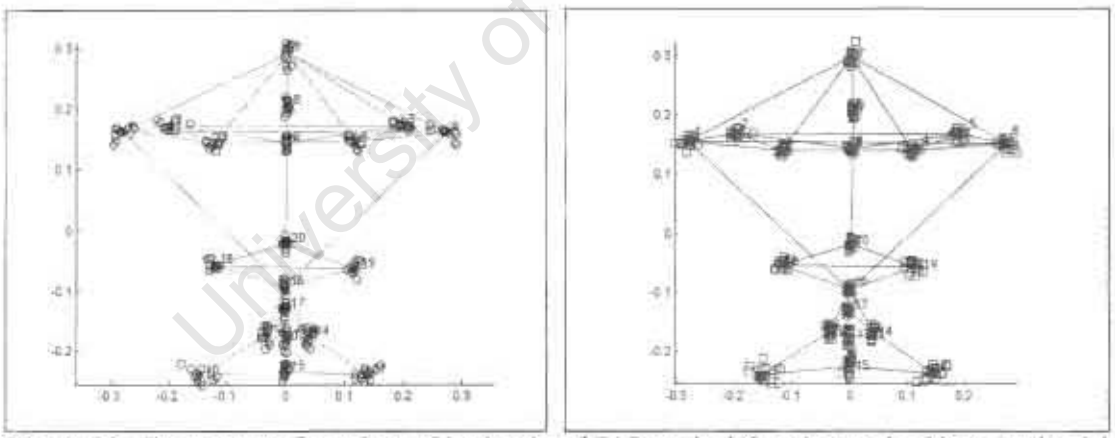


Figure 6.3: The mean configurations of landmarks of FAS on the left and normal subjects on the right. The data points shown are the full Procrustes fits superimposed on the mean. This is the coronal view with Z=0. The numbers correspond to the landmarks in the Figure 6.1 legend for the second study.

Superimposing the two mean configurations (Figure 6.4) allowed a visual comparison for both the coronal view (Z=0) and the sagittal view (X=0).

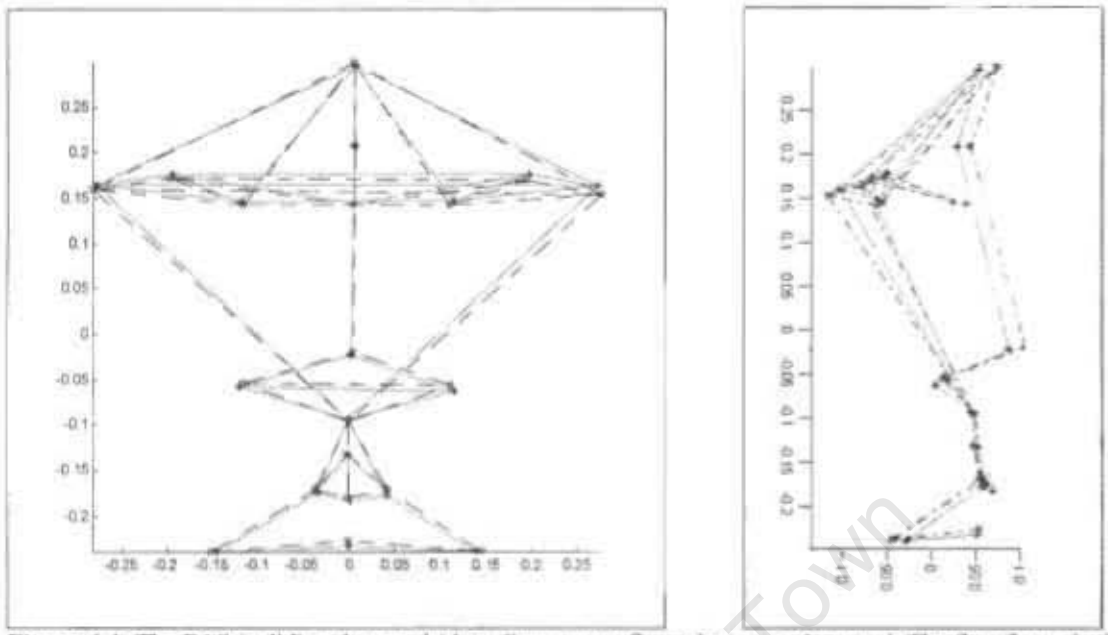


Figure 6.4: The FAS (solid) and normal (dotted) mean configurations superimposed. The first figure is the coronal view. (Z=0) The second view is the sagittal view (X=0). The numbers correspond to the landmarks in the Figure 6.1 legend for the second study.

6.3.4 Statistical analysis of difference in mean

Goodall's F test was used to examine differences in mean shape between the two groups. The results are presented in Table 6.6 below.

Between Group Distance	0.00558
F-Score	2.2295
df1	53
df2	1590
p	0.0000014134001

Table 6.6: Results of Goodall's F- test.

The *p* value reported is very significant; hence there is a clear shape difference between the two mean shapes, suggesting that FAS and normal subjects have different facial shapes. The small between-group distance is the distance between the two group means. The small value implies that although there is a significant difference between the two means the range between the two group extremes is very small.

6.3.5 Principal component analysis

The statistically significant difference in means for the two groups justified further analysis. The Procrustes residuals for all subjects were obtained using IMP software. These were used in principal component analysis (PCA) to assess shape variability. The usefulness of PCA is that the principal components (PCs) can be arranged in order of decreasing importance (how much of the total variance the PC accounts for). There is a number of ways of choosing how many components to retain. Most methods are informal and adhoc (Yeung and Ruzzo, 2000). The easiest method is by subjectively deciding what percentage of variance is necessary and choosing the smallest integer number of PCs such that this percentage is exceeded. Another method is using a scree-plot in which the *n*th eigen-value is plotted against the *k*th component. The component where the plot flattens is the cut-off and the components to its left are retained. The Jolliffe cut-off value for eigen-values is a method to show informally how many principal components should be considered significant (Jolliffe, 1986). This method retains the PCs associated with the covariance matrix that have eigen-values greater in magnitude than the average of all the eigen-values (Cangelosi and Goriely, 2006). PAST- PAlaeontologicalSTatistics, v1.35 (2005) retains a Jollife cut-off value during the analysis. Components with less than this value can be considered insignificant. The Jolliffe cut-off was 0.0001312 for the study sample. The first fifteen PCs had eigen-values greater than this cut-off value and so were considered for further analysis. The cumulative variance proportion of these components is 93.7% as shown in Table 6.7 below.

Principal Component	Eigen-value	% Individual Variance	% Cumulative Variance
1	0.00249	24.3	24.3
2	0.00113	11.1	35.4
3	0.00105	10.3	45.7
4	0.00097	9.4	55.1
5	0.00066	6.4	61.5
6	0.00058	5.7	67.2
7	0.00056	5.5	72.7
8	0.00048	4.7	77.5
9	0.00037	3.6	81.1
10	0.00034	3.3	84.4
11	0.00025	2.4	86.8
12	0.00023	2.3	89.0
13	0.00018	1.8	90.8
14	0.00015	1.5	92.3
15	0.00014	1.4	93.7

Table 6.7: Table of principal components showing percentage variation and eigen-values. The fifteen account for 93.7% of the variation. The total number of PCs is 60.

6.3.6 Discriminant analysis of principal components scores

Some of the shape variation represented by the PCs is responsible for the differences between facial shapes of FAS subjects and normal subjects. A discriminant analysis of the principal component scores for the above selected fifteen PCs was done to see which PCs had the most discriminating power between FAS and normal subjects. The use of discriminant analysis here was to ascertain which PCs had the most distance between group means relative to within-group variation. A visual assessment of the configuration of landmarks shows that the within-group variation of shape is relatively isotropic (see Figure 6.3), i.e. the landmarks for both groups are distributed evenly around the mean configuration. No test to prove this statistic explicitly was performed, although it is known that generally the within-group variation of landmarks of the morphometric data sets is usually very small (Klingenberg and Monteiro, 2005). The results are presented in

Table 6.8 below. Only PC1, PC3 and PC14 had significant discriminating power as reported by their significant *p*-values.

Discriminant Function Analysis Summary
Wilks' Lambda: 0.40349 approx. F (4,29)=10.718 **p<0.0000**

	Wilks' Lambda	Partial Lambda	F-remove	p-level	Toler.	1-Toler.
PC1	0.776302	0.519764	26.79460	0.000016	0.828716	0.171284
PC3	0.492692	0.818956	6.41093	0.017016	0.899160	0.100840
PC14	0.482782	0.835767	5.69867	0.023717	0.907740	0.092260

Table 6.8: Principal components indicated by discriminant analysis on the first 15 principal components to have significant discriminating power.

Scatter plots of PC1, PC3 and PC14 against each other are shown in Figure 6.5, 6.6 and 6.7 below. This method of presentation was used by Halazonetis and Odont (2004), to illustrate principal component differences for cephalometric diagnosis. Each subject is located at a particular point in these scatter plots (which here represent shape space as discussed in Section 5.3). Biological entities with similar shape should generally be located near each other in these scatter plots. PC1 seems to contain the most discriminating power from the scatter plots, i.e. there is greater separation between the two groups in the direction of PC1 (Figure 6.7 and 6.8). The FAS subjects generally lie to the right of the origin in the positive direction and the normal subjects are to the left. Together with the fact that PC1 contributes the most to shape variability it can be deduced that the shape variability along PC1 mostly accounts for the broad differences in shape in between the two groups biologically. PC3 and PC14 concentrate on the finer details in shape difference.

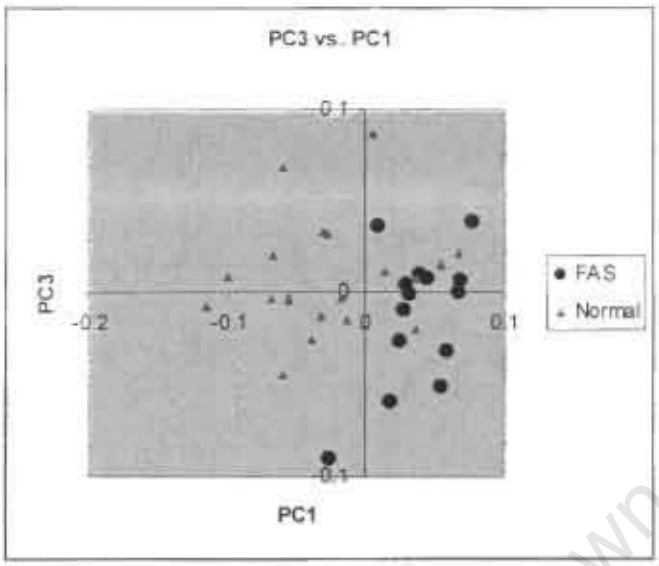


Figure 6.5: Scatter plot of subjects' principal score in the direction of PC1and PC3. It can be seen that FAS subjects generally lie to the right of the origin in the positive direction and the normal subjects are to the left. The orientation of the two groups in the vertical direction of PC3 is more ambiguous.

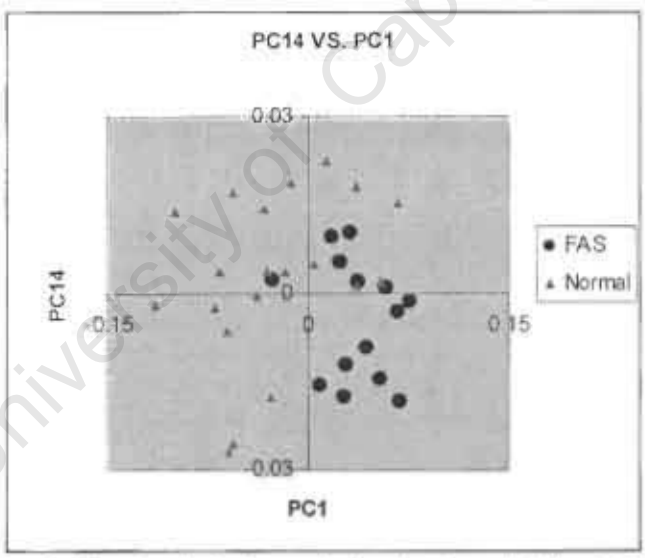


Figure 6.6: Scatter plot of subjects' principal score in the direction of PC1 and PC14. It can be seen that FAS subjects generally lie to the right of the origin in the positive direction and the normal subjects are to the left. The orientation of the two groups in the vertical direction of PC14 is more ambiguous.

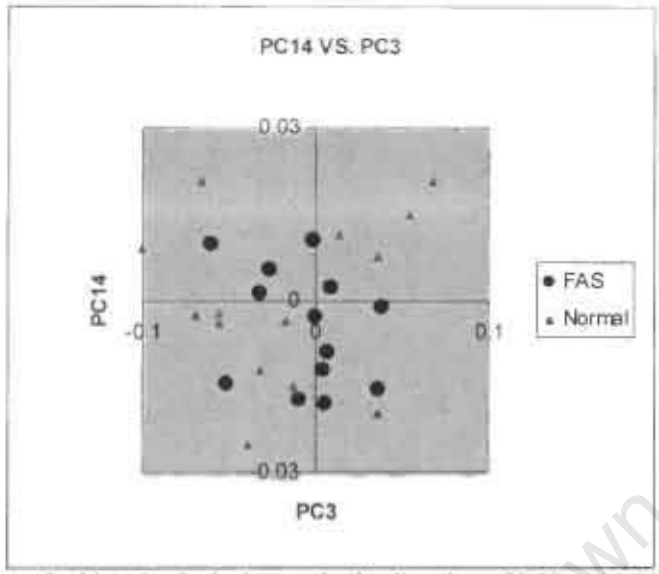


Figure 6.7: Scatter plot of subjects' principal score in the direction of PC3 and PC14. The horizontal axis represents PC3 and the vertical PC14. The range of values for normal subjects spans that of the FAS in the direction of PC14, although there are more normal subjects in the above the origin and more FAS subjects below the origin along PC14. It is also important to note that the variability range is very small (± 0.03 standard deviations for PC14, see Section 5.3).

6.3.7 Visual assessment of shape variability

The principal component loadings of each landmark point along the direction of each PC can be used to assess shape variability. The mean shape has by definition all PCs equal to zero. To illustrate the pattern of shape variability represented by each PC, the average or mean shape can be warped by moving the landmark points according to their loadings in the positive and negative direction of the PC by standard deviations in the negative and positive directions, as shown in Figure 6.8 for PC1. This method of shape visualization has been successfully used for visual assessment of shape (Stegmann and Gomez, 2002; Hennessy and Moss, 2001; Robinson, et al., 2001; Halazonetis, 2004). A description of the shape warping is presented in Table 6.9. The standard deviation range of values chosen in the warping is representative of the range of values in principle

scores, as shown in the scatter plots above. For completeness the warping of PC3 and PC14 is also presented in Figure 6.9. The variations in Figure 6.9 represent finer details about the shape differences between FAS and normal subjects. However, as the scatter plots in Figure 6.5, 6.6 and 6.7 illustrate, it is difficult to pick out the trend in clustering between the two groups along the direction of PC3 and PC14. Thus it is difficult to interpret the warped shapes in terms of how they discriminate between the two groups along the shape mode described by PC3 and PC14. A description of the warped shapes is presented in Table 6.10.

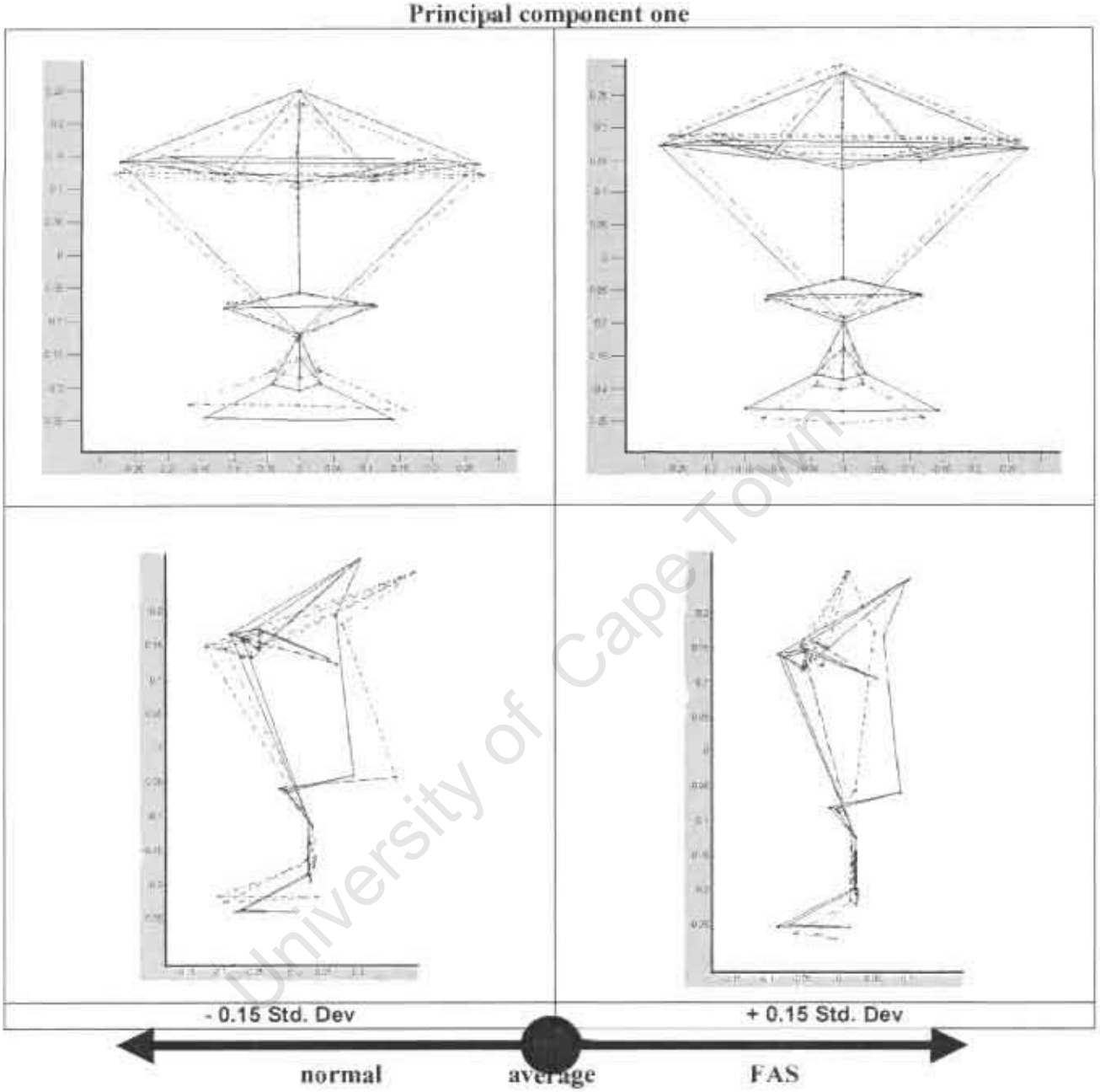


Figure 6.8: Warped shapes depicting the variation in PC1. The average shape is warped by an amount equal to -0.15 standard deviation in the negative direction (left) and 0.15 standard deviation in the positive direction (right). The solid line is the average and the dashed line is the superimposed warped shape.

REGION	PC1 (Mean -0.15 Std.Dev)	PC1 (Mean +0.15 Std.Dev)
Forehead	• Large posteroanterior protrusion of the glabella	• Marked anteroposterior retraction of the glabella to behind the nasion
Eyes	• ICD, OCD, PFL and IPD become slightly larger than the average. • Anteroposterior retraction of eye features	• Small movement in endocanthion laterally and decrease in IPD but great decrease in OCD hence indication of reduced PFL
Midface	• Great increase in nasal protrusion • No change in IAD	• Great reduction in nasal protrusion • No change in IAD
Philtrum/Lips	• Small PL • Great increase in MW but no increase in thickness • Slightly more prominent philtrum furrow	• Large PL • Decrease in MW although lip thickness is unchanged • Slightly smoother philtrum
Angles	• Increase in OGO and ONO but IGI remains the same	• Decrease in OGO and ONO but slight increase in IGI (This again suggests that the eye orbits are decreasing i.e. smaller PFL)
General	• Facial height is decreased • Facial width remains the same (suggesting there is an upper limit to facial width in normal subjects)	• Longer face in terms of vertical height • General flattening of the mid-facial region

Table 6.9: Description of the warped shapes long PC1. Features are described in Figure 6.1and 6.2. Inclusion of both the negative and positive direction of PC1 was necessary to show that some features have a limit to how they vary visually as you go from one extreme to another, e.g. Facial width remains the same as you go towards the negative side of PC1.

While most of the information about shape change in one column is the opposite in the other column, some of the changes that occur in one direction are not necessarily completely the opposite in the other direction. An example is facial width which seems to decrease towards the positive side if PC1 (the FAS direction) but seems to increase to an upper limit when going towards the negative side of PC1 (the normal side).

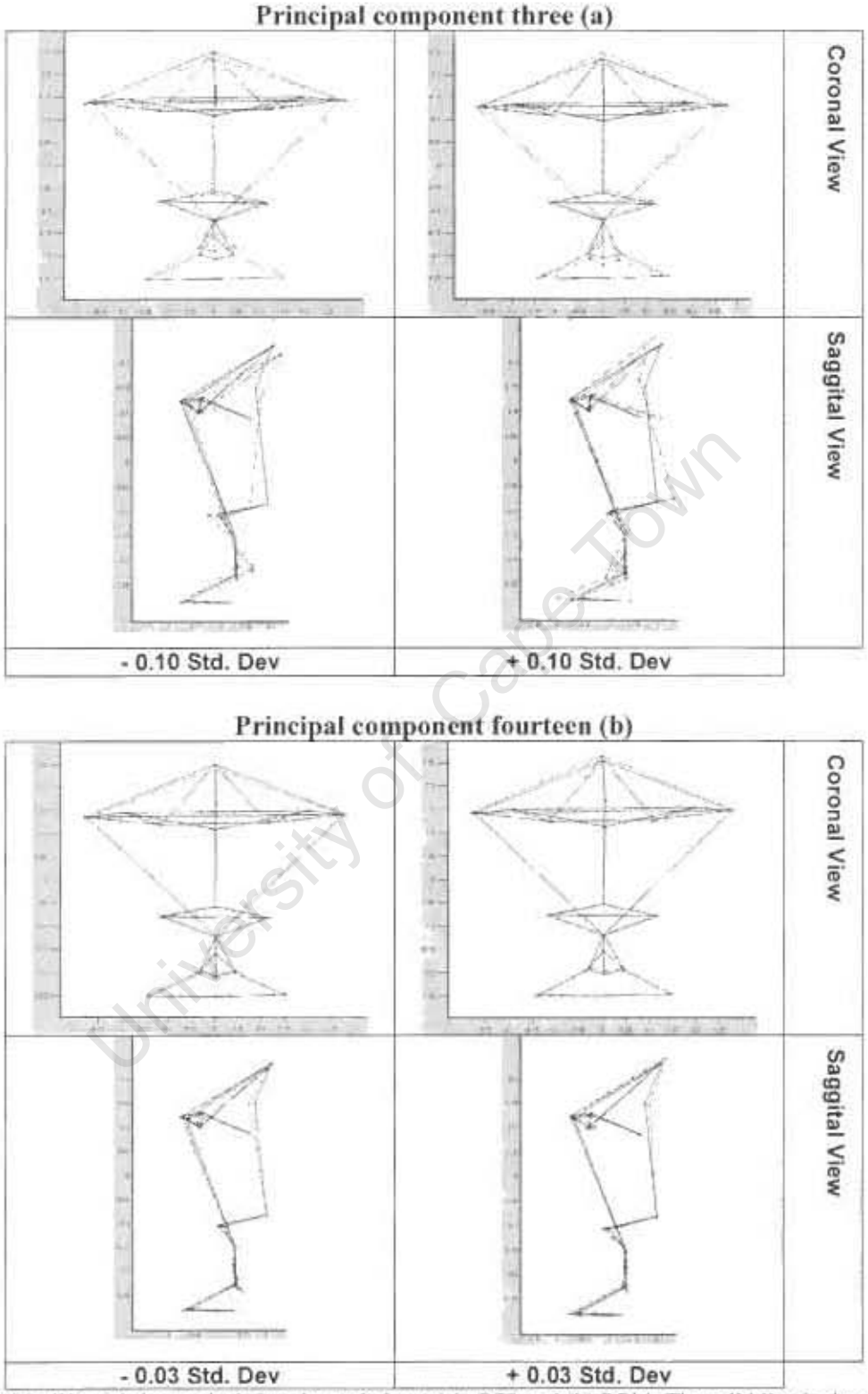


Figure 6.9: Warped shapes depicting the variation (a) in PC3 and (b) PC14. The solid line is the average and the dashed line is the superimposed warped shape.

REGION	PC3 (Mean ± 0.10 Std.Dev)	PC14 (Mean ± 0.03 Std.Dev)
Forehead	• Large anterior-posterior movement of seillion • Small anterior-posterior movement of nasion and glabella • Small superior-inferior movement of glabella	• Small anterior-posterior movement of nasion
Eyes	• ICD, OCD, and IPD display slight lateral-medial movement.	• No change
Midface	• Large anterior-posterior movement of pronasale	• Slight anterior-posterior movement of both alare
Philtrum/Lips	• Large anterior-posterior movement of upper vermillion border features (labiale superius and both crista philtri) • Large superior-inferior movement of upper vermillion border features (labiale superius and both crista philtri)	• Small superior-inferior movement of the stomion affecting upper lip thinness
Angles	• No change	• Slight change in OGO
General	• Greatly affects the mid-facial region especially the nose and mouth	• Slightly affects the forehead and the upper lip

Table 6.10: Description of the warped shapes along PC3 on left and PC14 on right.

6.3.8 Broad facial feature distinctions between FAS and normal subjects based on PC1.

Only PC1 has a clustering of subjects between the two groups that can easily be visually interpreted as shown in the scatter plot of PC3 vs. PC1 in Figure 6.5 or PC14 vs. PC1 in Figure 6.6. PC1 accounts for the largest shape variability, namely 24.3%. PC1 also has the greatest discriminating power between the FAS and normal subjects. These two factors combined means that any facial features affected by the mode of shape variation along PC1 are those that are most involved in the FAS facial phenotype. A description of these features is presented below in order of decreasing degree of manifestation. Since FAS subjects generally lie in the positive direction of PC1 it might be inferred that children with FAS have the following features:

- They have a smaller head circumference, if the measurement of circumference is between the glabella and the occipital protuberance (bony bulge on the occipital bone at the back of the skull), because there is retraction of the glabella to a more posterior position relative to the average shape.
- They have an underdeveloped mid face (midface hypoplasia). FAS faces tend to be flatter i.e. they seem to have their frontal facial features such as their noses, foreheads and lips approaching the same plane in the coronal view vertically. The nose is upturned and nasal protrusion is diminished. The nasal bridge length is increased.
- They have significantly longer philtrum lengths.
- FAS faces have shorter OCDs, but little lateral movement of the endocanthion accounting for the small PFLs that are evident in FAS subjects.
- They have slightly longer midfacial height; specifically, the upper facial height (described here as the distance from the nasion to the subnasale) is greatly increased. This, however, can also be attributed to the backward slanting of the nasion to a more posterior position.
- They tend to have lips that are small in width although there is no marked difference in thickness between normal and FAS faces along the direction of PC1.
- FAS faces have slightly smoother philtrums.

From the FAS direction of PC1 towards the direction where normal subjects lie (negative direction in Figure 6.5), the following shape variation occurs:

- Increase in nasal protrusion as the pronasale pushes forward as viewed from the sagittal plane. The nasion and subnasale remain in the same position as the average shape.
- The midfacial height decreases. As above, this could be also attributed to a more forward orientated nasion and glabella relative to the average.
- Anteroposterior retraction of eye features.
- The corners of the mouth become more retracted. Lips seem to be more visible from the side profile of their faces.
- ICD, OCD, PFL's and IPD become slightly larger than the average.
- The lips widen, although the thickness remains the same.
- Philtrum furrows become more pronounced.

6.4 Discussion

Application of the Procrustes approach to facial shape analysis is becoming more widespread in syndrome diagnosis (Hennessy, et al., 2002; Hammond, et al., 2004). The advantage of using PCA in conjunction with Procrustes analysis is the ability to give a comprehensive description of the overall facial shape with a small number of landmark measurements which are not conflicting because they are statistically unrelated (Halazonetis, 2004).

The mean shape difference between FAS and normal subjects in the first analysis is very small. Goodall's F test for difference in means, confirms that for the study sample, the facial shape range of FAS subjects' overlaps with the facial shape variability of normal

subjects for the features chosen in the first study. The within-group variability of the two groups is almost identical and together with the insignificant *p*-value (0.99968161) might suggest that there is no objective line of division between the FAS and normal groups. These findings seem to contrast with the findings of other researchers (Astley and Clarren, 1995; 1996). Ethnic variation could have been a factor in the findings presented here. Douglas and Viljoen, (2006) have reported that eye distance measurements taken from South African subjects do not consistently mirror published population norms from other parts of the world.

Figure 6.10 illustrates this overlap of features. The left image is a FAS subject and the right image a normal subject. Both subjects have smooth long philtrums and their upper lips have similar shapes. In the sample this occurrence is not unique. General Procrustes analysis attempts to compare shapes after removing scaling, rotation and position. In an ideal sample, it would be possible to delineate the typical FAS facial phenotype of smooth philtrum, small PFL and thin upper lip from that of normal subjects based on comparison of mean shape alone if these are indeed the most distinctive features between the two groups.

Figure 6.10: Left image; FAS subject exhibits smooth philtrum but normally developed upper lip. Right image; normal subject exhibits similar features as FAS subject.

There has not been much in the literature documenting assessment of mid-face hypoplasia from the saggital perspective in a quantitative way. Adding additional landmarks allowed for shape analysis of midfacial features in both the frontal and sagittal plane. Inclusion of additional landmarks relevant to the FAS facial phenotype resulted in better discrimination between the FAS and normal subjects. This was not surprising as the new configuration of landmarks was expected to be more representative of the FAS facial phenotype published in much of the literature. As more relevant landmarks are introduced in shape analysis the shape used in the analysis becomes more representative of the actual shape of the object. The same argument as in Section 6.4, Figure 6.11 is useful here. The difference between two shapes become more apparent as the shapes are better represented with more coordinates. The principal component (PC1) which accounts for the highest variation in shape across all subjects (24.3%) is also the mode of shape variability that discriminates best between the FAS and normal groups.

The flattening of the midface seems to dominate the mode of shape variability described by PC1. This agrees well with the facial dysmorphology associated with FAS in the literature (broad or depressed bridge of nose, short nose). Moore, et al., (2002) offer an explanation for the deficit in midfacial depth. They say there is a general reduction in all facial depth measurements but midfacial depth seems to be affected more than the upper facial and lower facial depths. The results presented in this chapter suggest that the anteroposterior movement of the nasion reduces the facial depth in the middle facial position of FAS subjects. Moore, et al., (2002), however, report frontal bossing (frontal bossing is the descriptive term for a prominent forehead) as suggested by near normal

values of minimal frontal breadth that they obtained for FAS subjects. This finding contrasts with the trend found here where the glabella seems to retract postero-anteriorly in FAS subjects.

The widespaced appearance of the eyes in FAS subjects is confirmed here (Astley and Clarren, 1995; 1996; Moore, et al., 2002). The endocanthion for the FAS face is slightly more laterally positioned from the facial midline. There is an accompanying great decrease in OCD. This gives the appearance of widely spaced eyes. The medial movement of the exocanthion is clearly evident, as PC1 is warped in the positive direction where the FAS subjects tend to lie (Figure 6.9). This, together with fact that the endocanthion is slightly more lateral, accounts for the reduction in palpebral fissures generally reported in FAS subjects.

PC3 seems to account more for the mode of shape variation that affects philtrum smoothness. The large overlap of principal scores between the FAS and normal groups in Figure 6.6 makes it difficult to assess the relative importance of philtrum smoothness in the FAS facial phenotype here. Along PC1, however, the philtrum in FAS subjects appears to be slightly smoother (i.e. the philtrum furrow and philtrum ridges become less distinguishable). While annotating more points along the philtrum ridges and philtrum furrow might have resulted in better shape analysis, the approximate shape here produced results consistent with the literature that FAS subjects generally have smooth long philtrums compared to normal people (Astley and Clarren, 1995; 1996; 2001). Philtrum length increases greatly in the FAS appearance as indicated by PC1. This has

also been reported as being phenotypical of FAS (Moore, et al., 2002). The explanation they offer that the ratio between philtrum length and nose size is so disproportionate that it makes individuals with FAS only seem to have long philtrums might not be accurate. The position of the subnasale in FAS subjects is slightly more superior than the average position in the frontal view of Figure 6.9(b). The upper vermillion border in FAS subjects is greatly more inferiorly positioned to that of the average mean shape in the same view. This is evidence of the philtrum being longer in FAS subjects than in normal subjects.

Microcephaly is shown here to be important in distinguishing FAS subjects but the FAS subjects were taken from a sample screened on the basis of growth retardation. The centroid sizes of the FAS subjects are smaller than those of normal subjects regardless of the normal subjects being on average younger (see Table 6.5). Centroid size is used here to indicate the relative smallness of the face since all landmark points are closer together if centroid size is reduced. Midface hypoplasia and philtrum length are also of great significance (see Figure 6.8) in the FAS facial shape. Palpebral fissure length is also a useful discriminating feature between the two groups. The increase in midfacial height for the FAS face seems to be in contrast with the literature, although its significance is evident in Figure 6.8. Philtrum smoothness is also clearly relevant. Upper lip thinness, which has become a common diagnostic criterion, does not seem to play as significant a role in distinguishing the FAS facial phenotype. The variation in lip thinness seems to be explained by PC3 (10.3% of total shape variability), although it is difficult to say where the FAS face lies compared to the normal face (i.e. does the FAS face have thicker or

thinner lips compared to the normal face), because of overlap of principle scores between the two groups along PC3.

The collection of landmark points was the greatest source of error in the facial shape analysis. An attempt was made to limit this error by assessing both intra-operator precision and reliability. The technical error of measurement *(TEM)* (1.2mm and 1.0mm for the first, and the second analysis, respectively) and the coefficient of relative variability (*CRV*) (1.03% and 0.69% for the first, and the second analysis, respectively) suggested a satisfactory precision. The intra-operator correlation coefficients for both studies also suggested a reasonable amount of reliability in obtaining landmark data (0.98 for the first study and 0.98 for the second study). Due to poor photographic quality, and in some cases subjectivity in landmark selection (estimating the position of some landmarks due the nature of their location on the face), some error was introduced. As mentioned in Chapter 5, however, Procrustes analysis treats all points equally and thus an error in the location of one point does not necessarily significantly affect the whole configuration. In contrast, conventional analyses might result in significant errors in measurement and therefore diagnosis, as a result of small errors in location of points, especially those that are used as reference points.

The limited number of landmark points that could be extracted was due to that fact few lay at points of extreme curvatures or contours. Increasing the number of landmark points in annotating features would increase the accuracy with which the shape of the feature could be described, thereby increasing the accuracy of shape analysis as shown in Figure 6.11. In this example, the shape of the upper vermillion border is more

accurately represented by increasing the number of landmark points compared to that presented in the study.

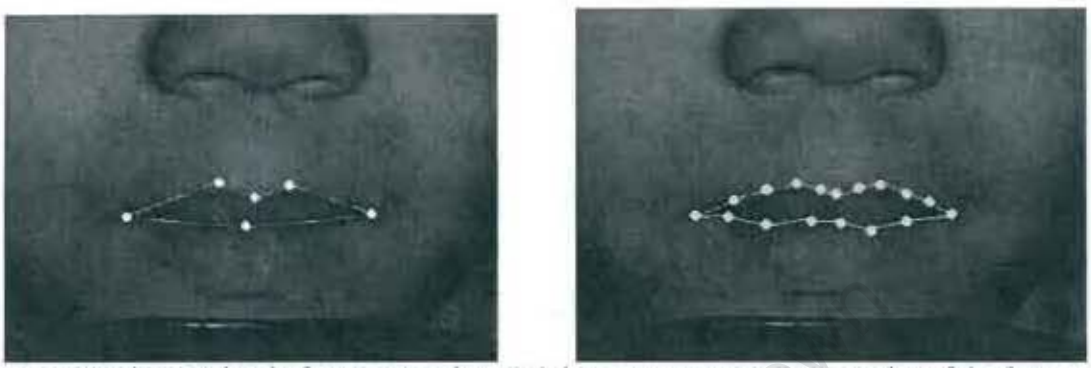


Figure 6.11: More points in feature extraction result in a more accurate representation of the feature as shown by the wire frame on the right.

It would be difficult, however, to match landmarks between subjects, as by definition, landmarks define easily identifiable extremes.

7 COMPARING LINEAR AND FACIAL METHODS

The discriminant analysis of facial landmark data presented in Chapter 4 was an attempt to find the minimal set of linear measurements that could be used in classifying subjects with FAS. In the process, the relative importance of each of the measurements/features was assessed compared to the others. The facial shape analysis attempted to find the modes of shape variation that differ between FAS and normal subjects. It is difficult to directly compare the two methods as the linear distance method uses inter-landmark distances, which contain form, i.e. shape plus size information and the facial shape method deals with shape only. With this in mind, when a quantity in shape analysis such as palpebral fissure length is said to be increasing, this is not in terms of absolute length, but is meant to convey that in the quantity appears to be increasing relative to the whole shape configuration. In the linear methods, absolute distances of quantities are being compared directly between the two groups for significant differences. The similarities between the two sets of results are now presented below in Table 7.1 for facial regions referred to in the literature as being affected in FAS (Astley and Clarren, 1995; 1996; Moore, et al., 2002).

The results of the discriminant analysis and the facial shape analysis are consistent. There is a correspondence between the relative importance of a feature in discriminant analysis and the degree to which it manifests itself in the facial shape analysis. Upper lip thinness is of little discriminating power compared to all the other measurements in the discriminant analysis. This is corroborated in facial shape analysis when the upper lip thinness is assessed along PC1 and found not to change greatly between the FAS and

normal faces. Inner canthal distance is shown to be relatively insignificant in discriminant analysis and does not greatly change between the FAS and normal faces in facial shape analysis. Features such as head circumference, and palpebral fissure length which have discriminating power in discriminant analysis, are shown to vary a great deal between the FAS and normal faces in facial shape analysis.

One feature shown to be important by shape analysis but which was not assessed by discriminant analysis is the decrease in the horizontal breadth of the upper lip in the FAS facial shape (see Figure 6.8). Upper lip circularity as a proxy measure for lip thinness increases with increase in the horizontal breadth. This might account for its inclusion in the discriminant analysis model of Section 4.4.1. Indeed, the average upper lip circularity of the normal subjects is greater than that of the FAS subjects (normal average is 23mm and the FAS average is 19.4mm). The significance of upper lip circularity, however, in discriminating between FAS and normal subjects, is small. A t-test for difference in mean values for upper lip circularity between FAS and normal subjects was not statistically significant (t-value=1.321817, p= 0.192627). An increase in philtrum length in the FAS face was also noted in facial shape analysis (see Figure 6.8) although not assessed in discriminant analysis.

It was possible to include landmarks that require some degree of estimation in photographs (Section 6.1.3) in the Procrustes analysis, because the method is robust to errors in the location of one point as it treats all points equally in a facial configuration.

The effects of errors in point localisation on the linear measurements used in discriminant analysis would affect the results of discriminant analysis more severely.

Discriminant analysis	Facial shape analysis (second study)
1) Head circumference	
This is the most discriminating measurement between the FAS and control groups. The FAS sample was, however, already screened for growth retardation	The anteroposterior retraction of the glabella is a large difference between the FAS and the normal groups along the first principal component
2) Midface Hypoplasia	
Not assessed	The broad or depressed bridge of nose and short upturned nose are very prominent in the FAS face described by first principal component
3) Abnormalities in the eye region	
Palpebral fissure length is the second most discriminating measurement, followed by inter-pupillary distance. The other eye measurement, ICD, does not seem to discriminate very well between the two groups.	Palpebral fissure length and inter-pupillary distance are greatly decreased in the FAS face compared to the normal face. Outer canthal distance is also greatly diminished in the FAS face. Inner canthal distance does seem to increase (giving the appearance of slightly widely spaced eyes for the FAS face) but not as significantly as the distances above
4) Upper lip region abnormalities	
Compared to the all the measurements used in the analysis except inner canthal distance, upper lip circularity (as a measurement of lip thinness) was not very successful as a discriminating factor between two groups. Philtrum smoothness and philtrum length were not included in the analysis	Philtrum length is increased greatly in the FAS face compared to the normal face. It is important to note that large philtrum length is considered midface hypoplasia, as the philtrum can also be considered as part of the midface. Upper lip thinness does not seem to be different between the FAS and normal face, although the horizontal length is increased greatly in the FAS face. The smoothening of the philtrum is evident in the FAS face.

Table 7.1: Similarities between the discriminant analysis and facial shape analysis results.

8 CONCLUSION AND RECOMMENDATIONS

8.3 Conclusions

Discriminant analysis was performed to determine those facial features (concentrating on the eyes) that can best be used to diagnose FAS. Occipital frontal circumference (OFC), palpebral fissure length (PFL) and interpupillary distance (IPD) alone could be used to classify subjects with a specificity of 86% and sensitivity of 100% in the first study for the hold out sample (number of subjects =25) after training using the analysis sample. A specificity of 95% and sensitivity of 100% was obtained across all 67 subjects using the discriminant equation, *D. Score* = 35.71743– 0.55104 (OFC) – 0.69307 (PFL) + 0.16932 (IPD). In the second study OFC, PFL and IPD and upper lip circularity (ULC) could be used to classify the hold out sample (number of subjects =20) with a specificity of 85.7% and sensitivity of 100% after training using the analysis sample. A specificity of 94% and sensitivity of 100% was obtained across all 49 subjects using the discriminant equation, *D.Score* = 41.42437- 0.64788 (OFC) - 0.64522 (PFL) + 0.14720 (IPD) - 0.04312(ULC). The same subject was misclassified in each of the discriminant analysis studies.

Facial shape analysis of stereo-photogrammetric image pairs was performed to see how the FAS facial shape differs from that of the normal face, and the extent of overlap of the two categories of faces was assessed.

The use of Procrustes analysis followed by principal component analysis to reduce dimensionality and explore shape variability was shown to be a successful method for analysing the modes of facial variation between FAS and non-FAS subjects. Head circumference and midfacial features together with eye features, were found to differentiate between FAS and normal subjects. The features reported to be phenotypic of FAS in the literature, with the exception of upper lip thinness, were shown to be relevant to FAS diagnosis on a local population of children. Ethnic variation may influence the cranio-facial morphological changes caused by facial syndromes (Farkas, 1994), and it thus becomes important to assess facial anomalies relative to the local population norms. Recently, Douglas and Viljoen (2006), found that linear eye measurements on a South African population of seven year old black children obtained using stereo-photogrammetry were different from those obtained in any other population. While normal population reference values for linear facial measurements may differ in South African children, facial shape analysis indicates that the shape variation associated with FAS in this country corresponds to that observed elsewhere.

Obtaining surface information from stereo-photogrammetric images has limitations. The landmark based, stereo-photogrammetric approach presented here was limited by the number of biologically homologous landmarks that are easily reproducible. Points of biological correspondence on soft tissue surfaces like the cheek, chin and forehead are not easy to find, although some shape information found on these surfaces could have been useful in delineating facial phenotypes of some syndromes (Hammond, et al., 2004). The advantages of using a full three-dimensional surface scan of the face become

obvious in light of the above limitation of the stereo-photogrammetric approach in this project. Three-dimensional surface scans allow the use of traditional landmark based geometric analysis with the added advantage that surface information of the face is easily obtainable and can be factored into the facial analysis. The dense surface model approach (Hammond, et al., 2005; Hammond, et al., 2004) using three-dimensional surface scans of the whole face or localised patches of the face has yielded accurate inter-syndrome discrimination rates. The disadvantage of full surface scans is the inhibiting cost of the scanners.

South Africa has many competing health issues including tuberculosis, sexually transmitted diseases, and particularly, acquired immunodeficiency syndrome (Viljoen, et al., 2003). For situations where prevalence of FAS is suspected to be very high, as is the case in some South African communities, cost, ease and accuracy (making diagnosis relevant to the ethnic population being surveyed) of surveillance techniques become of extremely important. Stereo-photogrammetry and the diagnosis methods using only landmark data explored here point towards an efficient way of large scale assessment of FAS prevalence in the Southern African context.

8.4 Recommendations

A larger study sample, particularly for discriminant analysis, would have produced more conclusive results and thus it is recommended that if the study is repeated, a larger study population be used. In their analyses, Astley and Clarren (1996) used a sample of 126 subjects (42 with FAS and 84 without FAS) and Moore, et al., (2001), 131 subjects (41

FAS, 59 with partial fetal alcohol syndrome and 31 controls) to explore the use of facial shape measurements for FAS diagnosis. The discriminant analysis in this project in comparison had relatively small study sample namely 67 subjects in the first analysis and 49 in the second analysis. The facial shape analysis methods used in the project however, are relatively robust to small study samples, and thus the reliability of the results is not seriously compromised. It is important to note however that the FAS and normal samples were not matched according to sex. The craniofacial shape of males has been shown to be markedly different to that of females in humans (Hennessy, et al., 2002). The effects of sexual dimorphism on the results could be further explored by using a sex-matched study sample for the facial shape analysis. The study population consisted of one racial ethnicity thus the conclusions drawn above can only be applied to individuals of mixed race and further work should to study incorporate the differences in the FAS facial shape between and racial groups.

Automated annotation of points or landmarks to populate areas between corners or contours of features would reduce the time taken by analysis. Advanced feature extraction and landmark detection algorithms could be utilized to obtain more accurate information from the images. The automatic feature extraction method developed by Douglas, et al., (2003) could be extended to other features on the face which display a large contrast ratio compared to the tissue surrounding them.

Marking bony or soft tissue landmarks by a trained person prior to taking the stereo images could make landmark identification more accurate but also more time consuming and intrusive. Automatic matching of landmarks in the right and left images would

contribute to the accuracy of three-dimensional landmark location. Another way to improve landmark detection is the use of three cameras instead of two to obtain a frontal perspective. Landmarks, which appear on the right or left view, are sometimes difficult to view on images of the opposite side.

Diagnosis of a patient using facial shape analysis could be done through calculating a subject's position in the shape space described by the principal components. The preceeding work has shown that the facial phenotype of the average FAS face is in a location different from the normal face shape in the shape space described by, especially, principal component one. If a measure of distance from the origin in this shape space is obtainable, then, using a threshold scheme on this distance, a subject could be classified as being at risk of FAS or not. Thus a quantitative diagnosis could be obtained in large scale surveillance efforts that can quickly and accurately assess FAS risk and allow the operator of the system to make recommendations for a full clinical diagnosis using traditional methods.

9 REFERENCES AND BIBLIOGRAPHY

Abdel-Aziz, Y. I. and Karara, H. M. (1971) Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In Proceedings of the Symposium on Close-Range Photogrammetry, pp. 1-18, American Society of Photogrammetry, Falls Church, VA

Astley, S. J. and Clarren, S. K. (1995) "A fetal alcohol syndrome screening tool". Alcohol Clin Exp Res **19**, 1565-1571

Astley, S. J. and Clarren, S. K. (1996) "A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome". J Pediatr **129**, 33-41

Astley, S. J. and Clarren, S. K. (2001) "Measuring the facial phenotype of individuals with prenatal alcohol exposure: correlations with brain dysfunction". Alcohol Alcohol **36**, 147-159

Astley, S. J.,Stachowiak, J.,Clarren, S. K. and Clausen, C. (2002) "Application of the fetal alcohol syndrome facial photographic screening tool in a foster care population". J Pediatr **141**, 712-717

Astley, S. J. (2004) Diagnostic Guide for Fetal Alcohol Spectrum Disorders: The 4-Digit Diagnostic Code. University Of Washington, Seattle, Washington

Bookstein, F. L. (1997) "Landmark methods for forms without landmarks: morphometrics of group differences in outline shape". Med Image Anal **1**, 225-243

Burd, L.,Martsolf, J. T.,Klug, M. G. and Kerbeshian, J. (2003) "Diagnosis of FAS: a comparison of the Fetal Alcohol Syndrome Diagnostic Checklist and the Institute of Medicine Criteria for Fetal Alcohol Syndrome". Neurotoxicol Teratol **25**, 719-724

Cangelosi, R. and Goriely, A. (2006) "Component retention in a principal component analysis with application to cDNA microarray data". Submitted to the Journal of Computational Biology on 1/10/2006 (in press)

Clarren, S. K.,Sampson, P. D.,Larsen, J., et al. (1987) "Facial effects of fetal alcohol exposure: assessment by photographs and morphometric analysis". Am J Med Genet **26**, 651-666

Douglas, T. S.,Martinez, F.,Meintjes, E. M.,Vaughan, C. L. and Viljoen, D. L. (2003) "Eye feature extraction for diagnosing the facial phenotype associated with fetal alcohol syndrome". Med Biol Eng Comput **41**, 101-106

Douglas, T. S. (2004) "Image processing for craniofacial landmark identification and measurement: a review of photogrammetry and cephalometry". Comput Med Imaging Graph **28**, 401-409

Douglas, T. S. and Viljoen, D. (2006) "Eye distance measurements in 7 year-old black South African children". Annals of Human Biology (in press)

Dryden, I. L. and Mardia, K. V. (1998) Statistical Shape Analysis. John Wiley & Sons, New York

Farkas, L. G. (1994) Anthroprometry of the Head and Face. Raven Press, New York

Garson, D. (2005) "PA 765 Statnotes: An Online Textbook". North Carolina State University, http://www2.chass.ncsu.edu/garson/pa765/statnote.htm, Last accessed September 2005

Halazonetis, D. J. (2004) "Morphometrics for cephalometric diagnosis". Am J Orthod Dentofacial Orthop **125**, 571-581

Hall, J. G.,Froster-Iskenius, U. G. and Allanson, J. E. (1989) Handbook of Normal Physical Measurements. Oxford University Press, New York

Hammer, O.,Harper, D. A. T. and Ryan, P. D. (2005).PAST- PAlaeontologicalSTatistics ver. 1.35, computer program, http://folk.uio.no/ohammer/past Last date accessed: 12 September 2005

Hammond, P.,Hutton, T. J.,Allanson, J. E., et al. (2004) "3D analysis of facial morphology". Am J Med Genet A **126**, 339-348

Hammond, P.,Hutton, T. J.,Allanson, J. E., et al. (2005) "Discriminating power of localized three-dimensional facial morphology". Am J Hum Genet **77**, 999-1010

Hennessy, R. J. and Moss, J. P. (2001) "Facial growth: separating shape from size". Eur J Orthod **23**, 275-285

Hennessy, R. J.,Kinsella, A. and Waddington, J. L. (2002) "3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: initial application to sexual dimorphism". Biol Psychiatry **51**, 507-514

Huang, J.,Jain, A.,Fang, S. and Riley, E. P. (2005) "Using Facial Images to Diagnose Fetal Alcohol Syndrome (FAS)". In IEEE Prcoceedings of International Conference on Information Technology: Coding and Computing pp. 66-71, IEEE

Hunter, A. G. (2002) "Medical genetics: 2. The diagnostic approach to the child with dysmorphic signs". Cmaj **167**, 367-372

Jamison, P. L. and Ward, R. E. (1993) "Brief communication: measurement size, precision, and reliability in craniofacial anthropometry: bigger is better". Am J Phys Anthropol **90**, 495-500

Jolliffe, I. (1986) Principal Component Analysis. Springer-Verlag, New York

Jones, K. L. and Smith, D. W. (1973) "Recognition of the fetal alcohol syndrome in early infancy". Lancet **2**, 999-1001

Klingenberg, C. P. and Monteiro, L. R. (2005) "Distances and directions in multidimensional shape spaces: implications for morphometric applications". Syst Biol **54**, 678-688

Lachenbruch, P. A. (1977) "Some misuses of discriminant analysis". Methods Inf Med **16**, 255-258

Mathworks Inc (2004).MATLAB ver. 6.50, computer program, www.mathworks.com Last date accessed: 20 January 2005

May, P. A.,Brooke, L.,Gossage, J. P., et al. (2000) "Epidemiology of fetal alcohol syndrome in a South African community in the Western Cape Province". Am J Public Health **90**, 1905-1912

Meintjes, E. M.,Douglas, T. S.,Martinez, F., et al. (2002) "A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome". Med Eng Phys **24**, 683-689

Moore, E. S.,Ward, R. E.,Jamison, P. L., et al. (2001) "The subtle facial signs of prenatal exposure to alcohol: an anthropometric approach". J Pediatr **139**, 215-219

Moore, E. S.,Ward, R. E.,Jamison, P. L., et al. (2002) "New perspectives on the face in fetal alcohol syndrome: what anthropometry tells us". Am J Med Genet **109**, 249-260

Robinson, D. L.,Blackwell, P. G.,Stillman, E. C. and Brook, A. H. (2001) "Planar Procrustes analysis of tooth shape". Arch Oral Biol **46**, 191-199

Rohlf, F. J. (2000) "Statistical power comparisons among alternative morphometric methods". Am J Phys Anthropol **111**, 463-478

Sheets, H. D. (2003).IMP: ThreeDPCA6 ver. beta, computer program, http://www2.canisius.edu/~sheets/moremorph.html Last date accessed: 25 August 2005

Sheets, H. D. (2004).IMP: Simple3 ver. beta, computer program, http://www2.canisius.edu/~sheets/moremorph.html Last date accessed: 25 August 2005

Spiegel, M. R. (1968) Mathematical Handbook of Formulas and Tables. McGraw-Hill New York

Statsoft.Inc (2004).STATISTICA (data software analysis system) ver. 7, computer program, www.statsoft.com Last date accessed: 15 July 2005

Statsoft.Inc (2004) "Electronic statistics Textbook". Tulsa,OK, http://www.statsoft.com/textbook/stathome.html., Last accessed 20 Septempber 2005

Stegmann, M. B. and Gomez, D. D. (2002) A Brief Introduction to Statistical Shape Analysis. pp. 15, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Denmark

Viljoen, D.,Graig, P.,Hymbaugh, M. P. H.,Boyle, C. and Blount, S. (2003) Fetal Alcohol Syndrome --- South Africa, 2001. In Morbidity and Mortality Weekly Report (Ward, R. E., ed.), pp. 660-662, Centers for Disease Control and Prevention, Atlanta

Ward, R. E. and Jamison, P. L. (1991) "Measurement precision and reliability in craniofacial anthropometry: implications and suggestions for clinical applications". J Craniofac Genet Dev Biol **11**, 156-164

Warren, K. R.,Calhoun, F. J.,May, P. A., et al. (2001) "Fetal alcohol syndrome: an international perspective". Alcohol Clin Exp Res **25**, 202S-206S

Yeung, K. Y. and Ruzzo, W. L. (2000) An empirical study on Principal Component Analysis for clustering gene expression data. Unievrsity of Washington, Seattle